Acta Geodaetica et Geophysica Hungarica

VOLUME 35, NUMBER 1, 2000

INCLUDING SECOND PART OF THE SIXTH WINTER SEMINAR SOPRON (WSS '98) "VARIATION OF THE EARTH ROTATION"

Akadémiai Kiadó, Budapest

AGG 35 (1) 1–103 (2000) HU ISSN 1217-8977

ACTA GEODAETICA ET GEOPHYSICA HUNGARICA

A Quarterly Journal of the Hungarian Academy of Sciences

Acta Geodaetica et Geophysica Hungarica (AGG) publishes original reports and reviews on geodesy and geophysics in English.

AGG is published in yearly volumes of four issues by

AKADÉMIAI KIADÓ
Prielle K. u. 4, H-1117 Budapest, Hungary
http://www.akkrt.hu

Manuscripts and editorial correspondence should be addressed to

AGG Editorial Office
Geodetical and Geophysical Research Institute
P.O. Box 5, H-9401 Sopron, Hungary
Phone: (36-99) 314-290
Fax: (36-99) 313-267
E-mail: actagg@ggki.hu
http://www.ggki.hu/agg/

Subscription information

Orders should be addressed to

AKADÉMIAI KIADÓ
P.O. Box 245, H-1519 Budapest, Hungary
Fax: (36-1) 464-8221
E-mail: kiss.s@akkrt.hu

Subscription price for Volume 35 (2000) in 4 issues US$ 180.00, including normal postage, airmail delivery US$ 20.00.

Printed in Hungary
PXP Ltd., Budapest

AGeod 35 (2000) 1

CONTENTS

Acta Geod. Geoph. Hung., Vol. 35(1), p. 1 (2000)

Editorial note

This issue contains the second part of the material of the Sixth Winter Seminar Sopron (WSS '98) "Variation of the Earth Rotation", namely the first three papers in this issue pp. 1–48, followed by general papers. The manuscript of C Denis' paper, the fourth announced in the previous first part of the WSS '98 material was not available.

J Verő

1217-8977/2000/$ 5.00

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 3–24 (2000)

CONTRIBUTIONS OF THE INTERNATIONAL GPS SERVICE FOR GEODYNAMICS (IGS) TO GLOBAL GEODYNAMICS, ASTRONOMY AND ATMOSPHERE SCIENCE: REVIEW AND OUTLOOK

G BEUTLER[1]

Keywords: GPS; International GPS Service; satellite orbits; station coordinates; terrestrial reference frames

Contents

The IGS is an International Scientific Service working under the auspices of the IAG (International Association of Geodesy). The IGS is closely cooperating with the IERS (International Earth Rotation Service). The IGS is a FAGS service since 1996.

IGS objectives and mission

The IGS terms of reference contained, e.g., in the IGS colleague directory state:

— IGS collects, archives and distributes GPS observation data sets

— These data sets are used by the IGS to generate:

- high accuracy GPS satellite ephemerides,
- Earth rotation parameters,
- coordinates and velocities of the IGS tracking stations,
- GPS satellite and tracking station clock information,
- atmosphere information.

[1]Astronomical Institute, University of Berne, Switzerland

1217-8977/2000/$ 5.00

The accuracies are sufficient to support current scientific objectives including the improvement and the realization of the accessibility to the ITRF, the International Terrestrial Reference Frame.

The IGS accomplishes its mission through the following components

— a network of tracking stations,

— Data Centers, where Operational, Regional, and Global Data Centers are associated with stations, regions, and the global IGS network,

— Analysis Centers generating daily global products without interruption,

— Associated Analysis Centers generating unique products like ionosphere models, regional networks, etc., possibly only for a limited amount of time per year

— an Analysis Center Coordinator,

— a Central Bureau, and

— an International Governing Board, consisting of 15 members.

IGS development

Chronicle of IGS-events 1989–1998

Date	Event
Aug 89–Feb 90	IAG General Meeting in Edinburgh. Original ideas by I I Mueller, G Mader, W G Melbourne, B Minster and R E Neilan
16 Mar 90	IAG Executive Committee Meeting in Paris decides to establish a Working Group to explore the feasibility of an IGS under IAG auspices with I I Mueller as chairman
25 Apr 90	The Working Group is redesignated as the IAG Planning Committee for the IGS in Paris
02 Sep 90	Planning Committee Meeting in Ottawa to prepare the Call for Participation (CFP)
01 Feb 91	CFP mailed. Letters of intent due 1 April 1991
01 Apr 91	CFP attachments mailed to interested parties
01 May 91	Proposals due
24 Jun 91	Proposals evaluated and accepted in Columbus, Ohio
17 Aug 91	Planning Committee reorganized and renamed IGS Campaign Oversight Committee (OSC) at 20st IUGG General Assembly in Vienna
24 Oct 91	First IGS Campaign Oversight Committee Meeting in Greenbelt, MD. Preparation of the 1992 IGS Test Campaign, scheduled for 21 June – 23 September 1992 and for a two weeks intensive campaign called Epoch 92

Chronicle of IGS-events 1989–1998 (contd.)

Date	Event
17 Mar 92	Second IGS OSC Meeting at OSU, Columbus, Ohio
04 May 92	Communication tests
21 May 92	IGS e-mailbox established at University of Berne
21 Jun 92	Start of 1929 IGS Test Campaign
01 Jul 92	First results about 2 weeks after beginning of campaing
27 Jul 92	Start of Epoch-92 (2 weeks duration)
23 Sep 92	Official end of campaign; data collection and processing continue
15 Oct 92	Third IGS OSC Meeting at GSFC in Greenbelt, MD
01 Nov 92	Start of IGS PILOT Service. Regular weekly orbit comparisons performed by IGS Analysis Center Coordinator
24 Mar 93	1993 IGS Workshop and fourth IGS OSC Meeting at University of Berne
27 May 93	Fifth IGS OSC Meeting at AGU spring meeting in Baltimore, MD
09 Aug 93	IAG Symposium in Beijing. Official IAG approval of the IGS
12 Oct 93	Analysis Center Workshop in Ottawa
18 Oct 93	Network Operations Workshop and first IGS Governing Board Meeting in Silver Spring, MD
08 Dec 93	GB Business Meeting in San Francisco during AGU meeting
01 Jan 94	Start of official IGS. Production of combined IGS Orbits, Central Bureau Information System (CBIS) installed
21 Mar 94	Combined IERS/IGS Workshop in Paris (one week)
25 Mar 94	Second IGS Governing Board Meeting in Paris
30 Nov 94	IGS Workshop densification of the ITRF through regional GPS analyses in Pasadena
06 Dec 94	Third IGS Governing Board Meeting in San Francisco
15 May 95	IGS Workshop Special Topics and New Directions in Potsdam
06 Jul 95	Fifth IGS Governing Board Meeting in Boulder
12 Dec 95	Fifth IGS Governing Board Meeting in San Francisco
19 Mar 96	IGS Analysis Center Workshop in Silver Spring
30 Jun 96	Major Review of Processing (use of ITRF94, subdaily ERP terms, start of activities for new AACs, compliance to IERS Conventions 96)
16 Oct 96	Sixth IGS Governing Board Meeting in Paris
12 Mar 97	IGS Analysis Center Workshop in Pasadena
05 Sep 97	Sevent IGS Governing Board Meeting in Rio de Janeiro
11 Dec 97	Eighth IGS Governing Board Meeting in San Francisco
12 Dec 97	IGS retreat 1997 in San Francisco
09 Feb 98	IGS Analysis Center Workshop in Darmstadt

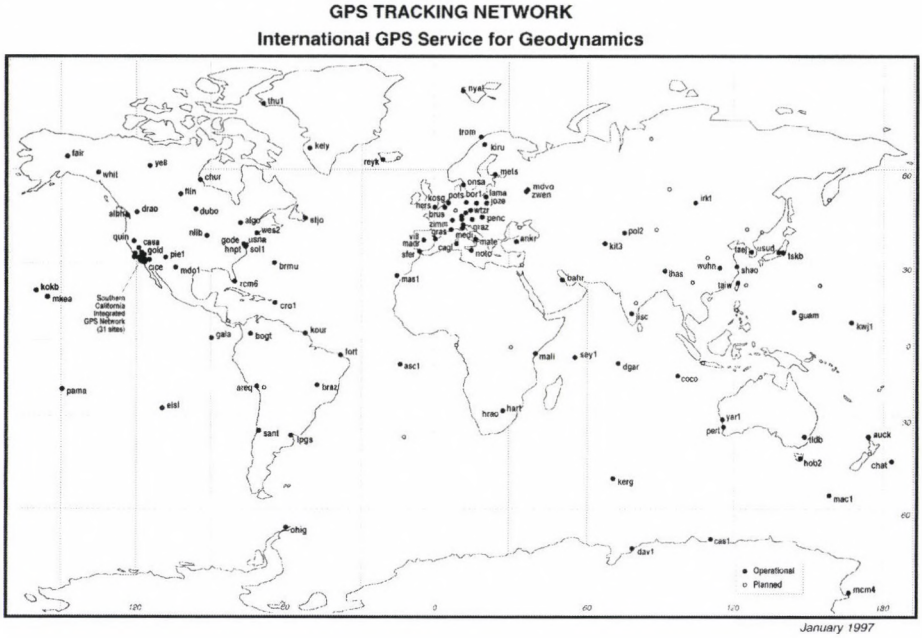


Fig. 1. GPS tracking network. International GPS service for geodynamics

Present structure of the IGS

Today the IGS consists of

— a Global Network of about 80 stations (Fig. 1),

— three global Data Centers

  – CDDIS (Crustal Dynamics Data Information System at Goddard Space Flight Center, USA),
  – IGN (Institut Geographique National, France),
  – SIO (Scripps Institution of Oceanography),

— seven Analysis Centers, namely CODE, EMR, ESA, JPL, GFZ, NGS and SIO,

— the Analysis Center Coordinator Jan Kouba, Natural Resources, Canada,

— the Central Bureau with Ruth Neilan as director,

— the International IGS Governing Board with Gerhard Beutler as chairman.

The IGS Network

Since 1992 the IGS Network went through a considerable development. Important empty areas like those in South America and in Russia could be filled with permanent receivers. At present Africa is considered as the most critical IGS area.

Table I. Some IGS institutions

Abbreviation	Institution	Function	Participation
CODE	at Astronomical Institute, Uni Berne Collaboration of AIUB, IfAG, IGN and Swiss Federal Office of Topography	A	since 21 June 1992
NRCan	Natural Resources, Canada (former EMR)	A	since September 1992
ESOC	European Space Agency, Germany	A	since 21 June 1992
GFZ	Geoforschungszentrum, Germany	A	since 21 June 1992
JPL	Jet Propulsion Laboratory, USA	A	since 21 June 1992
NOAA	Nat. Oceanic and Atmosph. Adm. USA	A	since March 1993
SIO	Scripps Institution of Oceanography, USA	A	since 21 June 1992
UTX	University of Texas at Austin, USA	A	21 June to 23 Sep 1992
CDDIS	Goddard Space Flight Center, USA	D	since 21 June 1992
IGN	Institut Geographique National, France	D	since August 1992
SIO	Scripps Institution of Oceanography, USA	D	since 21 June 1992
OSU	Ohio State University	C	21 June 92 – Dec 1993
NRCan	Natural Resources, Canada (former EMR)	C	since January 1994
JPL	Jet Propulsion Laboratory, USA	B	since 21 June 1992

A: Analysis Center, D: Global Data Center, C: Coordinator, B: Central Bureau

Present structure of the IGS: The Central Bureau Information System (CBIS)

The Central Bureau Information System (CIBS) was installed with the start of the official service on January 1, 1994.

The CBIS contains

— information concerning the availability of IGS observational data,

— information concerning IGS data and analysis centers,

— information concerning IGS stations,

— information concerning status of the GPS,

— the official IGS orbits and pole coordinates,

— IGS-mail, IGS-reports.

The CBIS is available through internet (ftp: igscb.jpl.nasa.gov, dir: /igscb, world wide web, (http://igscb.jpl.nasa.gov) and e-mail).

Table II

Perturbation	Acceleration m/s^2	Orbit error after one day (m)
Two-body term of Earth's gravity field	0.59	∞
Oblateness of Earth	$5 \cdot 10^{-5}$	10000
Lunar gravitational attraciton	$5 \cdot 10^{-6}$	3000
Solar gravitational attraction	$2 \cdot 10^{-6}$	800
Other terms of Earth's grav. field	$3 \cdot 10^{-7}$	200
Radiation pressure (direct)	$9 \cdot 10^{-8}$	200
y-bias	$5 \cdot 10^{-10}$	2
Fixed body tides	$1 \cdot 10^{-9}$	0.3

IGS products and quality: satellite orbits

Forces (accelerations) acting on GPS satellites. Radiation pressure (direct, y-bias and others) have to be modeled in the orbit determination process. Interesting new RPR models were presented at the 1998 Darmstadt IGS Workshop.

IGS makes available through its Global Data Centers the orbits of the seven IGS Analysis Centers. In addition each week the official IGS orbit is produced:

— A weight (corresponding to the standard deviation of the transformation of the individual solution to the unweighted mean orbit of the seven centers) is assigned to each analysis center.

— The official IGS orbit consists of the satellite positions computed as the weighted mean of the positions as computed by the IGS analysis centers.

— The mean errors and the transformation parameters of the individual solutions with respect to the IGS orbit are made available every week for each day of the (preceeding) week in an IGS-Report.

Motivation

Bauersime (1983) computed the error Δx in a component of a baseline of length l as a function of an orbit error of size ΔX:

$$\Delta x = \frac{l}{d} \cdot \Delta X \approx \frac{l \text{ (km)}}{25000 \text{ (km)}} \cdot \Delta X \tag{1}$$

d is the mean distance between the survey area and the satellite system.

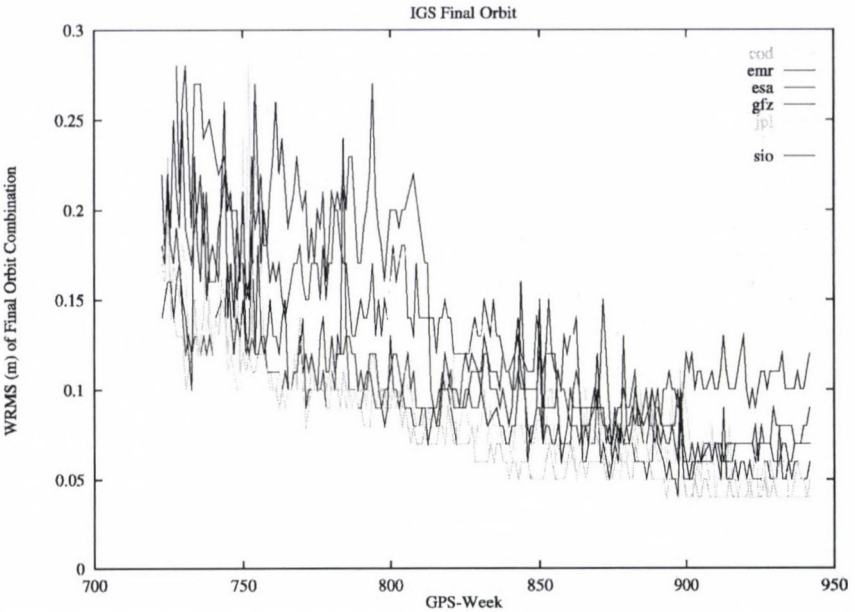


Fig. 2. IGS final orbit quality from November 1993 – February 98 as computed by the IGS Analysis Center Coordinator. Most IGS analysis centers are now consistent to better than 10 cm rms per satellite coordinate for the IGS final orbits

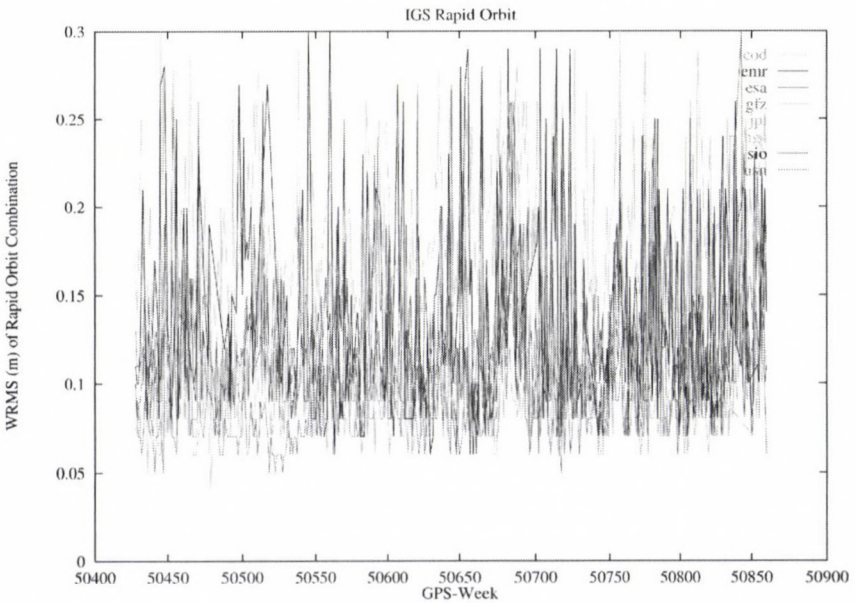


Fig. 3. IGS rapid orbit quality from December 96 – February 98 as computed by the IGS Analysis Center Coordinator. Consistency level 10–15 cm

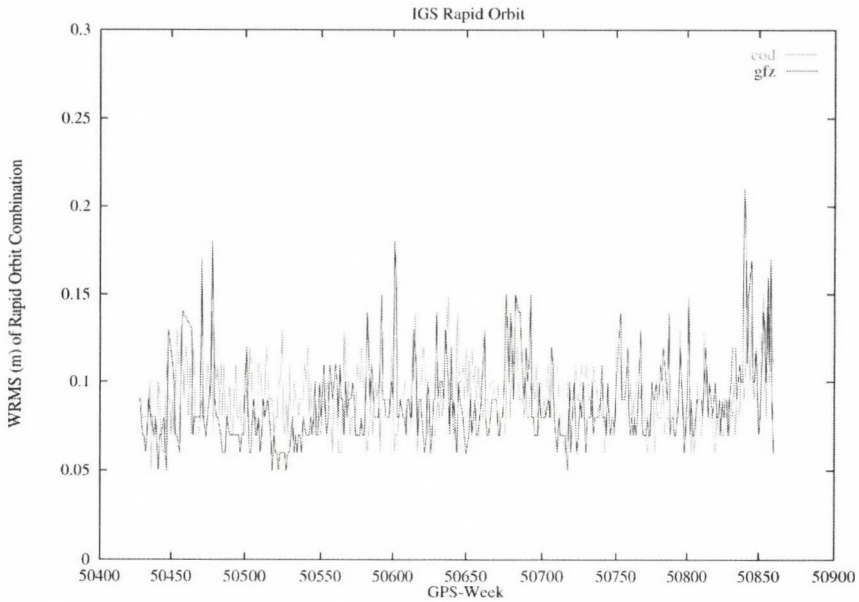


Fig. 4. IGS rapid orbit quality from December 96 – February 98 as computed by the IGS Analysis Center Coordinator. Only GFZ and CODE rapid orbits shown

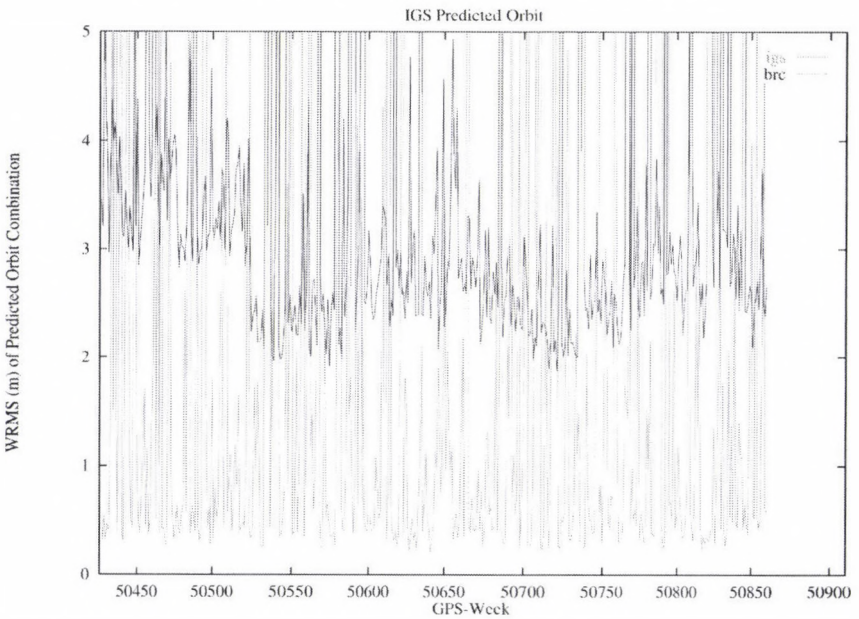


Fig. 5. IGS predicted orbit quality from December 96 – February 98 as computed by the IGS Analysis Center Coordinator. Prediction quality over second day of predictions

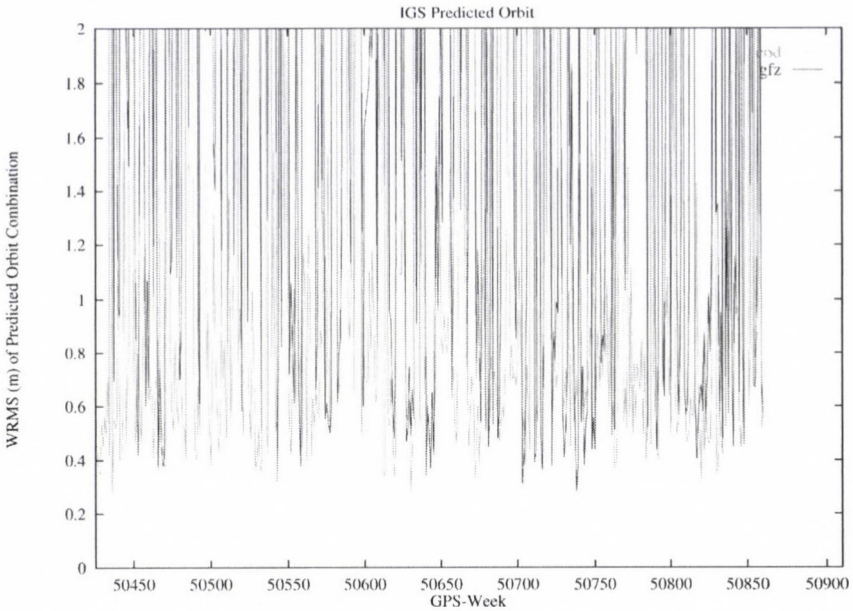


Fig. 6. Predicted orbit quality from December 96 – February 98 as computed by the CODE and GFZ Analysis Centers. Prediction quality over second day of predictions

Table III. Impact of orbit errors on the estimated station coordinates for different baseline lengths

Orbit error m	Baseline length km	Baseline error ppm	Baseline error mm
25	1	1	1
25	10	1	10
25	100	1	100
25	1000	1	1000
2.5	10	.1	1
2.5	100	.1	10
2.5	1000	.1	100
.25	100	.01	1
.25	1000	.01	10
.05	100	.002	–
.05	1000	.002	.5

Table IV. Orbit types available today

Orbit type	Quality m	Delay of availability
Broadcast orbit	3.0	real time
CODE predicted orbit	.25	real time
IGS predicted orbit	.25	real time
CODE rapid orbit	.10	after 12 hours
IGS rapid orbit	.07	after 24 hours
IGS final orbit	.05	after 11 days

Orbit type	Quality m	Available at
Broadcast orbit	3.0	broadcast message
CODE predicted orbit	.25	real time, IGS Data Centers
IGS predicted orbit	.25	real time, CODE
CODE rapid orbit	.10	CODE through FTP
IGS rapid orbit	.07	IGS Data Centers
IGS final orbit	.07	IGS Data Centers

IGS product changes on June 30, 1996

All IGS combined products (orbits, EOPs, station coordinates) are based on ITRF94, realized by the published positions and velocities of the stations TROM, MADR, KOSG, WETT, HART, FAIR, KOKB, YELL, ALGO, GOLD, SANT, TIDB, YAR1.

The IGS preliminary orbit/clocks became the rapid orbit, produced daily within 24 h. The IGS rapid orbit became the IGS final orbit. The former IGS final orbit is discontinued. All IGS solutions employ the 1996 IERS diurnal and semidiurnal EOP model. The operational broadcast ephemerides are included in the IGS rapid orbit summaries for information (they have no weight in the combination). Polar motion rates are taken into account for the combination process.

IGS predicted orbits

Several IGS Analysis Centers started predicting GPS orbits based on the rapid orbits (available within 12 to 24 hours after the observations. The Analysis Coordinator started comparing and combining these predictions with GPS week 895. There is an official predicted IGS orbit available for every day 30 minutes before start of new day. The quality of the IGS predictions (for "normal" satellites) is close to what the quality of final IGS orbits was back in 1992! In general the IGP orbits are an order of magnitude better than the broadcast orbits. The predicted orbits are of importance for time-critical applications (for meteorological community).

IGS products and quality: station coordinates

So far, station coordinates and the associated estimated velocities were sent by the individual Analysis Centers on an annual basis to the IERS. They contributed — together with the corresponding results from VLBI, SLR, DORIS, PRARE — to the realization of the current ITRF. The ITRF94 is the latest ITRF available. VLBI and SLR with about 120 sites each, GPS with about 70 sites, and DORIS with about 60 sites were the main contributors. The ITRF96 being prepared right now!

In view of the growing number of permanent regional GPS networks (Australia, Canada, Europe, Japan, USA) the systematic densification of the ITRF using GPS becomes more and more important.

IGS products and quality: densification of the ITRF

The IGS strategy was defined at the Workshop Densification of the IERS Terrestrial Reference Frame through Regional GPS Analyses in December 1994 at JPL in Pasadena. It was decided to proceed along the following lines:

— The densified IGS network shall consist of about 200 globally well distributed sites.

— The densification shall be realized in cooperation with other groups setting up permanent GPS networks.

— Regional analyses are performed by (new) IGS Associate Analysis Centers. These make extensive use of the global products (orbits, ERPs, ITRF coordinates and velocities of sites well established in the Global IGS Network). These AACs produce free network solutions on a weekly basis.

— The AACs make available their solutions to specialized Associated Analysis Centers which compare and combine the individual solutions.

The pilot phase of the IGS Densification Project started in September 1995 and will last till the end of 1997.

The Associate Analysis Centers combining the solutions are

— JPL,

— MIT and

— University of Newcastle.

Initially only the seven IGS Analysis Centers were making available their free network solutions on a weekly basis. Several Regional Analyses Centers contribute solution since 1996. A special format, the SINEX format was created to make the solutions and the associated covariance information available. The Pilot Phase should be terminated by the end of 1997. The Governing Board will decide on the future of the activities.

The following table shows (in abbreviated form) an extraction from the MIT report:

Transformation of Loose solutions to COMB

	X-tran (mm)	Y-tran (mm)	Z-tran (mm)	Scale (ppb)	rms (mm)
EMR	1.8	111.5	–108.2	–1.920	4.3
JPL	–8.3	–1.0	–45.1	–2.569	2.6
GFZ	–12.6	–13.0	–61.9	–2.701	4.6
COD	–16.9	9.2	–20.5	–2.797	3.5
SIO	–20.2	6.5	–72.2	–1.330	2.1
NGS	4.1	214.8	–148.0	–1.962	8.3
ESA	3.1	–45.0	–144.4	–2.663	4.1

(Mailed from: Tom Herring <tah@mtglas.mit.edu>)

The report reflects the results of GPS Week 898. The reports of all three centers (JPL MIT, Newcastle) indicate a consistency between the individual coordinate series of well below the cm for the horizontal, of about 1 cm for the vertical components.

Today the following institutions are contributing solutions as RNAACs on a weekly basis:

AUS: Australian Surveying and Land Information Group (AUSLIG)
EUR: EUREF-Solution (Euref Subcommission of IAG) with many contributors (following page)
GIA: Geophysical Institute, University of Alaska, Fairbanks
GSI: Geographical Survey Institute, Japan
PGC: Natural Resources Canada, Pacific Geoscience Centre, Canada
SIR: SIRGAS Solution prepared by the Deutsches Geodätisches Forschungsinstitut, Abt. I (DGFI/I).

EUR: EUREF-Solution Contributors:

ASI: Nuova Telespazio S.p.A., Space Geodesy Center
BEK: International Commission for Global Geodesy of the Bavarian Academy of Sciences (BEK)
COE: European solution created at CODE (Center for Orbit Determination in Europe)
GOP: Geodetic Observatory Pecny, Czech Republic
IFG: Institute for Applied Geodesy in Germany (IfAG)
LTP: Bundesamt für Landestopographie (L+T), Switzerland
NKG: Nordic Geodetic Commission (NKG)
OLG: Observatory Lustbühel Graz (OLG)
ROB: Royal Observatory of Belgium (ROB)
WUT: Warsaw University of Technology (WUT).

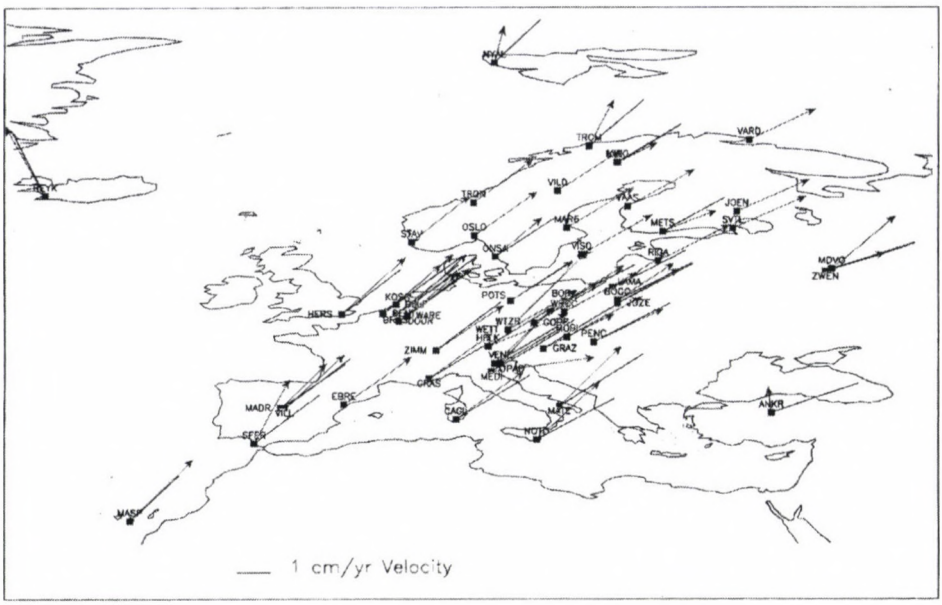


Fig. 7. Estimation of velocities of EUREF sites based on data from GPS week 834 (first week of year 1996) to GPS week 898 (end of March 1997)

IGS products and quality: densification of the ITRF EUREF solution

IGS products and quality: Earth orientation

An uninterrupted series of Earth orientation parameters is available starting with 21 June 1992. A combined IGS pole is available since 28 May 1995! With the increasing number stations and with growing experience the quality of EOP-estimates grew from about 1–2 mas in 1992 to 0.1–0.2 mas in 1998.

LOD (or UT1-UTC-drift values) are available with a one-day resolution from IGS Analysis Centers. LOD values corresponding to combined series are available as well!

Exploiting the GPS/IGS for fundamental astronomy

At the CODE processing center the attempt is made to exploit the IGS network for the following purposes:

— Not only the UT1-UTC-drift but also the drifts in the nutation in longitude $\dot{\Delta}\Psi \sin\epsilon$ and in obliquity $\dot{\Delta}\epsilon$ are estimated. (If you are solving for length-of-day there is no reason not to solve for these nutation drifts!)

— The polar wobble components x and y and UT1–UTC-drift are estimated with a high time resolution (2 hours). It is thus possible to extract ocean

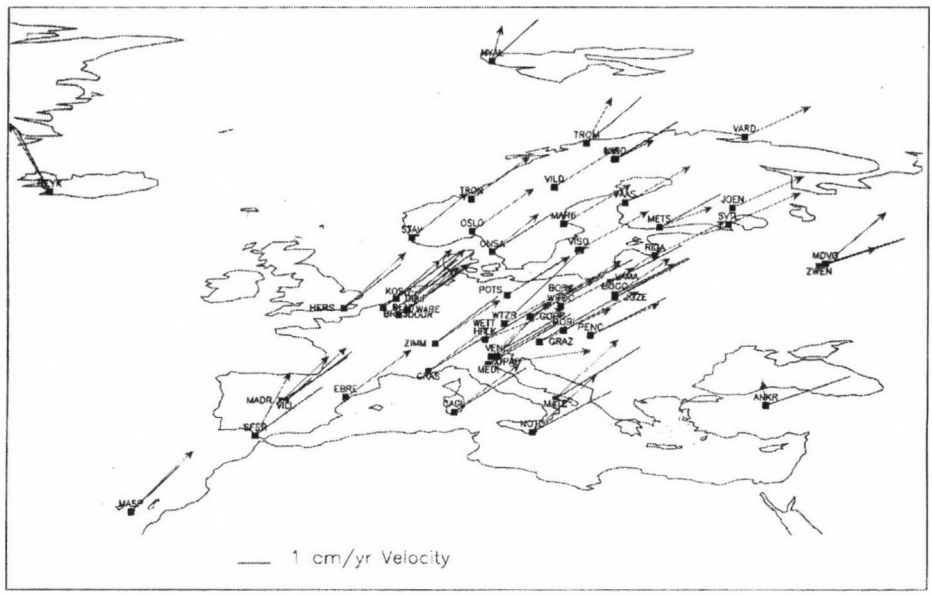


Fig. 8. Estimation of velocities of EUREF sites based on data from GPS week 834 (first week of year 1996) to GPS week 898 (end of March 1997). Mean velocity for Europe subtracted

tide models from these series. More precisely: models for tide-driven oceanic angular momenta.

So far Markus Rothacher (AIUB), Robert Weber (TU Vienna), Jan Hefty (TU Bratislava), Tim Springer (AIUB) and Tom Herring (MIT) were analysing these series. All IGS Analysis Centers have to model tropospheric refraction in order to obtain good results. Different techniques are used. Some Analysis Centers use stochastic models (based on Kalman filters), others just solve for station- and time-specific troposphere parameters.

Gendt and Beutler (1996) showed that the consistency of estimates stemming from different IGSS Analysis Centers is relatively high. It is thus posible to extract on a routine basis the precipitable water content for the entire IGS network with a high temporal resolution (two hours or finer) – provided high accuracy barometers are deployed in the IGS network. The issue was decided at the 1998 AC Workshop in Darmstadt to come up with an official IGS troposphere product (troposphere zenith delays).

The ionosphere is a dispersive medium:

$$dr_I(\nu) = \pm\alpha \cdot \frac{E}{\nu^2} \,. \tag{2}$$

The difference between the L1 and L2 GPS observations may thus be used for ionosphere modeling.

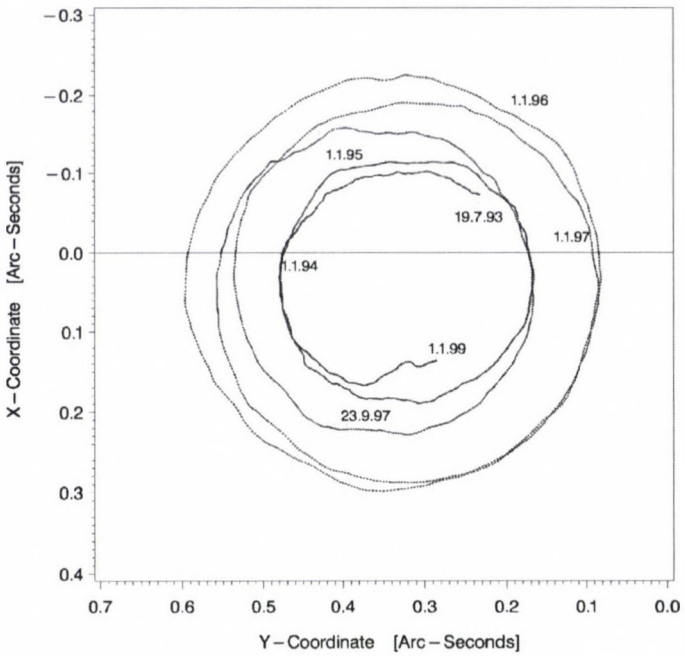


Fig. 9. Polar motion (19 July 1993 – 13 February 1998), as estimated by CODE

The motivation to produce ionosphere models is manyfold:

— local or regional models may be used to remove (reduce) biases in single frequency surveys,

— regional or global models may be used to calibrate altimeter measurements (ERS-1 is an example),

— there are pure research projects (ionosphere maps, extraction of geomagnetic indices).

CODE, ESA, JPL plus University of New Brunswick (UNB) and DLR Neustrelitz started producing ionosphere models. The example (Fig. 19) shows a single layer model produced by CODE, where the density within the shell is computed as series of harmonical functions. A common format, IONEX, was accepted.

New IGS products: antenna calibration tables

Zenith-dependent phase center variations exist for each antenna type, small variations even exist between antennas of the same type. Although the Dorne-Margolin antenna is dominant in the IGS network, there is a considerable number of Trimble-, Ashtech-antennas in the network.

It must be considered a constructive achievement that it was decided to produce tables of phase center variations relative to the Dorne-Margolin antenna (Rothacher

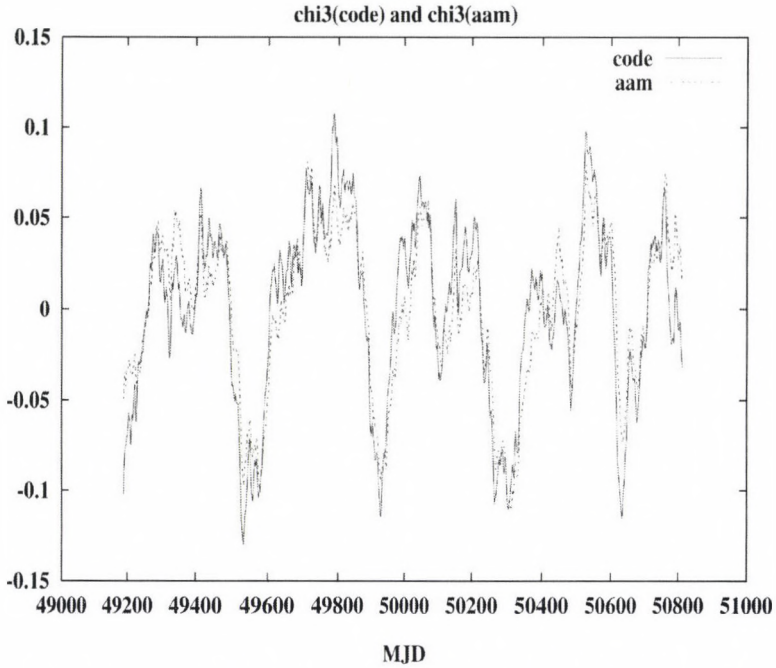


Fig. 10. Axial angular momentum calculated from CODE UT1-UTC-drift estimates and from IERS-AAM-Subbureau values (NCAR)

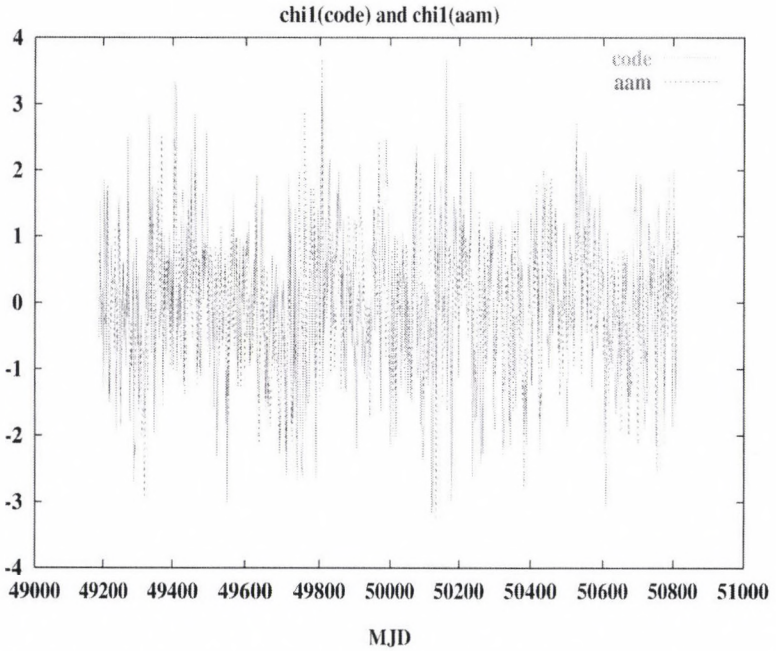


Fig. 11. First equatorial angular momentum calculated from CODE x-, and y- pole estimates and from IERS-AAM-Subbureau values (NCAR). Smoothing applied over three days

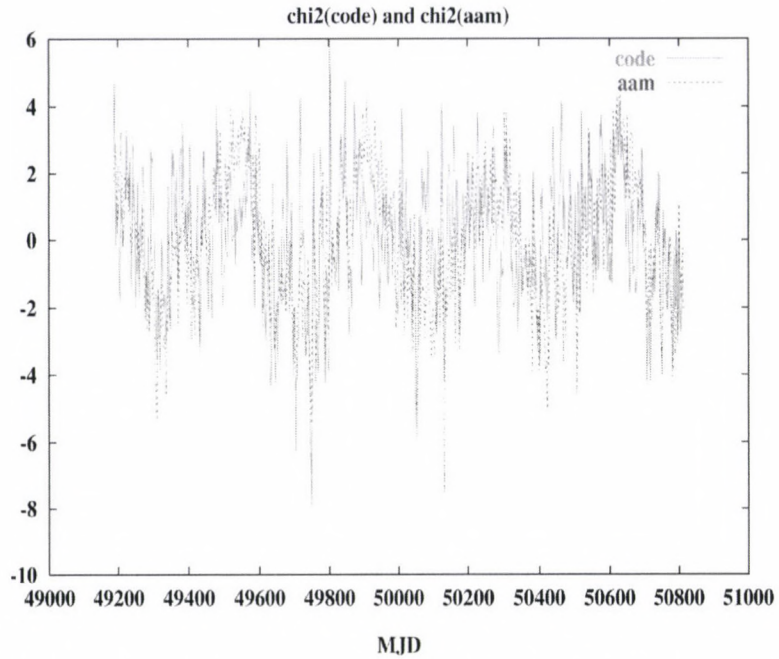


Fig. 12. First equatorial angular momentum calculated from CODE x-, and y- pole estimates and from IERS-AAM-Subbureau values (NCAR). Smoothing applied over three days

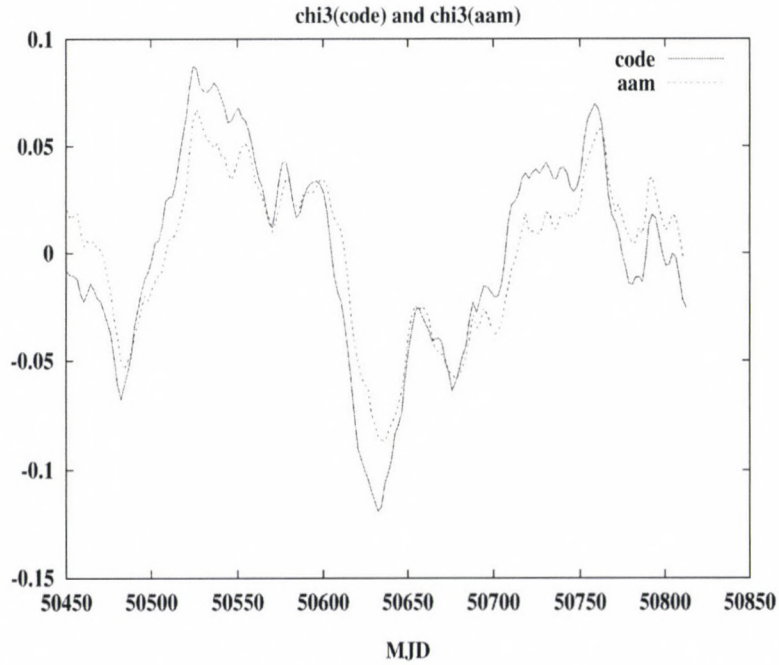


Fig. 13. Axial angular momentum calculated from CODE UT1-UTC-drift estimates and from IERS-AAM-Subbureau values (NCAR). Results for year 1997. AAM-series smoothed over one day

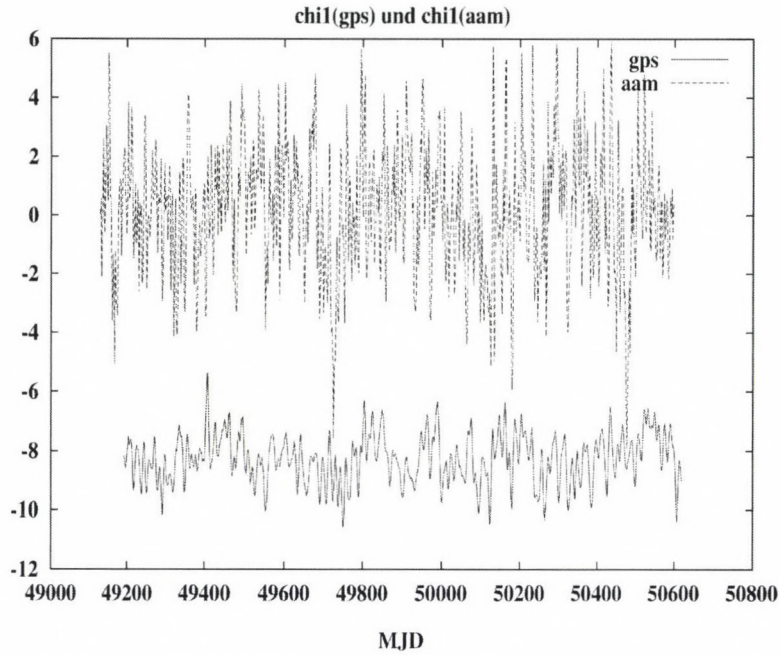


Fig. 14. First equatorial angular momentum calculated from CODE x-, and y- pole estimates and from IERS-AAM-Subbureau values (NCAR). Results for 1997. Smoothing applied over one day

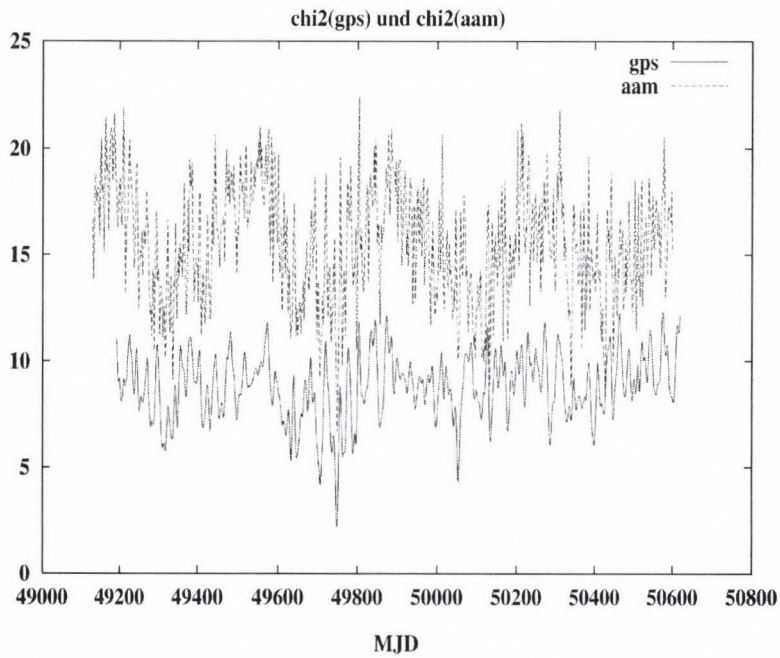


Fig. 15. Second equatorial angular momentum calculated from CODE x-, and y- pole estimates and from IERS-AAM-Subbureau values (NCAR). Results for 1997. Smoothing applied over one day

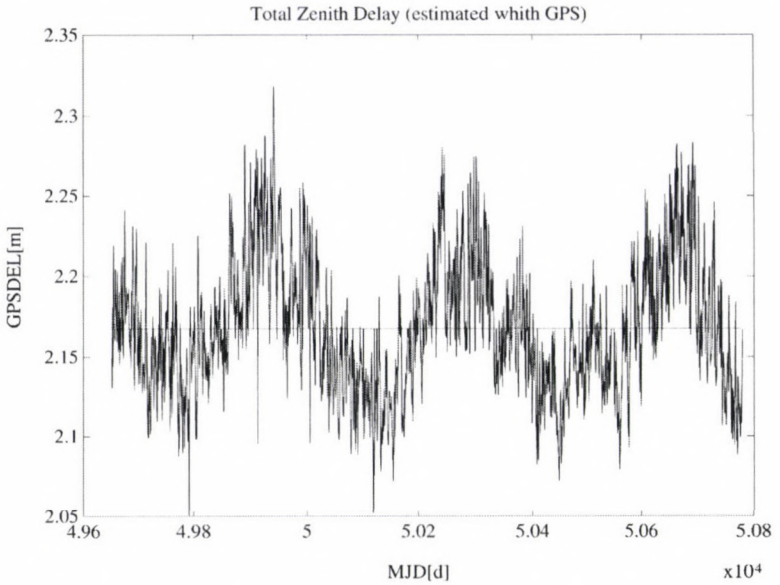


Fig. 16. GPS estimates from CODE (6 hours binning) for tropospheric zenith delays for Zimmerwald, October 1994 to November 1997

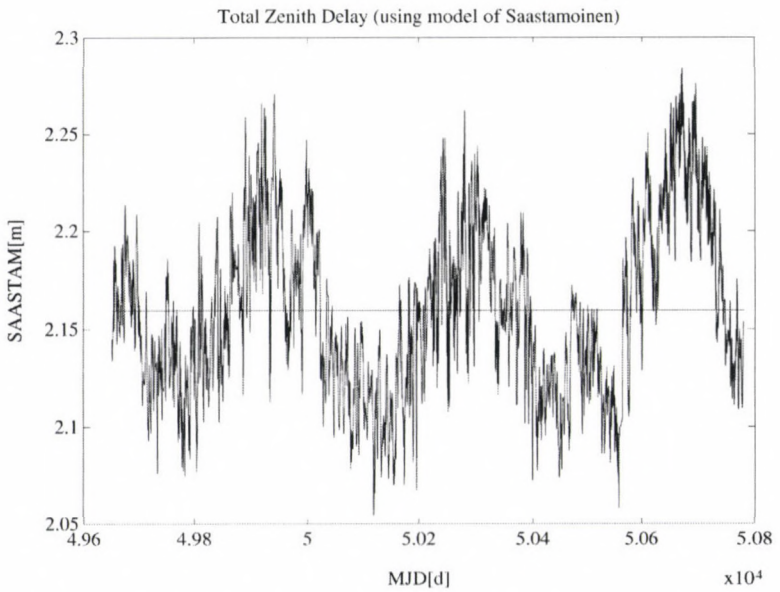


Fig. 17. Tropospheric zenith delays for Zimmerwald using surface met measurements for October 1994 to November 1997

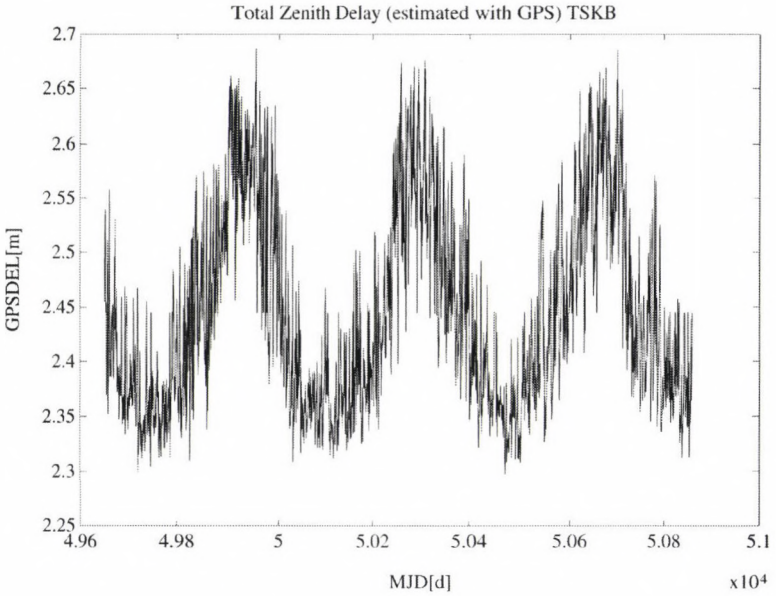


Fig. 18. GPS estimates from CODE (6 hours binning) for tropospheric zenith delays for Tsukuba (Japan), October 1994 to February 1998. Observe that the annual signal is much more pronounced than for Zimmerwald!

and Mader 1996) and to use them within the IGS starting with GPS week 860. More information is available at the CBIS, directory igscb/station/general.

News from the 1998 Darmstadt Workshop and Governing Board meetings

A call of participation will be issued in February 1998 for the IGEX-98. IGEX-98, the International GLONASS Experiment by the end of 1998, is a three months GLONASS test campaign. CSTG is coordinating this effort (chair Pascal Willis).

The IGS adopts the ITRF96. The IGS realization of ITRF96 will consist of about 50 GPS stations. This should lead to much more consistent EOPs (rapid products in particular).

The IGS/BIPM Time transfer experiment is well underway. The proposals answering this call for participation are due in March 1998.

The recommendations from the 1997 IGS retreat are discussed and viewed positively by the Governing Board. Formal approval will take place in May 1998 in Boston.

It was decided to establich and maintain a Global IGS Ionosphere model.

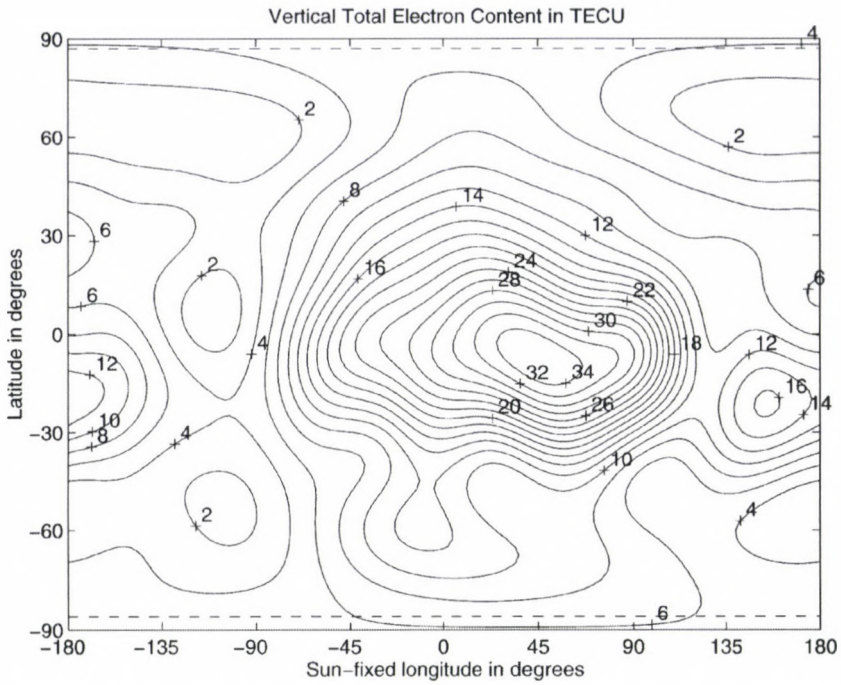


Fig. 19. Global ionosphere model for day 073 of Year 1996

Summary and conclusions

The IGS Orbits allow it to perform regional GPS analyses of highest accuracy without further orbit refinement. The results refer to the best available realization of the ITRF.

The IGS pole positions x and y are important contributions to the IERS ERP-series (Bulletins A and B), today.

The IGS significantly contributes to the realization and densification of the ITRF.

IGS solution series containing estimates of the time derivatives of nutation may significantly contribute to the establishment of nutation terms with periods < 40 days.

A re-processing starting from raw observations of the years 1992–1998 would lead to much more consistent ERP- and other series.

Important characteristics of the IGS:

— The IGS actually is an international and a multi-agency service.

— Although the IGS is a service with the goal to facilitate research, a considerable amount of research is taking place within the IGS. In this respect it is essential that there is a high redundancy in all IGS components (network, data centers, analysis centers).

— The IGS products (even those of the Analysis Centers) always were freely available to the scientific community. This led, e.g., to the detection of biases in the official IERS ERP-series.

— The IGS always was and is open to cooperate with other groups operating permanent networks.

— IGS standards for data formats and GPS station monumentation are observed by many other network operators today.

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 25–36 (2000)

GLOBAL GEODYNAMICS: BASIC CONSIDERATIONS

G BEUTLER[1]

Keywords: Earth models; Euler equations; Euler-Liouville equations; non-rigid Earth; three-body problem

Contents

The three body problem Earth-Sun-Moon

All considerations concerning Earth rotation are particular aspects of the *three body problem Earth-Sun-Moon.* At least Earth and Moon have to be modelled as *finite mass distributions.*

Assuming that the Earth is a rigid body it is possible to define an Earth-fixed coordinate system with the origin in the center of mass of the Earth and the axes coinciding with the principal axes of inertia (Figs 1 and 2).

The motion of *one* rigid body is completely known, if we know

— the motion of its center of mass in inertial space,

— the three Euler angles and the components of the angular velocity vector $\vec{\omega}$

as a function of time. Both problems are coupled, however. If they are artificially de-coupled, this is done through some approximations. That this may be done successfully is due to the fact that the coupling (taking place over the moments of inertia) is weak.

The first part of the problem is usually attributed to celestial mechanics, the latter to geodynamics.

[1]Astronomical Institute, University of Berne, Switzerland

1217-8977/2000/$ 5.00

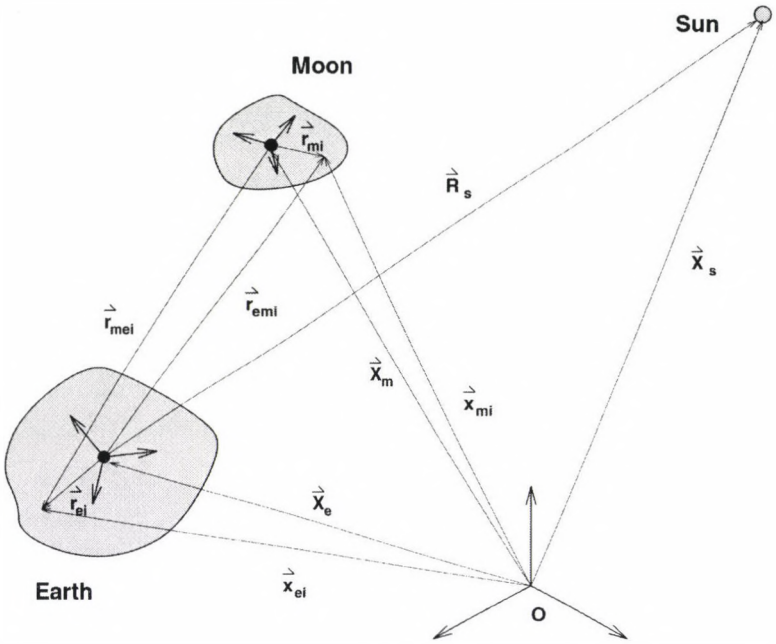


Fig. 1

Both problems are described by six scalar quantities, namely components of position and velocity vectors in the first case, components of angular velocity vector and the three Euler angles in the second case.

The general three body problem is thus described by 24 independent scalar quantities, namely

— the 2 × (3+3) components of the geocentric position and velocity for the centers of mass of Sun and Moon,

— the 3+3 components of angular velocities of Earth and Moon, and

— the 3+3 Euler angles describing the orientation (attitude) of the finite bodies Earth and Moon in inertial space.

Because all 24 quantities are functions of time we are looking for an ordinary differential equation system with the time as independent argument. The differential equations for these quantities may be derived by setting up the Newtonian equations of motion for each mass element of each celestial body in inertial space.

Let us give an example for mass particle m_i of the Earth:

$$\begin{aligned} m_{ei} \cdot \left(\ddot{\vec{X}}_e + \ddot{\vec{r}}_{ei} \right) \quad = \quad & -Gm_{ei}M_s \cdot \frac{\vec{x}_{ei} - \vec{X}_s}{|\vec{x}_{ei} - \vec{X}_s|^3} - \\ & -Gm_{ei} \cdot \sum_{k=1}^{n_m} m_{mk} \frac{\vec{x}_{ei} - \vec{x}_{mk}}{|\vec{x}_{ei} - \vec{x}_{mk}|^3} + \sum_{k=1, k \neq i}^{n_e} \vec{F}_{eik} \, . \end{aligned} \tag{1}$$

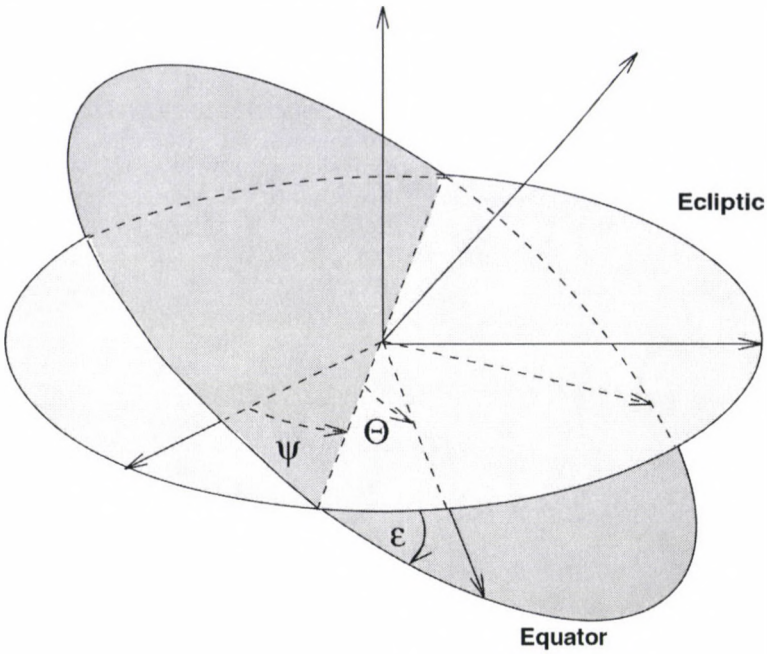


Fig. 2

The last term in the above equation symbolizes the internal forces acting on particle i. Adding all equations for all mass particles of the Earth leads in a very natural way to the equations of motion for the center of mass of the Earth.

Multiplying each of the above equations vectorially with $\vec{r}_i\times$ and adding all equations for all mass particles of the Earth leads in a natural way to the differential equation for the angular momentum of the Earth.

The Euler equations

The equation for the angular momentum reads as:

$$\begin{aligned} \sum_{i=1}^{n_e} m_{ei}\cdot\vec{r}_{ei}\times\ddot{\vec{r}}_{ei} &= GM_s\cdot\sum_{i=1}^{n_e} m_{ei}\cdot\frac{\vec{r}_{ei}\times\vec{R}_{es}}{|\vec{R}_{es}-\vec{r}_{ei}|^3}+ \\ &\quad +G\cdot\sum_{i=1}^{n_e} m_{ei}\cdot\sum_{k=1}^{n_m} m_{mk}\frac{\vec{r}_{ei}\times\vec{r}_{emk}}{|\vec{r}_{emk}-\vec{r}_{ei}|^3} \\ \dot{\vec{d}}_e &= \vec{l}_m+\vec{l}_s\,. \end{aligned} \tag{2}$$

where $\vec{l}_m$ and $\vec{l}_s$ are the torques excerted by Moon and Sun on the equatorial bulge of the Earth.

The sums may easily and approximately be replaced by the elements of the Earth's inertia tensor. The equation assumes a very simple and handy form when written in the body-fixed coordinate system.

The idea of transforming the equations for the angular momentum is attributed to Leonhard Euler (1707–1783). We use the tensor relation

$$\vec{d}_e = I \cdot \vec{\omega}\,, \tag{3}$$

where I is the inertia tensor of the Earth, and we know that the inertia tensor is particularly simple when expressed in the coordinate system defined by the principal axes of inertia:

$$I = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix} \approx \begin{pmatrix} A & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & C \end{pmatrix} \tag{4}$$

where A, B, C are the principal moments of inertia. Using $B = A$ means to assume rotational symmetry of the Earth, an assumption which is amply justified for our order of magnitude considerations.

Written in components for the case $A = B$ the Euler equations read as:

$$\begin{aligned} \begin{pmatrix} \dot{\omega}_{Fe,1} + \gamma \cdot \omega_{Fe,2}\omega_{Fe,3} \\ \dot{\omega}_{Fe,2} - \gamma \cdot \omega_{Fe,3}\omega_{Fe,1} \\ \dot{\omega}_{Fe,3} \end{pmatrix} &= 3 \cdot \tfrac{GM_s}{R_{es}^5} \cdot \begin{pmatrix} \gamma \cdot R_{Fes,3}R_{Fes,2} \\ -\gamma \cdot R_{Fes,1}R_{Fes,3} \\ 0 \end{pmatrix} + \\ &\quad + 3 \cdot \tfrac{GM_m}{R_{em}^5} \cdot \begin{pmatrix} \gamma \cdot R_{Fem,3}R_{Fem,2} \\ -\gamma \cdot R_{Fem,1}R_{Fem,3} \\ 0 \end{pmatrix} \end{aligned} \tag{5}$$

where

$$\gamma = \frac{C - A}{A} \approx \frac{1}{300}\,. \tag{6}$$

The third equation may easily be solved to become

$$\omega_{Fe,3} = \omega \tag{7}$$

Introducing this result into the Euler equation leads to a system of non-homogeneous linear differential equations of first order for the components $\omega_{Fe,1}$ and $\omega_{Fe,2}$ of the angular velocity vector $\vec{\omega}$ in the Earth-fixed system.

It is easily possible to express the components of ω as a function of the first time derivatives of the Euler angles:

$$\vec{\omega}_{Fe} = \begin{pmatrix} -\sin\epsilon_e \sin\Theta_e, & +\cos\Theta_e, & 0 \\ -\sin\epsilon_e \cos\Theta_e, & -\sin\Theta_e, & 0 \\ \cos\epsilon_e, & 0, & 1 \end{pmatrix} \cdot \begin{pmatrix} \dot{\Psi}_e \\ \dot{\epsilon}_e \\ \dot{\Theta}_e \end{pmatrix}. \tag{8}$$

Together with the preceeding Euler equations these equations completely describe the rotational motion of the rigid Earth. If we assume that the motions of Moon and Sun around the Earth are known, the Euler equations and the above equations for the Euler angles may be integrated independently on the corresponding equations for Sun and Moon. We will perform a few experiments using this technique.

Messing around with Earth rotation

We use a computer program to numerically integrate Euler's equations and the associated equations for the Euler angles. Using "reasonable values" for masses, moments of inertia, and orbital elements for Sun and Moon we want to study

— polar motion,

— the lumped effects of precession and nutation,

a) over a short time interval of a few months, b) over 20 years.

We look at the cases

— Sun and Moon "turned off",

— only Moon "turned on",

— only Sun "turned on",

— Sun and Moon "turned on".

The non-rigid Earth

Plate motions are of the order of a few centimeters per year. They under line that the Earth is not a rigid body. Earth surface is (almost) and equipotential surface in the combined gravitational and centrifugal potential. That is, for forces constant over millions of years, the Earth is an elastic body.

Due to the polar wobble the rotation axis slightly differs from the axis of maximum inertia. The polar tides try to transform the solid ellipsoids into the dotted ellipsoid (Fig. 4). The effect of polar tides is responsible for the difference between the Euler and the Chandler period. The deformation in Fig. 5 is, e.g., responsible for tidal variation of length of day.

The Euler-Liouville equations

If the Earth is not rigid, expressions like Earth-fixed become obsolete in the strict sense. It turns out that the equations for the angular momentum remain unchanged in the inertial system:

$$\dot{\vec{d}}_e = \vec{l}_m + \vec{l}_s \,. \tag{9}$$

It is not possible, of course, to transform into the Earth-fixed system, *but it is possible to refer the equations to a rigid, rotating coordinate system.* We may virtually take over all the developments made for the rigid Earth, *but* we have to take into account that deformations with respect to this rigid frame are possible. The Euler-Liouville equations read as:

$$\frac{d}{dt}\left(I_{Fe}\cdot\vec{\omega}_{Fe}+\vec{h}_{Fe}\right)+\vec{\omega}_e\times\left(I_{Fe}\cdot\vec{\omega}_{Fe}+\vec{h}_{Fe}\right)=\vec{l}_{Fe}\,,$$

where $\vec{h}_{Fe}$ stands for the angular momentum due to deformation.

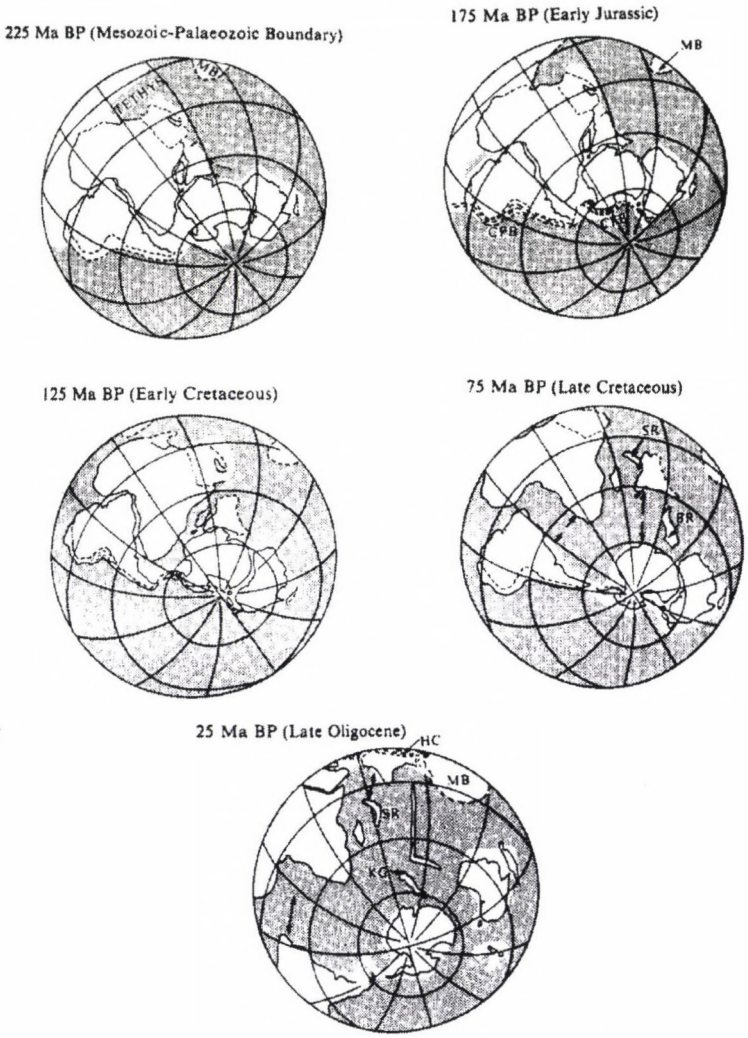


Fig. 3. Development of the Southern Continent Gondwana (from Lambeck 1988)

If we follow Felix Tisserand's (1845–1896) idea to *define the rotating system in such a way that the angular momentum due to deformations vanishes* we obtain the Euler-Liouville equations in the following form

$$\frac{d}{dt}\left(I_{Fe}\cdot\vec{\omega}_{Fe}\right)+\vec{\omega}_e\times\left(I_{Fe}\cdot\vec{\omega}_{Fe}\right)=\vec{l}_{Fe}\,. \tag{10}$$

As a matter of fact the Euler equations which we wrote explicitely for the rigid body may be written in exactly the above form. We "only" have to take into account that now the inertia tensor is no longer diagonal and that its elements are varying with time!

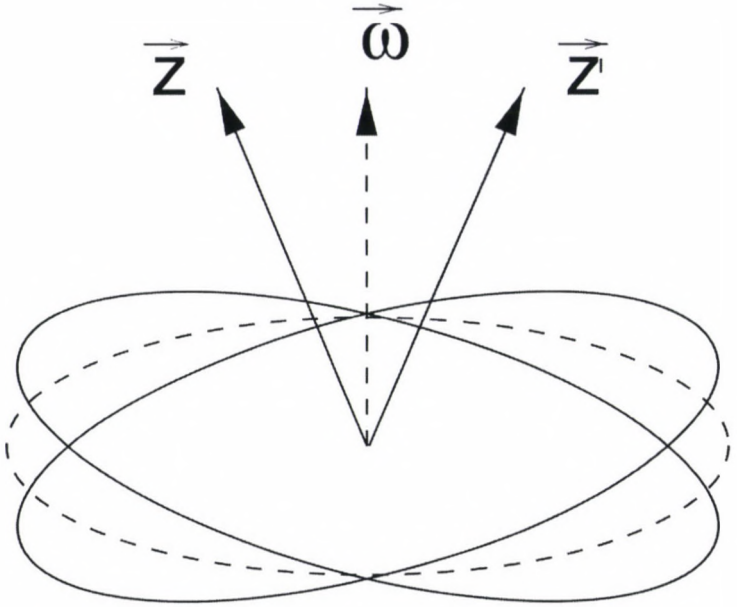


Fig. 4

As opposed to the case of the rigid body the deformable Earth must produce, e.g., length-of-day variations.

Explicitely, the equations of motion for an elastic Earth with rotation symmetry may be written in the form.

$$\begin{aligned}
\dot{\omega}_{Fe,1} + \gamma\omega_{e0}\left(1 - \tfrac{k}{k_s}\right)\cdot\omega_{Fe,2} &= -\frac{\delta\dot{I}_{gez,Fe,13}}{A_0}\cdot\omega_{e0} + \frac{\delta I_{gez,Fe,23}}{A_0}\cdot\omega_{e0}^2 + l_{Fe,1} \\
\dot{\omega}_{Fe,2} - \gamma\omega_{e0}\left(1 - \tfrac{k}{k_s}\right)\cdot\omega_{Fe,1} &= -\frac{\delta\dot{I}_{gez,Fe,23}}{A_0}\cdot\omega_{e0} - \frac{\delta I_{gez,Fe,13}}{A_0}\cdot\omega_{e0}^2 + l_{Fe,2} \\
\dot{\omega}_{Fe,3} &= -\frac{\delta\dot{I}_{gez,Fe,33}}{C_0}\cdot\omega_{e0}
\end{aligned}$$

where $k \approx 0.3$ is the Love constant, $k_s \approx 1$ the secular Love constant. The third equation may be separated from the first two equations. The above differential equations are linear, the coefficients are constant. The free motion is ruled by:

$$\begin{aligned}
\dot{\omega}_{Fe,1} + \gamma\omega_{e0}\left(1 - \tfrac{k}{k_s}\right)\cdot\omega_{Fe,2} &= 0 \\
\dot{\omega}_{Fe,2} - \gamma\omega_{e0}\left(1 - \tfrac{k}{k_s}\right)\cdot\omega_{Fe,1} &= 0 \\
\dot{\omega}_{Fe,3} &= 0\,.
\end{aligned}$$

The Chandler period is given by

$$P = \frac{2\cdot\pi}{\omega_{e0}\cdot\left(1 - \tfrac{k}{k_s}\right)} \approx 430 \text{ days}$$

for $k \approx 0.3$.

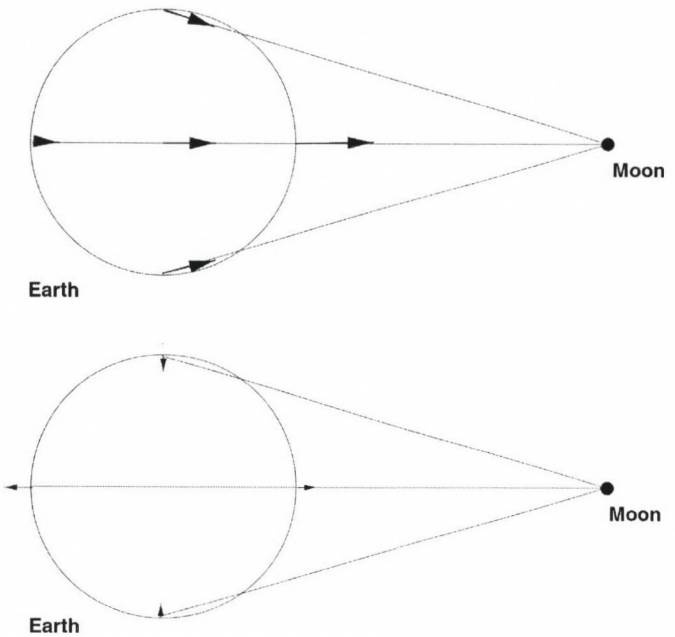


Fig. 5. Deformation of the Earth due to tidal attraction of the Earth; top: forces in inertial system, bottom: forces in geocentric system

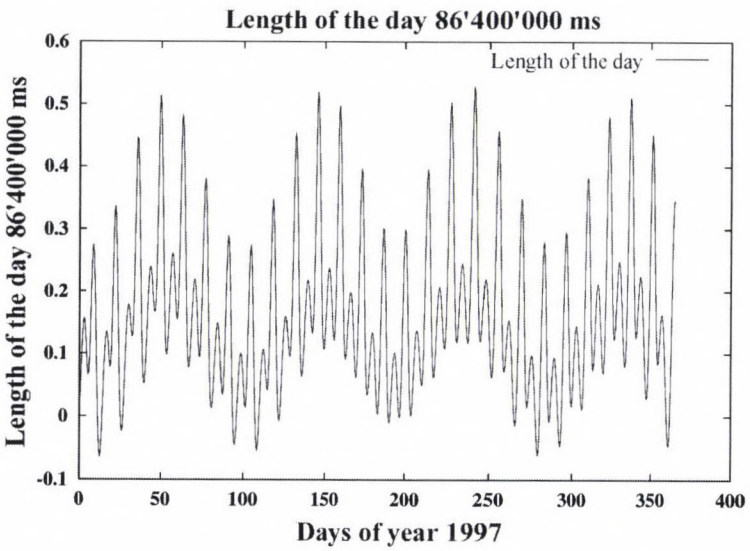


Fig. 6. Length of day variation due to lunar and solar tides

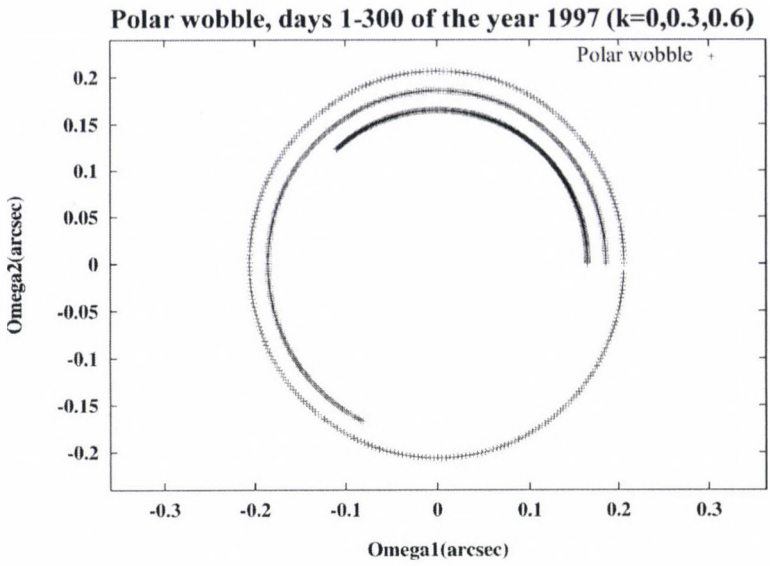


Fig. 7. Polar wobble for different values of Love constant k

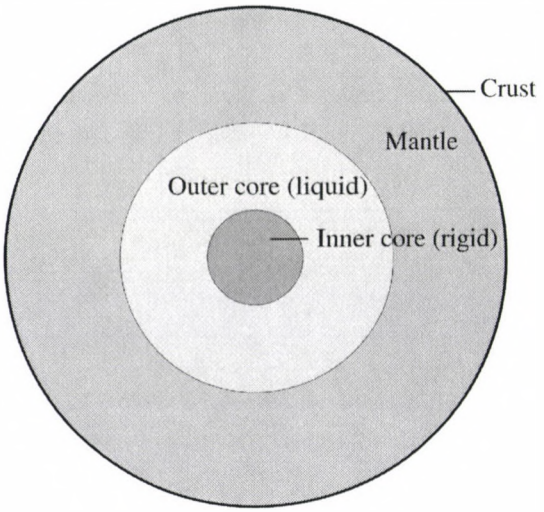


Fig. 8

Composite Earth models

From seismic and other terrestrial measurements we know that neither a rigid, nor an elastic Earth are good approximations for the actual Earth. In Fig. 3 we make the distinction between a rigid inner and a liquid outer core, the mantle and the crust.

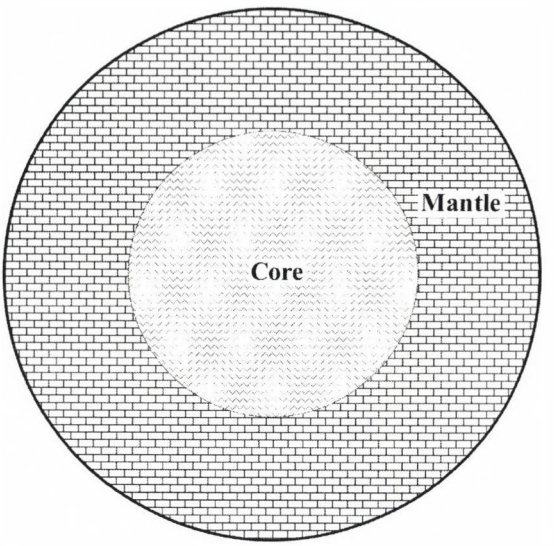


Fig. 9

The Poincaré Earth models

In the Poincaré model we make the distinction between a rigid mantle plus crust and a liquid core (Fig. 9). There may be a net motion "rotation" of the liquid core relative to the mantle. We have to set up equations for the core and for the mantle.

Equations of motion for Poincaré model:

$$
\begin{array}{lll}
I_{e1}\dot{\omega}_{Fe1} + \tilde{I}_{c1}\dot{\chi}_{F1} + \gamma_{e1}I_{e1}\omega_{Fe2}\omega_{Fe3} + \tilde{I}_{c3}\omega_{Fe2}\chi_{F3} - \tilde{I}_{c2}\omega_{Fe3}\chi_{F2} & = & l_{Fe1} \\
I_{e2}\dot{\omega}_{Fe2} + \tilde{I}_{c2}\dot{\chi}_{F2} + \gamma_{e2}I_{e2}\omega_{Fe1}\omega_{Fe3} + \tilde{I}_{c1}\omega_{Fe3}\chi_{F1} - \tilde{I}_{c3}\omega_{Fe1}\chi_{F3} & = & l_{Fe2} \\
I_{e3}\dot{\omega}_{Fe3} + \tilde{I}_{c3}\dot{\chi}_{F3} + \gamma_{e3}I_{e3}\omega_{Fe1}\omega_{Fe2} + \tilde{I}_{c2}\omega_{Fe1}\chi_{F2} - \tilde{I}_{c1}\omega_{Fe2}\chi_{F1} & = & l_{Fe3} \\
\hline
\tilde{I}_{c1}\dot{\omega}_{Fe1} + I_{c1}\dot{\chi}_{F1} + \tilde{I}_{c2}\omega_{Fe2}\chi_{F3} - \tilde{I}_{c3}\omega_{Fe3}\chi_{F2} - I_{c1}\gamma_{c1}\chi_{F2}\chi_{F3} & = & 0 \\
\tilde{I}_{c2}\dot{\omega}_{Fe1} + I_{c1}\dot{\chi}_{F2} + \tilde{I}_{c3}\omega_{Fe3}\chi_{F1} - \tilde{I}_{c1}\omega_{Fe1}\chi_{F3} - I_{c2}\gamma_{c2}\chi_{F3}\chi_{F2} & = & 0 \\
\tilde{I}_{c3}\dot{\omega}_{Fe3} + I_{c3}\dot{\chi}_{F3} + \tilde{I}_{c1}\omega_{Fe1}\chi_{F2} - \tilde{I}_{c2}\omega_{Fe2}\chi_{F1} - I_{c3}\gamma_{c3}\chi_{F1}\chi_{F2} & = & 0 \\
\hline
-\sin\epsilon_e \sin\Theta_e\dot{\Psi}_e + \cos\Theta_e\dot{\epsilon}_e & = & \omega_{Fe1} \\
-\sin\epsilon_e \cos\Theta_e\dot{\Psi}_e - \sin\Theta_e\dot{\epsilon}_e & = & \omega_{Fe2} \\
\cos\epsilon_e\dot{\Psi}_e + \qquad\qquad \dot{\Theta}_e & = & \omega_{Fe3}
\end{array}
$$

Moments of inertia with index c refer to the core, those with index e refer to the entire Earth. $\chi_{\ldots}$ define the motion of the core relative to mantle. Approximately, the $\chi_{\ldots}$ are the components of the angular velocity of the core relative to the mantle.

The first three equations are generalized Euler equations (for $\chi_{\ldots} = 0$ we obtain the Euler equations), the last three equations are the equations defining the motion of the rotation axis of the mantle in space.

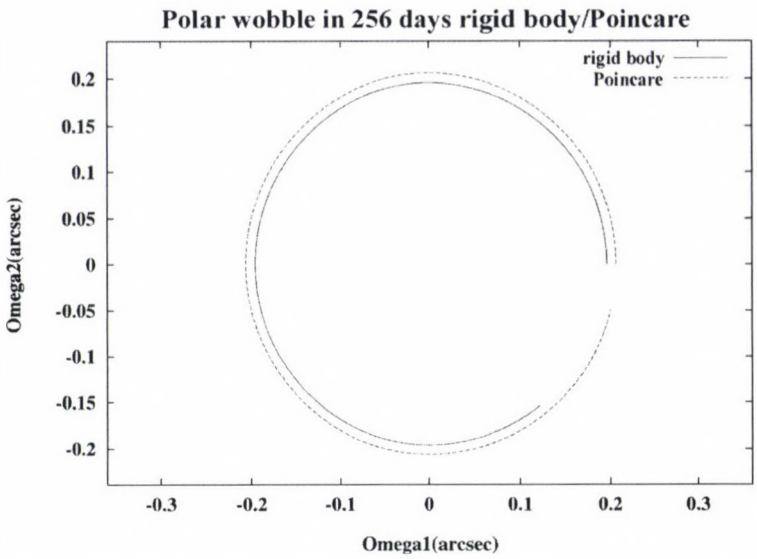


Fig. 10. Free motion for Poincaré and rigid Earth model with spherical core-mantle boundary

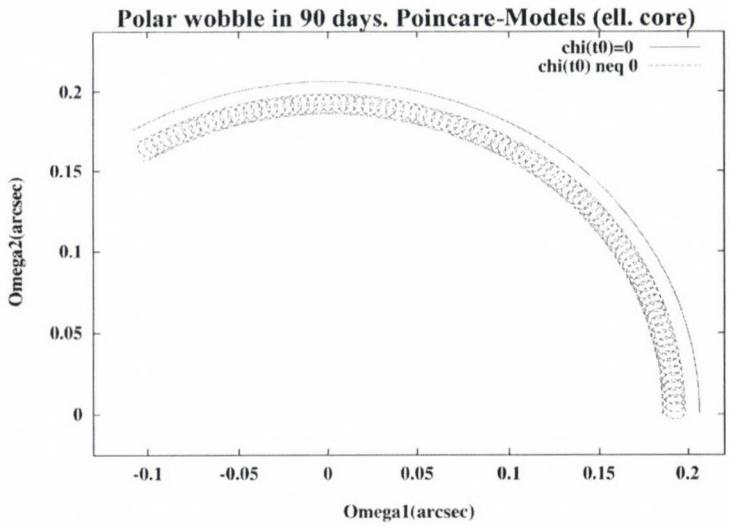


Fig. 11. Free motion for Poincaré and rigid Earth model, elliptical core mantle boundary, 20" offset between mantle and core rotation axes

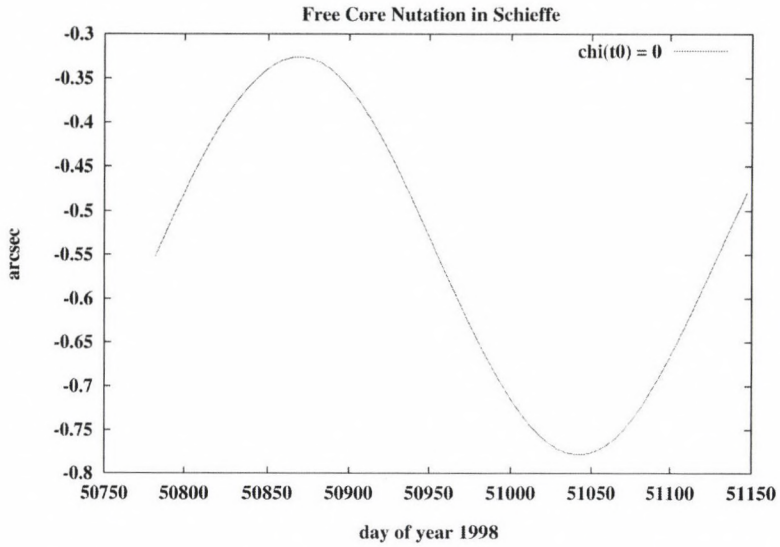


Fig. 12. Free motion in inertial system for the Poincaré model, elliptical core mantle boundary, 2" offset between mantle and core rotation axes

The resulting motion is called free core nutation (FCN). It has a period of

$$P_{I,fcn} = \frac{I_m}{I_e} \cdot \frac{2\pi}{\gamma_c \omega_0} \approx 350 \text{ days}$$

and an amplitude of

$$a_{I,fcn} = \frac{I_c}{I_e} \cdot a_{F,fcn} \approx 0.11 \cdot a_{F,fcn} \,,$$

where $a_{F,fcn}$ is the amplitude of the free core nutation χ of the core relative to the mantle.

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 37–48 (2000)

THERMOVISCOELASTIC MODELS OF THE DEFORMATIONS AND GRAVITY CHANGES DUE TO ANOMALOUS SOURCE OF HEAT

L BRIMICH[1]

The paper presents basic formulae for the stress and strain components of the deformation and gravity changes due to source of heat of prismatic shape embedded in the elastic and viscoelastic halfspace (lithosphere) of the Kelvin type. Numerical results are presented in the numerous graphs which show gradual approaching stress, strain and gravity changes to the steady state. It is shown that the surface displacement causes a pronounced dome in the epicentral region of the source of heat and that the surface thermoviscoelastic stresses are mainly of tensile type and their values are well comparable with critical stresses required to cause creep or fractures of the surface of the halfspace.

Keywords: deformation; gravity change; heat source; thermoelastic model

1. Introduction

As is known, a magmatic intrusion in the Earth's crust will cause a series of effects related to its mass as well to the pressurization of the chamber due to overfilling or temperature changes. In deformation modelling and prediction the most interesting effects are those which in principle could be detected on the surface before the eruption, including surface gravity changes and deformations.

There are more important goals of modern crustal deformation studies. One of these goals is to understand the transient post event ground deformation sometimes seen following large earthquakes and dike emplacement events. Hofton et al. (1995) studied the horizontal surface deformation due to dike emplacement in an elastic-gravitational layer overlying a viscoelastic-gravitational halfspace.

For understanding crustal deformation in volcanic regions there are two fundamental and equally important aspects: theoretical modeling of the effects caused by the natural phenomenon under study (in this case a magmatic intrusion in the crust) and the practical aspect of continuous observation and appropriate processing of the data obtained.

The literature on deformation modelling is extensive, and most of it does not consider the existence of the ambient gravitational field, the crust being represented by a homogeneous, purely elastic halfspace. It is impossible and would be tedious to present a complete list of references in this field. Pressure changes within the magma chamber produce deformation and gravity changes at the surface of the Earth. Modelling of these deformations and gravity changes was discussed by Rundle (1980, 1982, 1983), Fernández (1992), Fernández and Rundle (1994) and many other authors.

[1]Geophysical Institute Slovak Academy of Sciences, Dúbravská cesta 9, 842 28 Bratislava, Slovak Republic, e-mail: geofbrim@savba.sk

1217-8977/2000/$ 5.00

The problem of viscoelastic-gravitational displacement due to rectangular dipping thrust fault in a layered Earth model was solved in Fernández et al. (1996).

In the Geophysical Institute of the Slovak Academy of Sciences Hvoždara and Rosa (1980) started with the model of the thermoelastic deformation field due to a point source of heat. Theoretical models for gravity anomalies caused by thermoelastic deformations in the vicinity of magmatic bodies was discussed by Hvoždara and Brimich (1991, 1995).

Hvoždara (1992) considered the model of the viscoelastic halfspace with the point source of heat. But point source models are not acceptable if we want to study the deformation in the very proximity of the source. Dietrich and Decker (1995) worked out finite-element models representing finite-size magma chambers in an elastic halfspace. A series of elastic models with increasing complexity has also been proposed, accounting for various geometries of the source (Davis 1986) and various spatial distributions of elastic properties (Bianchi et al. 1987).

Even after so many different deformation models have been developed, there are still discrepancies in data interpretation, in particular for gravity data. This fact, together with the interest and need to further knowledge of all aspects of volcanic phenomena, mean that more complex calculations, which include effects not present in other models, or not present together (e.g. consideration of gravitational potential, layering, non-point sources, visco-elastic and inelastic properties of the medium, etc.) are required.

In this paper we present the thermo-viscoelastic deformation field due to the source of heat of prismatic shape embedded in the viscoelastic halfspace (lithosphere) of Kelvin's type.

The formulae for the gravity perturbation due to the volume dilatation connected with thermo-viscoelastic deformations are derived as well. Numerical results are presented in the graphs which show stress and strain gradually approaching to the steady state.

In the following sections we discuss the technique involved in synthesising displacements, deformation, potential and gravity changes due to volcanic loading considered the described medium. Results of the calculations are then discussed for the aspects mentioned above.

2. Point source of heat in a viscoelastic halfspace

The time evolution of heating of the halfspace (lithosphere) and deformation associated with it can be mathematically calculated by means of the theory of thermoviscoelastic deformation. If we consider the non-steady point source of heat to be located in the depth ζ of the viscoelastic halfspace $z > 0$ using Nowacki's monograph (1962) we can write fundamental equations for the uncoupled thermoviscoelastic problem for the point source of heat buried in the depth ζ. The temperature disturbance field $T(x, y, z, t)$ due to this source must obey the equation:

$$\lambda_T \nabla^2 T + w\delta(x)\delta(y)\delta(z - \zeta)H(t) = c_p \rho \frac{\partial T}{\partial t} \tag{1}$$

where λ_T is heat conductivity, c_p is specific heat under the constant pressure, ρ is the material density, w is the power of the heat source, δ is the Dirac function, $H(t)$ is Heaviside's unit step function $H(t) = 0$ for $t < 0$, $H(t) = 1$ for $t > 0$ and ζ is the depth of the point source of heat.

If the surface of the halfspace is kept at a constant temperature which can be taken to be zero then on the surface $z = 0$ we have the following boundary condition:

$$T(x, y, z, t)|_{z=0} = 0. \tag{2}$$

The initial temperature disturbance in all points of the halfspace is zero then for $t = 0$ we have initial condition:

$$T(x, y, z, t)|_{t=0} = 0. \tag{3}$$

Then the solution of Eq. (1) is known (Carslaw and Jaeger 1959) in the form:

$$T(r, z, t) = \frac{w}{4\pi\lambda_T} \left\{ R_1^{-1} \mathrm{erfc}\left(\frac{R_1}{\sqrt{4\kappa t}}\right) - R_2^{-1} \mathrm{erfc}\left(\frac{R_2}{\sqrt{4\kappa t}}\right) \right\}, \tag{4}$$

where $R_1 = \left[r^2 + (z - \zeta)^2\right]^{1/2}$, $R_2 = \left[r^2 + (z + \zeta)^2\right]^{1/2}$ being the horizontal distance from the polar axis z and $\kappa = \lambda_T/(c_p\rho)$. The complementary error function erfc (s) is defined as follows:

$$\mathrm{erfc}(s) = 1 - \frac{2}{\sqrt{\pi}} \int_0^s e^{-u^2} \mathrm{d}u. \tag{5}$$

The time and space variable temperature disturbance (3) causes in the viscoelastic halfspace time-space variable stresses and displacements. Since the process of temperature changes is much slower in comparison with propagation time of elastic waves, it is sufficient to consider the static equilibrium equation for a viscoelastic body:

$$\sum_{j=1}^{3} \frac{\partial \sigma_{ij}}{\partial x_j} = 0 \qquad\qquad i = 1, 2, 3, \tag{6}$$

where σ_{ij} are the components of the viscoelastic stress tensor.

In order to obtain the actual time behaviour of the displacements and stresses we have to calculate Laplace transform of them. The detailed calculation was performed in Hvoždara (1992).

The calculation was performed for the Kelvin body, for which the generalized Duhamel-Neumann relation is of the form:

$$\begin{aligned} \sigma_{ij}(x_r, t) \quad = \quad & 2\mu \left(1 + t^* \frac{\partial}{\partial t}\right) e_{ij}(x_r, t) + \\ & + \delta_{ij} \left\{ \frac{1}{3} \left[3K - 2\mu \left(1 + t^* \frac{\partial}{\partial t}\right) \Theta(x_r, t) \right] - 3K\alpha_t T(x_r, t) \right\}, \end{aligned} \tag{7}$$

where μ is the modulus of rigidity (Lamé constant), $K = \lambda + \frac{2}{3}\mu$ bulk modulus, $t^* = \eta/\mu$ is decay time, η the viscosity of material, α_T being the thermal coefficient of the linear expansion and $\Theta(x_r, t)$ the dilatation.

The time dependence for displacements and stresses on the surface of the viscoelastic halfspace we have the following formulae:

$$u_r(r,0,t) = \frac{Qr}{\pi}\int_0^t V(t-\tau)S_1(r,\tau)\mathrm{d}\tau,$$

$$\begin{aligned} u_z(r,0,t) = \; & -Q2\pi\Big\{\zeta R_0^{-3}b(t)+ \\ & +\frac{2\zeta}{\sqrt{\pi}}\int_0^t b(t-\tau)\tau^{-1}(4\kappa\tau)^{-3/2}\exp\left[\frac{-R_0^2}{4\kappa\tau}\right]\mathrm{d}\tau- \\ & -\int_0^t W(t-\tau)S_2(r,\tau)\mathrm{d}\tau\Big\}, \end{aligned} \tag{8}$$

$$\sigma_{rr}(r,0,t) = \frac{2Q}{\pi}\left\{\int_0^t B(t-\tau)S_0(r,\tau)\mathrm{d}\tau - \int_0^t U(t-\tau)S_1(r,\tau)\mathrm{d}\tau\right\},$$

$$\sigma_{\varphi\varphi}(r,0,t) = \frac{2Q}{\pi}\left\{\int_0^t N(t-\tau)S_0(r,\tau)\mathrm{d}\tau + \int_0^t U(t-\tau)S_1(r,\tau)\mathrm{d}\tau\right\}.$$

The stresses σ_{zz} and σ_{rz} are zero according to known boundary condition of the theory of elasticity. Explanation of the symbols in Eqs (8) is following:

$$Q = \frac{w\kappa}{\lambda_T}, \quad R_0 = \left[r^2+\zeta^2\right]^{1/2},$$

$$U(t) = B(t) - N(t), \quad V(t) = b(t) - M(t), \quad W(t) = b(t) - 2M(t),$$

$$B(t) = a_1[t - \beta_1\kappa_1^{-1}(1-e^{\kappa_1 t})],$$

$$N(t) = a_1a_3\kappa_3^{-1}[t - (\kappa_3^{-1}+\beta_1/\kappa_1)(1-e^{-\kappa_3 t} + \beta_1\kappa_3(\kappa_1\kappa_3-\kappa_1^2)^{-1}(e^{-\kappa_1 t}-e^{-\kappa_3 t}),$$

$$b(t) = a_2[t - \kappa_1^{-1}(1-e^{-\kappa_1 t}),$$

$$M(t) = a_2\cdot a_3\kappa_3^{-1}[t - (\kappa_1^{-1}+\kappa_3^{-1})(1-e^{-\kappa_3 t}) + \kappa_3(\kappa_1\kappa_3-\kappa_1^2)^{-1}\cdot(e^{-\kappa_1 t}-e^{-\kappa_3 t}),$$

$$a_1 = \mu a_2,$$

$$a_2 = 9K\alpha_T/(3K+4\mu),$$

$$a_3 = 9K/(2\mu t^*),$$

$$\kappa_1 = (3K+4\mu)/(4\mu t^*),$$

$$\kappa_3 = (3K+4\mu)/(\mu t^*),$$

$$\beta_1 = 1 - \kappa_1 t^*,$$

$$\begin{aligned} S_0(r,t) \; = \; & R_0^{-3}(3\zeta^2R_0^{-2}-1)\delta(t) - \frac{2\zeta}{t\beta^4}\exp\left(-\frac{R_0^2+\zeta^2}{2\beta^2}\right)\cdot \\ & \cdot\left[\left(1-\frac{r^2}{\beta^2}\right)I_0\left(\frac{r^2}{2\beta^2}\right)+\frac{r^2}{\beta^2}I_1\left(\frac{r^2}{2\beta^2}\right)\right], \end{aligned}$$

$$S_1(r,t) = R_0^{-3}\delta(t) - \frac{\zeta}{t\beta^4}\exp\left(-\frac{R_0^2+\zeta^2}{2\beta^2}\right)\left[I_0\left(\frac{r^2}{2\beta^2}\right) - I_1\left(\frac{r^2}{2\beta^2}\right)\right],$$

$$S_2(r,t) = \zeta R_0^{-3}\delta(t) - \frac{2\zeta}{t\beta^3\sqrt{\pi}}\exp(-R_0^2/\beta^2),$$

$\beta = \sqrt{4\kappa t}$, $I_1 = \dfrac{\mathrm{d}I_0(s)}{\mathrm{d}s}$ is modified Bessel function of the first kind, order 1, I_0 is the same, order 0, $\delta(t)$ is the Dirac function.

The heat flow anomaly due to temperature field is:

$$q_z(r,0,t) = \lambda_t\left[\frac{\partial T}{\partial z}\right]_{z=0} = \frac{w}{2\pi}\frac{\zeta}{R_0^3}\left\{\mathrm{erfc}\left[\frac{R_0}{4\kappa t} + R_0(\pi\kappa t)^{-1}\exp\left(\frac{-R_0^2}{4\kappa t}\right)\right]\right\}. \quad (9)$$

3. Perturbation of gravity

There are two principal reasons for the gravity change:

— the change of the density ρ_0 by increment $\Delta\rho$ due to volume dilatation,

— the free-air change and Bouguer correction as an effect of vertical uplift of the surface above the source of heat.

The density change $\Delta\rho$ generates perturbation of gravity potential, which for $z \geq 0$ obeys the Poisson equation. For the upper halfspace this potential analytically continues into harmonic potential (obeys Laplace equation).

For the gravity change due to volume dilatation was derived in Hvoždara (1998):

$$\Delta g_{TVE}(r,0,t) = \frac{1}{2}G\rho_0 Q\left\{W(t)\frac{-\zeta}{R_0^3} + 2\zeta\int_0^t W_2(t-\tau)\frac{e^{-R_0^2/(4\kappa\tau)}}{(2\kappa\tau)\sqrt{4\pi\kappa\tau^3}}\mathrm{d}\tau\right\}, \quad (10)$$

where $G = 6.67\times10^{-11}\mathrm{kg}^{-1}\mathrm{m}^3\mathrm{s}^{-2}$ is Newton's gravity constant

and $W_2(t) = \dfrac{9K\alpha_T}{6K+2\mu}\left[t - \kappa_3^{-1}(1-e^{-\kappa_3 t})\right]$.

The gravity effect due to upward doming of the surface of the Earth which was originally plane $z = 0$ is given by the sum of free-air change of gravity and Bouguer correction:

$$\Delta g_{FAB} = \left[-\frac{\partial g_0}{\partial z} + 2\pi G\rho_0\right]h(r), \quad (11)$$

where $h(r) = -u_z(0,r,t)$ and $\dfrac{\partial g_0}{\partial z} = 3.086\times10^{-6}\mathrm{m\,s}^{-2}/\mathrm{m}$ is the vertical gradient of normal gravity and $2\pi G\rho_0$ is the Bouguer correction.

4. Finite volume source of heat in viscoelastic halfspace

According to the principle of superposition we can employ the solution for the point source of heat such as the Green function, in order to calculate the effect due to the finite volume source of heat, i.e. for a certain magmatic body which occupies the volume V of the halfspace $z > 0$.

Let us suppose that the heat power due to the unit volume of the magmatic body is constant in the region V and equals to f_0. The temperature field outside the body in the time t is then given by the integral:

$$T(x,y,z,t) = \frac{f_0}{4\pi\lambda_T} \iiint_V \left[R_1^{-1} \operatorname{erfc}\left(\frac{R_1}{\sqrt{4\kappa t}}\right) - R_2^{-1} \operatorname{erfc}\left(\frac{R_2}{\sqrt{4\kappa t}}\right) \mathrm{d}V \right] \tag{12}$$

where

$$R_1 = \left[(x-x')^2 + (y-y')^2 + (z-z')^2\right]^{1/2},$$

$$R_2 = \left[(x-x')^2 + (y-y')^2 + (z+z')^2\right]^{1/2},$$

The calculation point is $P \equiv (x,y,z)$, $z > 0$, a source location point $Q \equiv (x',y',z')$ is now arbitrary point inside V. Formula (12) was generalized from the formula for point source of heat.

In this way we can also obtain displacements and stresses, but first it is necessary to determine the Cartesian components of displacements and stresses. According to the rules of vector and tensor transformations we can transform the components of displacements vector $\mathbf{u}$ and stress tensor $\boldsymbol{\sigma}$ from the cylindrical co-ordinate system (r,φ,z) to Cartesian co-ordinates (x,y,z). For the point source of heat we obtain:

$$\begin{aligned}
u_x &= u_r \cos\varphi \\
u_y &= u_r \sin\varphi \\
u_z &= u_z \\
\sigma_{xx} &= \sigma_{rr}\cos^2\varphi + \sigma_{\varphi\varphi}\sin^2\varphi, \\
\sigma_{yy} &= \sigma_{rr}\sin^2\varphi + \sigma_{\varphi\varphi}\cos^2\varphi, \\
\sigma_{zz} &= \sigma_{zz} \\
\sigma_{xy} &= (\sigma_{rr} - \sigma_{\varphi\varphi})\sin\varphi\cos\varphi, \\
\sigma_{xz} &= \sigma_{rz}\cos\varphi, \\
\sigma_{yz} &= \sigma_{rz}\sin\varphi,
\end{aligned} \tag{13}$$

where $\cos\varphi = (x-x')/r$,
$\sin\varphi = (y-y')/r$,
$r = [(x-x')^2 + (y-y')^2]^{1/2}$.

We now integrate over the volume of the source region V to obtain the total effect for displacements and stresses. We denote the integrated values of displacements by U_x, U_y, U_z and the integrated stress tensor components by S_{ij}, where $i,j = x,y,z$.

For the time dependence of the components of the displacements on the surface of the viscoelastic halfspace $z = 0$ we have the following formulae:

$$
\begin{aligned}
U_x(t) &= \tfrac{Q_1}{\pi} \iiint_V (x - x') \left[\int_0^t V(t-\tau) S_1(r,\tau) \mathrm{d}\tau \right] \mathrm{d}V' \\
U_y(t) &= \tfrac{Q_1}{\pi} \iiint_V (y - y') \left[\int_0^t V(t-\tau) S_1(r,\tau) \mathrm{d}\tau \right] \mathrm{d}V' \\
U_z(t) &= -\tfrac{Q_1}{2\pi} \left\{ \iiint_V (z - z') R_0^{-3} b(t) \mathrm{d}V' + \tfrac{2}{\sqrt{\pi}} \iiint_V (z - z') \cdot \right. \\
&\quad \cdot \left[\int_0^t b(t-\tau) \tau^{-1} (4\kappa\tau)^{-3/2} \exp\left(\frac{-R_0^2}{4\kappa\tau} \right) \mathrm{d}\tau \right] \mathrm{d}V' - \iiint_V (z - z') \cdot \\
&\quad \left. \cdot \left[\int_0^t W(t-\tau) \left\{ R_0^{-3} \delta(t) - \frac{2}{t\beta^3\sqrt{\pi}} \exp\left(\frac{-R_0^2}{\beta^2} \right) \right\} \mathrm{d}\tau \right] \mathrm{d}V' \right\},
\end{aligned}
\tag{14}
$$

where $Q_1 = \dfrac{f_0 w}{\lambda_T}$.

Components S_{zz}, S_{xz}, S_{yz} are zero at the surface $z = 0$. For the other components of the stress tensor we obtain:

$$
\begin{aligned}
S_{xx} &= \tfrac{2Q_1}{\pi} \iiint_V \left(\tfrac{2Q}{\pi} \right)^{-1} \left[\sigma_{rr} \frac{(x-x')^2}{r^2} + \sigma_{\varphi\varphi} \frac{(y-y')^2}{r^2} \right] \mathrm{d}V', \\
S_{yy} &= \tfrac{2Q_1}{\pi} \iiint_V \left(\tfrac{2Q}{\pi} \right)^{-1} \left[\sigma_{rr} \frac{(y-y')^2}{r^2} + \sigma_{\varphi\varphi} \frac{(x-x')^2}{r^2} \right] \mathrm{d}V', \\
S_{xy} &= \tfrac{2Q_1}{\pi} \iiint_V \left(\tfrac{2Q}{\pi} \right)^{-1} (\sigma_{rr} - \sigma_{\varphi\varphi}) \frac{(x-x')(y-y')}{r^2} \mathrm{d}V'.
\end{aligned}
\tag{15}
$$

As regards the stress tensor we are interested in the principal stresses and the maximum shear stress. We performed this analysis on the $z = 0$ plane. According to tensor analysis formulae the principal stresses in the x, y plane are given by the formula:

$$
S_{1,2} = \frac{1}{2}(S_{xx} + S_{yy}) \pm \frac{1}{2} \left[(S_{xx} - S_{yy})^2 + (2S_{xy})^2 \right]^{1/2}, \tag{16}
$$

S_1 and S_2 being the maximum and minimum stresses, respectively.

The maximum shear stress $\tau_{\max}$ is given by the difference:

$$
\tau_{\max} = \frac{1}{2}(S_1 - S_2) = \frac{1}{2} \left[(S_{xx} - S_{yy})^2 + (2S_{xy})^2 \right]^{1/2}. \tag{17}
$$

It is also interesting to evaluate the associated anomaly of the heat flow which is due to the temperature field:

$$
q_z = \frac{f_0}{2\pi} \iiint_V \frac{z'}{R_0^3} \left\{ \operatorname{erfc} \left[\frac{R_0}{4\kappa t} + R_0 (\pi\kappa t)^{-1} \exp\left(\frac{-R_0^2}{4\kappa t} \right) \right] \right\} \mathrm{d}V'. \tag{18}
$$

By using Hook's law we can express the strain components for the surface of the

visco-elastic halfspace:

$$\begin{aligned} \varepsilon_{xx} &= \frac{1}{2\mu} S_{xx} - \frac{\lambda}{\mu(6\lambda + 4\mu)}(S_{xx} + S_{yy}), \\ \varepsilon_{yy} &= \frac{1}{2\mu} S_{yy} - \frac{\lambda}{\mu(6\lambda+4\mu)}(S_{xx} + S_{yy}), \\ \varepsilon_{zz} &= \frac{\lambda}{\mu(6\lambda + 4\mu)}(S_{xx} + S_{yy}), \\ \varepsilon_{xy} &= \frac{1}{2\mu} S_{xy}. \end{aligned} \tag{19}$$

Dilatation $\Theta = \operatorname{div} \mathbf{u}$ can be expressed:

$$\Theta = \frac{1}{3\lambda + 2\mu}(S_{xx} + S_{yy}). \tag{20}$$

5. Numerical calculation and discussion

As a reference model we have used a point source of heat at the depth $\zeta = 10$ km, its intensity (power) $w = 2.6384 \times 10^7 W$, in order to achieve the epicentral heat flow anomaly $q_z(0) = 42$ mW/m^2, since $q_z(0) = w(2\pi\zeta^2)^{-1}$. For the elastic parameters of halfspace we put:

λ	=	7.05×10^{10} MPa	Lamé elastic constant,
μ	=	6.075×10^{10} MPa	Lamé elastic constant,
K	=	1.11×10^{11} MPa	bulk modulus,
ν	=	0.26857	Poisson ratio,
ρ	=	3×10^3 km.m^{-3}	density.

The decay time for the Kelvin's type of the viscoelastic body we put $t^* = 3.3 \times 10^{12}$ s. $t^* = \dfrac{\eta}{\mu}$, where η is the mean viscosity of crustal rocks.

For the thermal parameters of the medium we put following values:

$\lambda_T = 3$ W.m^{-1}K^{-1}	heat conductivity
$c_p = 840$ J kg^{-1}K^{-1}	specific heat under constant pressure
$\alpha_t = 10^{-6}$ K^{-1}	coefficient of linear thermal expansion
$\kappa = \lambda_T \cdot (c_p \cdot \rho)^{-1} = 1.1905 \times 10^{-4}$m^2 s^{-1}	coefficient of thermal conductivity.

The time evolution of the displacements u_r, u_z and terms of gravity anomaly Δg_{TVE}, Δg_{FAB} and their sum Δg_{SUM} as well as q_z was calculated for various time using multiplies of the characteristic heat disturbance time $t_\kappa = -\zeta^2(4\kappa\tau)^{-1}$, which correspond value e^{-1} of the known heat propagation factor $\exp(-\zeta^2/4\kappa\tau)$ in the

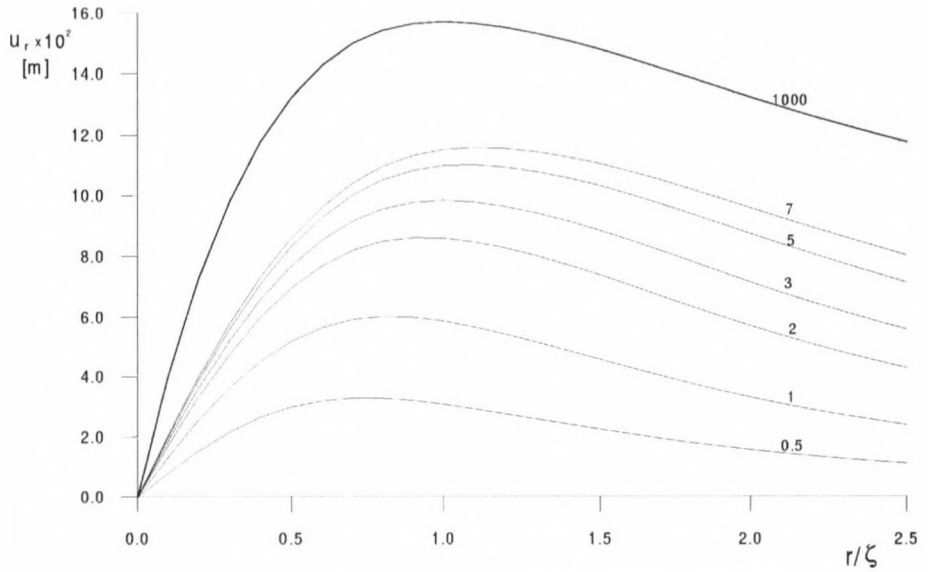


Fig. 1. Time variation of the thermo-viscoelastic displacement u_r on the surface of the viscoelastic halfspace

epicentre of the heat source. The results for the depth $\zeta = 10$ km ($t_\kappa = 2.1 \times 10^{13}$ s) are presented in Figs 1–6.

The curves for $t/t_\kappa = 0.5, 1.0, 2.0, 3.0, 5.0, 7.0$ gradually approach the curves for $t/t_\kappa = 1000$, which were calculated by means of formulae for stationary thermo-elastic problem (Hvoždara and Brimich 1991).

The most rapid approaching to the stationary value is in heat flow anomaly q_z in which curves formulae $t/t_\kappa = 5.$ and $7.$ are almost identical with the static curve $t/t_\kappa = 1000$. This effect can be explained by the formula (9), where is no term influenced by the decay time t^* of Kelvin's rheological body. The displacements, and gravitational anomalies approach their static values slowly because of viscoelastic behaviour of the halfspace which is mathematically expressed by the convolutional integrals in the previous chapter.

6. Conclusions

This paper is about the techniques of computing simple analytical models of surface displacements and gravity changes in a viscoelastic halfspace with a heat source. It is seen that in the viscoelastic lithosphere process of the thermoelastic deformation can be considered as a part of slow deformation of the earth's crust. Even though the model is an idealization of the real situation we have obtained useful results, which are in agreement with actual observations in geothermally disturbed areas (Okube and Watanabe 1984).

The directions and magnitudes of the displacements result in a pronounce dome in the epicentral region. The horizontal displacement points radially away from

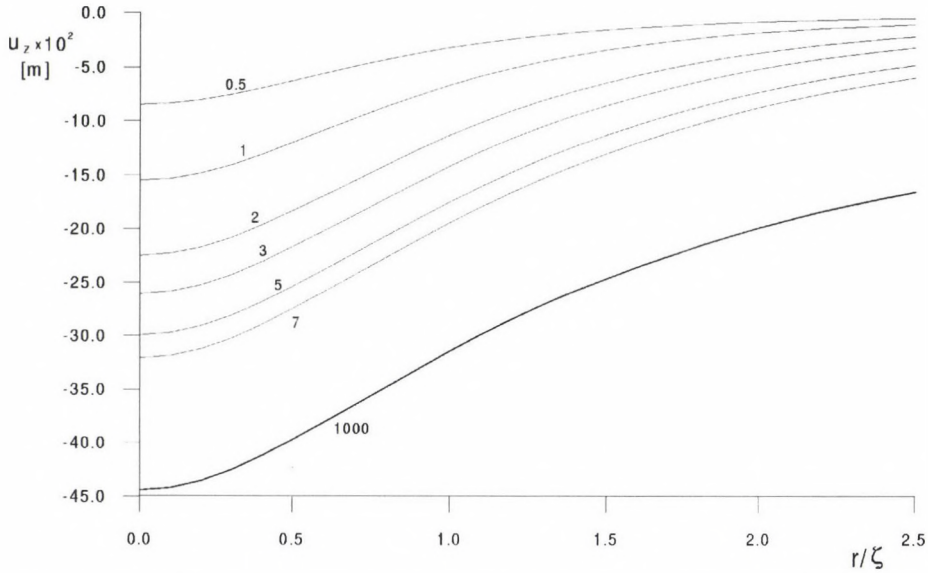


Fig. 2. Time variation of the thermo-viscoelastic displacement u_z on the surface of the viscoelastic halfspace

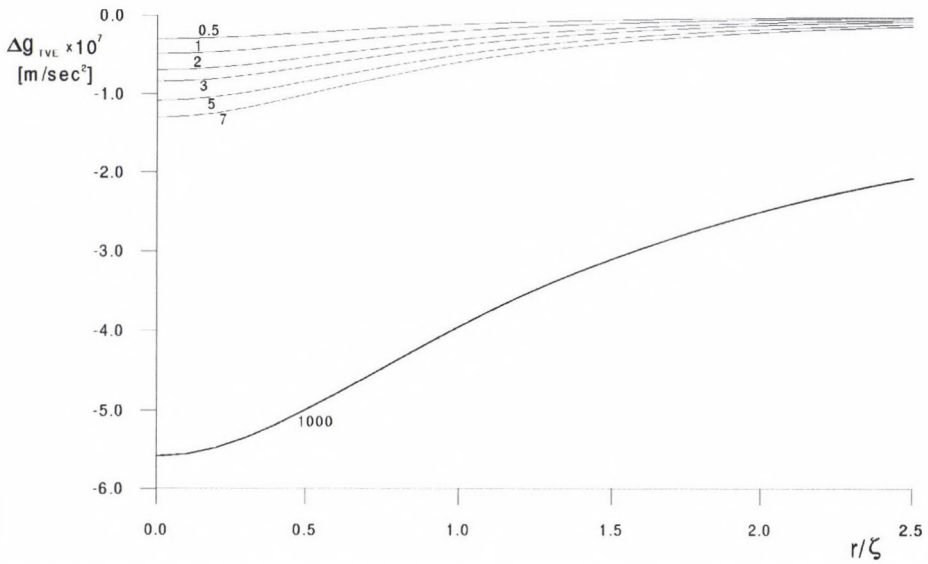


Fig. 3. Time variation of the gravity changes Δg_{TVE} due to volume dilatation on the surface of the viscoelastic halfspace

the epicentre. The main deformation is due to thermal effects produced by new magmatic injections or temperature changes in the magma chamber.

The gravity anomalies of order of 10^{-6} m.s^{-2}, connected with the deformation

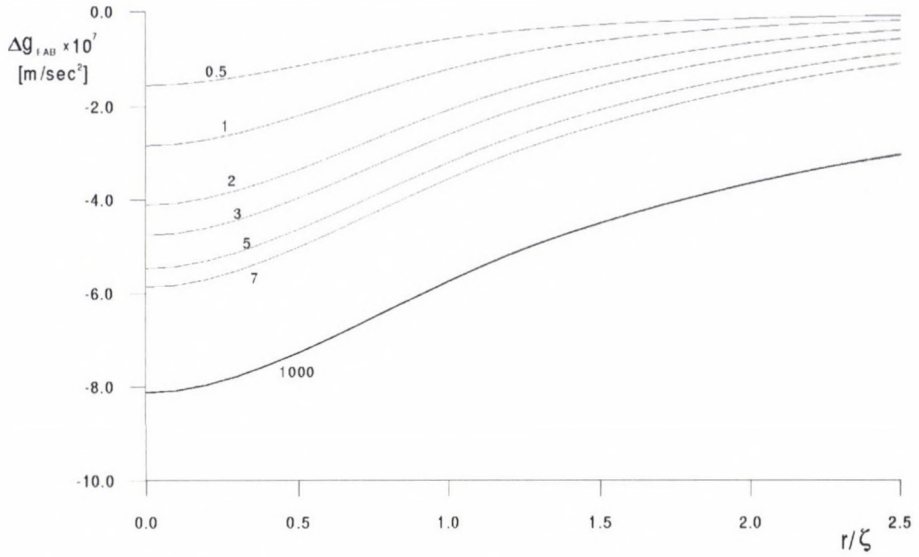


Fig. 4. Time variation of the gravity changes Δg_{FAB} due to upward doming on the surface of the viscoelastic halfspace

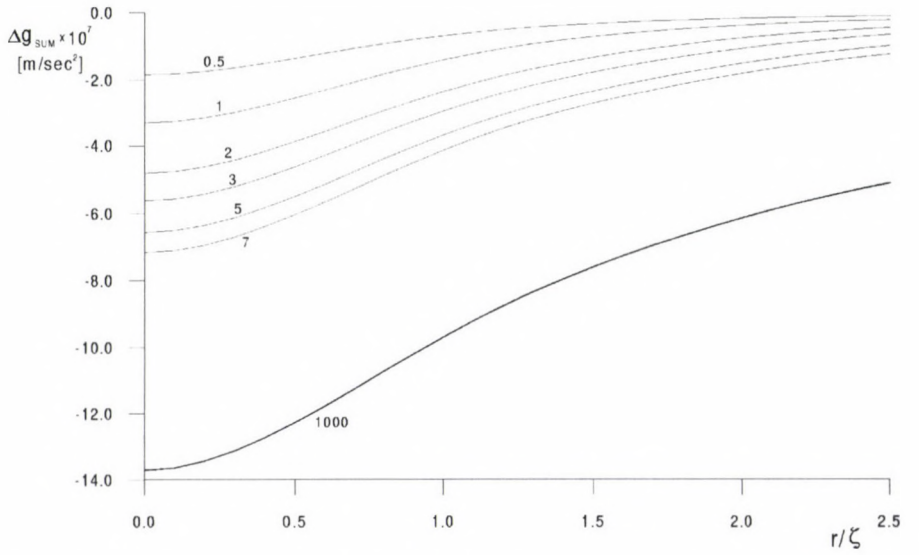


Fig. 5. Time variation of the total gravity changes Δg on the surface of the viscoelastic halfspace

fields, offer a valuable tool for understanding the geodynamic processes related to magmatic activity at an early stage.

An additional gravity effect can be expected due to transport of magmatic material of density different from crustal density, but this effect is not subject of this paper.

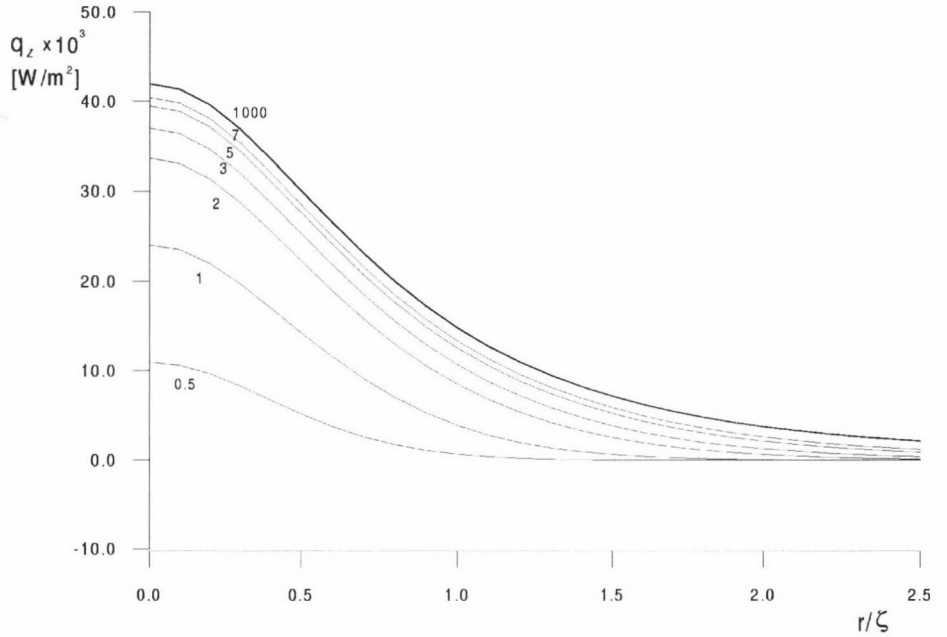


Fig. 6. Time variation of the heat flow anomaly q_z due to temperature field on the surface of the viscoelastic halfspace

References

Bianchi R, Coradini A, Federico C, Gibberti G, Lancian P, Pozzi J P, Sartoris G, Scandone R 1987: *Jour. Geoph. Res.*, 92, 14139–14150.

Carslaw H S, Jaeger J C 1959: Conduction of heat in solids. Oxford, Clarendon Press

Davis P M 1986: *Jour. Geoph. Res.*, 91, 7429–7438.

Dietrich J H, Decker R N 1995: *Jour. Geoph. Res.*, 80, 4094–4107.

Fernández J 1992: Técnicas geodésicas y geodinámicas aplicadas a la investigación del riesgo volcánico en la isla de Lauzarote. Doctoral disertation, Universidad Complutense de Madrid

Fernández J, Rundle J B 1994: *J. Geophys. Res.*, 99, 2737–2746.

Fernández J, Ting-To Y, Rundle J B 1996: *J. Geophys. Res.*, 101, 13581–13594.

Hofton M A, Rundle J B, Foulger G R 1995: *J. Geophys. Res.*, 100, 6329–6338.

Hvoždara M 1992: *Contr. Geoph. Inst. SAS*, 22, 48–66.

Hvoždara M 1998: Density and gravity changes due to point heat source buried in the viscoelastic halfspace. *Contr. Geoph. Geod.*, 28, 123–137.

Hvoždara M, Brimich L 1991: *Contr. Geophys. Inst. Slov. Acad. Sci.*, 21, 59–78.

Hvoždara M, Brimich L 1995: *Cahier du Centre Européan de Géodynamique et de Séismologie*, 8, 337–348.

Hvoždara M, Rosa K 1980: *Studia geophys. geod.*, 24, 51–60.

Nowacki W 1962: Thermoelasticity. Oxford, Pergamon Press

Okube S, Watanabe H 1989: *Geoph. Res. Lett.*, 16, 445–448.

Rundle J B 1980: *J. Geophys. Res.*, 85, 5355–5363.

Rundle J B 1982: *J. Geophys. Res.*, 87, 10729–10744.

Rundle J B 1983: *J. Geophys. Res.*, 88, 10647–10652.

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 49–61 (2000)

1D ANISOTROPIC VERSUS 2D ISOTROPIC MEDIA IN MAGNETOTELLURICS

F A M SANTOS[1] and L A MENDES-VICTOR[1]

The paper compares the characteristics of the distorted impedance tensor obtained over 1D anisotropic media with those of the impedance tensor from 2D media, studying the applicability of the current proceedings on dimensionality and directionality analysis. The results show that magnetotelluric data acquired over 1D media containing a deep anisotropic layer have rotational properties similar to those ones from 2D media and that it is not possible to distinguish between those media using only the impedance tensor. The results also show that distorted data from 1D media containing shallow and deep anisotropic layers have a complex behaviour dependent on the characteristics of the shallow anisotropy and that, in this case, the study of the dimensionality and directionality of the data is not a trivial subject.

Keywords: electrical anisotropy; galvanic distortion; magnetotellurics; structural dimensionality

1. Introduction

The magnetotelluric method uses surface measurements of natural electromagnetic fields to obtain information about the subsurface electrical conductivity distribution of the earth. Dimensionality and directionality indicators of the main geological structures are two important targets that can be reached using magnetotelluric data.

Most of that information comes from the study of the impedance tensor, which is obtained at various frequencies, from measurements of the two components of the natural earth's magnetic field and of two horizontal components of the electric field. At a given frequency ω (see, e.g., Vozoff 1972)

$$\vec{E}^o = Z^o \vec{H}^o$$

where $\vec{E}$ and $\vec{H}$ are the electric and magnetic fields, respectively and Z is the 2×2 impedance tensor. The superscript o indicates observed quantities.

The characteristics of the impedance tensor are only dependent on the observation site, measurement direction, frequency and on the electrical structure of the earth's subsurface. It is known that on an ideal layered earth (with isotropic layers) the impedance tensor is rotacionally invariant and has two null components, $Z_{xx} = Z_{yy} = 0$. In this case the off-diagonal elements of the impedance tensor are $Z_{xy} = -Z_{yx}$.

Over a two-dimensional earth (heterogeneous media), and when the measurement directions are not parallel and perpendicular to the geoelectric strike, all

[1]Departamento de Física da Universidade de Lisboa, Centro de Geofísica da Universidade de Lisboa, R. Escola Politécnica, 58, 1269–102 Lisboa, Portugal, e-mail: dfams@fc.ul.pt

1217-8977/2000/$ 5.00

components of the impedance tensor are different from zero with $Z_{xx} + Z_{yy} = 0$ and, $Z_{xy} \neq Z_{yx}$. If the acquisition directions are parallel and perpendicular to the strike, the diagonal elements are nulls. Therefore, over a two-dimensional earth one can obtain the geoelectric strike rotating the magnetotelluric impedance tensor up to reach the condition expressed above. However, as it is known, this result is only valid for noise-free and undistorted data set. In the presence of galvanic distortions the analysis of the impedance tensor is much more complex and different approaches were developed based on physical models (Larsen 1977, Zhang et al. 1986, Bahr 1988, Groom and Bailey 1989, amongst others). Accordingly, the observed magnetotelluric impedance tensor can be expressed by

$$Z^o = r\, C\, Z_{2\mathrm{D}}\, R^T$$

where R is a rotation matrix, $Z_{2\mathrm{D}}$ is the not distorted two-dimensional regional impedance tensor, C is the distortion tensor and T represents transpose. If the distortions are galvanic, the elements of the 2×2 tensor C are real and frequency independent

$$C = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} .$$

A question arises: what happens if the earth presents electric anisotropy? More specifically, it is possible to distinguish between magnetotelluric data acquired over a two-dimensional earth from data acquired over a one-dimensional anisotropic earth using only the impedance tensor? Cantwell (1960), Bostick and Smith (1962), Mann (1965) and many others have pointed the importance of anisotropic resistivity in magnetotelluric studies out several years ago. In fact, magnetotelluric field data can exhibit anisotropic behaviour. These anisotropies may arise due to crystal anisotropies of minerals or when regular bending of rocks with different resistivities result in an apparent anisotropy (Keller and Frischknecht 1966). Anisotropic phenomena have been reported by several workers, e.g., by Rasmussen (1988) and more recently by Sénéchal et al. (1996), Ádám (1996) and Pek and Verner (1997).

The main objectives of this note are to compare the characteristics of the distorted impedance tensor obtained over an one-dimensional anisotropic earth with those of the tensor from a two-dimensional earth and to study the applicability of the most current methods on dimensionality, directionality and distortion analysis, to anisotropic data. We start making a summary of the main characteristics of the impedance tensors obtained over two-dimensional media. The impedances from three cases of one-dimensional anisotropic earth are then analyzed. These studies were performed using synthetic data generated by a program based on the algorithm proposed by Loewenthal and Landisman (1973).

In this work the MT data will be depicted using the tensor apparent resistivities and phases derived from all the four elements of the impedance tensor instead of the elements tensor itself. Althought the significance of the diagonal elements resistivities and phases are not the same of those of the anti-diagonal, it is advised to use all those components when dealing with anisotropic data. In such a case, the different elements contain relatively independent information. This was the

procedure of many workers (see, e.g., Loewenthal and Landisman 1973, Reddy and Rankin 1975, Chakridi et al. 1992).

2. Summary of the characteristics of the distorted impedance tensor obtained over a two-dimensional earth

The 3D/2D model

Let us consider distorted magnetotelluric data acquired over a two-dimensional earth. If the distortions of the impedance tensor are originated for a three-dimensional thin sheet overlying the two-dimensional regional earth (the usually called 3D/2D model of distortion), the observed impedance tensor, in a co-ordinate system rotated accordingly with the regional strike, is (see, e.g., Chakridi et al. 1992)

$$Z^o = \begin{bmatrix} c_2 Z_{yx} & c_1 Z_{xy} \\ c_4 Z_{yx} & c_3 Z_{xy} \end{bmatrix}$$

where Z_{ij} $(i, j = x, y)$ represents the elements of the undistorted two-dimensional impedance tensor. Then column elements have identical phases and the corresponding apparent resistivity curves are parallel to each other, differing by a real multiplicative factor. This important result can be used to define the dimensionality of the impedance tensor and to determine the regional strike (Zhang et al. 1986, Chakridi et al. 1992).

The 2D/2D model

If the distortions of the impedance tensor are originated by two-dimensional structures with a strike (called local strike) different from the regional one, the observed impedance tensor expressed in the direction of the local strike has the diagonal elements proportional, i.e., related by a real constant (Zhang et al. 1986)

$$Z^o_{xx} = \alpha Z^o_{yy}$$

and, therefore, the phases of those elements are identical.

3. Characteristics of the impedance tensor from one-dimensional anisotropic media

In this study only one-dimensional earth with transverse anisotropy within some layers (Fig. 1-B) will be considered. Within an anisotropic layer and in the principal anisotropy directions, the electrical conductivities are $\sigma_1 \neq \sigma_2$. The strike of the anisotropy is indicated by the angle Θ_o between the co-ordinate acquisition system and the system of principal anisotropy directions. For this kind of models the vertical conductivity has no effect on the magnetotelluric response.

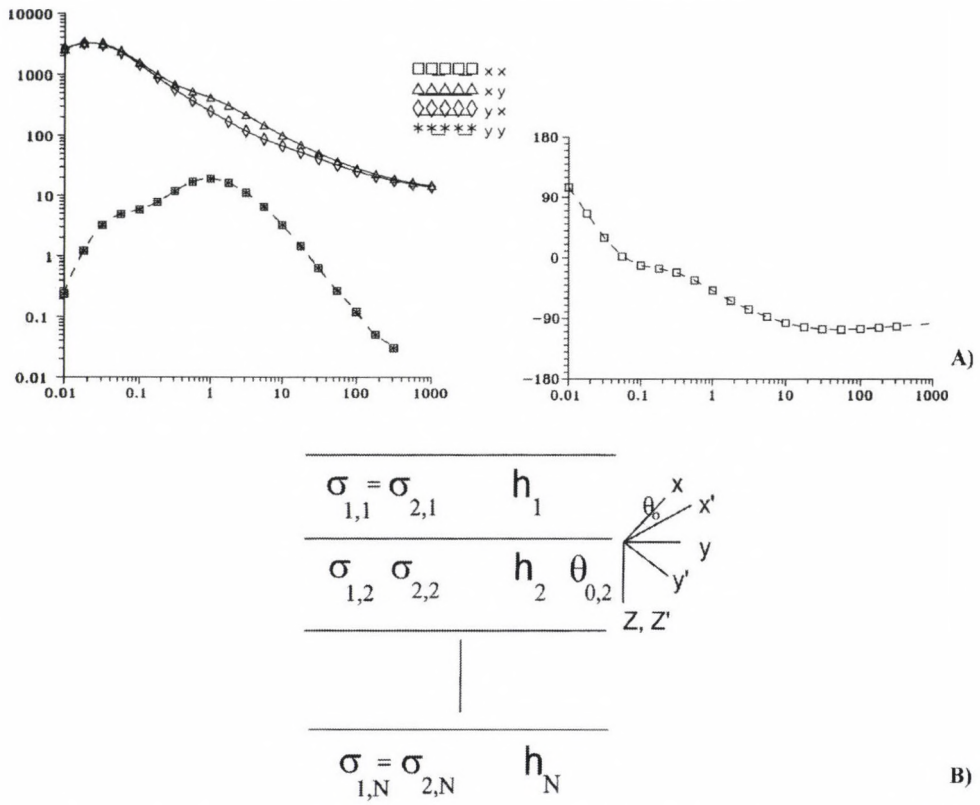


Fig. 1. A) Magnetotelluric responses (apparent resistivity and phase of the elements of the impedance tensors obtained in the x,y system) for the anisotropic 1D model shown in Table I. B) 1D anisotropic model: σ represents the electrical conductivity in the principal directions (x', y', z'), h is the thickness of the layer and Θ_0 the strike of the anisotropy relatively to the measurement directions (x, y, z)

3.1 1D earth with a deep anisotropic layer

The DA model

Let us analyse the response of a layered medium with a deep anisotropic layer (DA model, Table I and Fig. 1-A). In an arbitrary co-ordinate acquisition system, not aligned with principal anisotropic directions, the observed impedance tensor (noise-free and undistorted) has all its four components not null, with $Z_{xx} + Z_{yy} = 0$ and is not rotationally invariant.

If the data are acquired in a system rotated in accordance with the principal directions of the anisotropy, the diagonal elements of the tensor become nulls. It should be noted that this are similar characteristics to those shown by a two-dimensional tensor measured in a co-ordinate system rotated accordingly with the two-dimensional strike.

Table I. Conductivity model (DA type) used in the calculation of the data in Fig. 1 A

Layer	σ_1 S/m	σ_2 S/m	h km	Θ_0 deg
1	0.005	0.005	0.1	0
2	0.0001	0.0001	4.0	0
3	0.1	0.001	2.0	60
4	0.001	0.001	4.0	0
5	0.1	0.1	–	0

The 3D/DA model

Let us consider that a thin three-dimensional sheet distorts the impedance tensor, from a layered earth with a deep anisotropic layer (corresponding to undistorted regional impedance). In our case the distortion was incorporated using the tensor presented in Chakridi et al. (1992), which is

$$C = \begin{bmatrix} 1.26 & 0.44 \\ 0.53 & 0.86 \end{bmatrix}.$$

Those authors calculated this tensor from a 3D/1D model. The interested reader can find details in the mentioned reference. It should be noted that our basic one-dimensional model, before the introduction of the anisotropic layers, is the same of those authors (Table I).

In an acquisition co-ordinate system rotated according to the principal anisotropy directions, the observed impedance tensor becomes:

$$Z^o(\Theta_0) = \begin{bmatrix} c_2 Z_{yx} & c_1 Z_{xy} \\ c_4 Z_{yx} & c_3 Z_{xy} \end{bmatrix}.$$

Then, the columns have identical phases and the correspondent resistivity curves are parallel (Fig. 2). These features are similar to those presented by distorted impedance tensors acquired over two-dimensional media. Therefore, the distinction between the two media is not possible using only the rotational analysis of the impedance tensor. Nevertheless, one can distinguish the two-dimensional earth from the one-dimensional anisotropic one using the vertical magnetic component, which is null in the latest case but not null over a two-dimensional earth. The problem arises when the vertical magnetic component is not available, as frequently happen.

Figure 3 shows the results of the Groom-Bailey (G-B) and Bahr's decompositions of the impedance tensor obtained from the model in Table: considering the 90° ambiguity, we can say that the G-B or the Bahr's regional strike estimations are identical to the direction of the deep anisotropy. In the G-B decomposition case the values of the twist and shear parameters are $+6°$ and $-24°$, respectively.

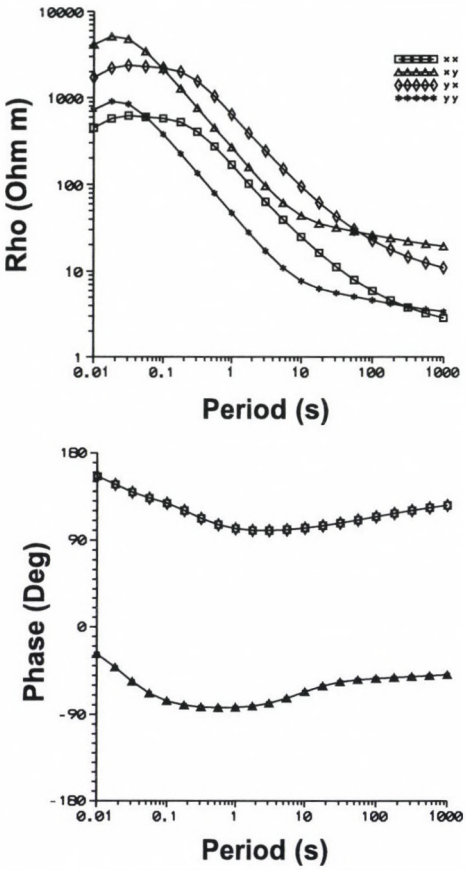


Fig. 2. Distorted apparent resistivities and phases of the impedance tensor in a co-ordinate system oriented according to the principal directions of the deep anisotropy. Those data were obtained from the model shown in Table I and distorted using the tensor presented by Chakridi et al. (1992)

3.2 1D earth with deep and shallow anisotropic layers

The (SA+DA) model

Let us now consider the data (noise-free and undistorted) generated by a layered earth containing two anisotropic layers with different principal directions and at different depths. The observed diagonal elements of the impedance tensor rotated accordingly with the deep anisotropy strike depend only on the shallow anisotropy. All layers influence the non-diagonal elements, then the general form of the tensor is:

$$Z^o(\Theta_0) = \begin{bmatrix} Z^s_{xx} & Z^{s+d+o}_{xy} \\ Z^{s+d+o}_{yx} & Z^s_{xy} \end{bmatrix}.$$

Here $Z_{xx} = -Z_{yy}$. s and d mean that the element is influenced by shallow and deep anisotropic layers, respectively and o represents the influence of isotropic layers.

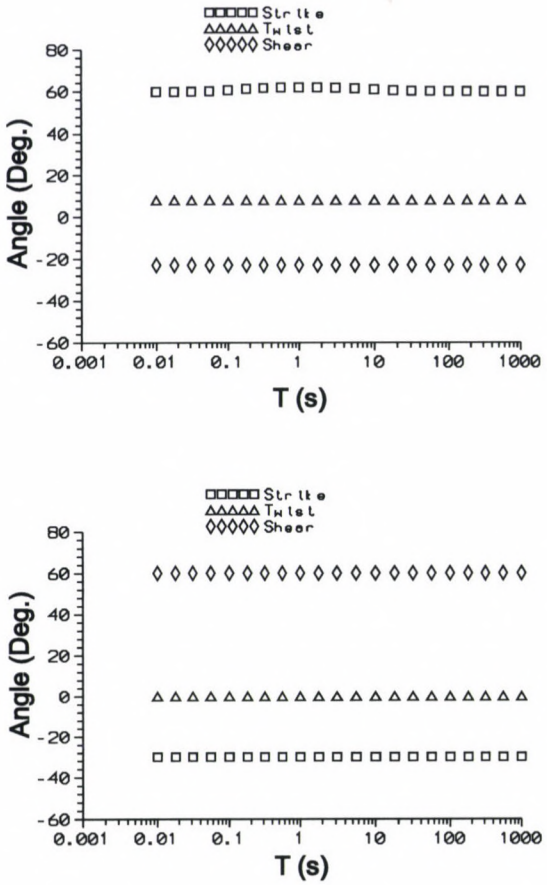


Fig. 3. Groom-Bailey's (upper) and Bahr's (bottom) decompositions of the impedance tensor shown in Fig. 2 (3D/DA model). The decomposition parameters are represented as following: squares-strike; triangles-twist; diamond-shear

Figure 4 shows the apparent resistivities and the phases of the elements of the impedance tensor rotated accordingly with the strike of the deep anisotropic layer (case A in Table II). In this example the shallow anisotropic layer has a strike of 25° while the strike in the deep layer is of 60°. Note that the high conductivity direction within the shallow anisotropic layer is −65°.

As expected the apparent resistivities and phases of the diagonal elements are identical, due to the shallow anisotropy. In fact, as the acquisition co-ordinate system has the direction of the deep anisotropy the contribution of this layer to those components of the tensor is nulls. The phases of the column elements are different, except for a short period range (up to 0.03 s). Similar results were obtained from other (SA+DA) models. These results show that the behaviour of the shallow anisotropic layer is not similar to that of a galvanic thin sheet. If it was the case the apparent resistivities curves of the column elements of the rotated tensor should be parallels and its phases identical.

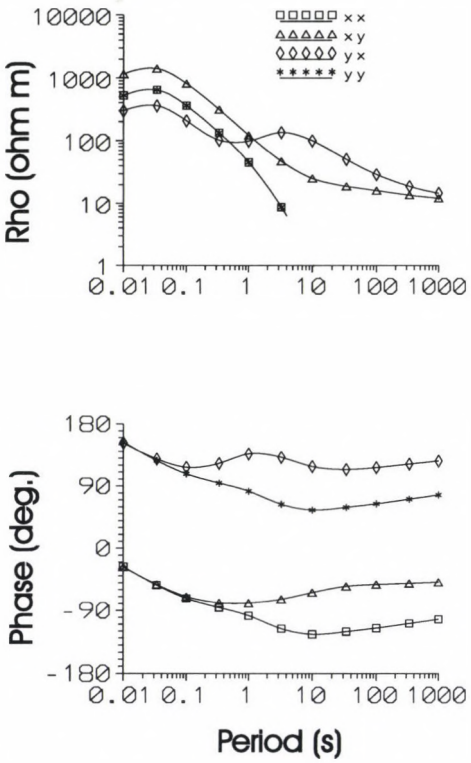


Fig. 4. Not distorted apparent resistivity and phase data obtained from model A shown in Table II, in a reference system oriented in accordance with the deep anisotropy

3D/(SA+DA) models

Let us consider the impedance tensor from a model (SA+DA) distorted by a thin 3D sheet. In an co-ordinate system rotated accordingly the deep anisotropy the general form of the impedance tensor becomes:

$$Z^o(\Theta_0) = \begin{bmatrix} c_1 Z^s_{xx} + c_2 Z^{s+d+o}_{yx} & c_1 Z^{s+d+o}_{xy} + c_2 Z^s_{yy} \\ c_4 Z^{s+d+o}_{yx} + c_3 Z^s_{xx} & c_4 Z^s_{yy} + c_3 Z^{s+d+o}_{xy} \end{bmatrix} .$$

This expression of the tensor rotated accordingly the deep strike shows that in general, 1. the phases of the impedance tensor from a 3D/(SA+DA) case are not identical to the phases of the (SA+DA) case and, 2. at long periods, when the influence of the shallow anisotropy does not exist, the phases of the column elements are identical.

Figure 5 shows the apparent resistivities and the phases of the four elements of the perturbed impedance tensor rotated accordingly with the principal directions of the deep anisotropy. Three cases were considered (Table II). In the first case (A) the shallow anisotropic layer is the uppermost layer. In the second case (B) the uppermost anisotropic layer is underlying a thin (100 m) isotropic layer. The

third case (C) differs from the second on the principal directions of the shallow anisotropic layer: in the case C the horizontal anisotropic axes are rotated of 90°, relatively to those ones in case B.

In the case A, the apparent resistivity curves for the column elements are parallel at frequencies greater than 10 s. A similar feature is observed at long periods in the case B. It can be noted that the behaviour of the phases of the column elements are also strongly dependent on the characteristics of the shallow anisotropic layer. The expected identicalness of the phases of the column elements in both cases, A and B, is perturbed in a frequency range that depends on the characteristics (depth and thickness) of the shallow anisotropy.

In the case C the behaviour of the phases seems to be less complex and for both short and long periods the identity between the phases is almost reached. Note that the angular distance between the two principal directions corresponding to the high conductivity in the anisotropic layers, is lesser in the model C (35°) than in cases A and B (125°).

These results indicate that in cases like A and B, where the anisotropic layers show distinct strikes, one can expect to distinguish the principal direction within the shallow layer from that one in the deep layer using, e.g., the G-B method. However, increasing the proximity of the anisotropic layers increases the difficulty on that distinction.

Groom-Bailey decomposition

Figure 6 shows the results obtained from the Groom-Bailey's (G-B) decomposition of the three above cases, leaving all the parameters free to change. Relatively to strikes we conclude that in all three studied cases it was possible to recover the strike corresponding to the shallow anisotropic layers (25°). However, the deep strike was only recovered at long periods in the case A (−30° or 60°). In the case B, and in the period range used in this study, it was not possible to recover the correct strike of the deep layer, even at long periods: the estimated value of strike is −10°. In the case C, the behaviour of the strike at the longest periods (+45°) seems to indicate that it would be possible to recover the deep strike if periods greater than 1000 s was used. These features mean that it might be possible to recover the strikes associated with anisotropic layers, the shallow (at the shortest periods) and deep ones (at the longest periods). However, the recover of the deep strike is more difficult when the depth (or the thickness) of the uppermost anisotropic layer increases.

Quite constant values of the G-B parameters shear and twist, are obtained at some period ranges separated by periods where the parameters present more perturbed features. These parameters represent the amount of angular distortion that regional electric field experiments. In our case those distortions are combinations of the distortion due to the distortion tensor and due to the shallow anisotropic layer.

According to Groom and Bailey (1989), the shear develops anisotropy in the axes that bisect the regional induction axes. In the cases A and B the shear is quite constant at periods ranging from 0.2 to 10 s with values ranging from −18° to −20°.

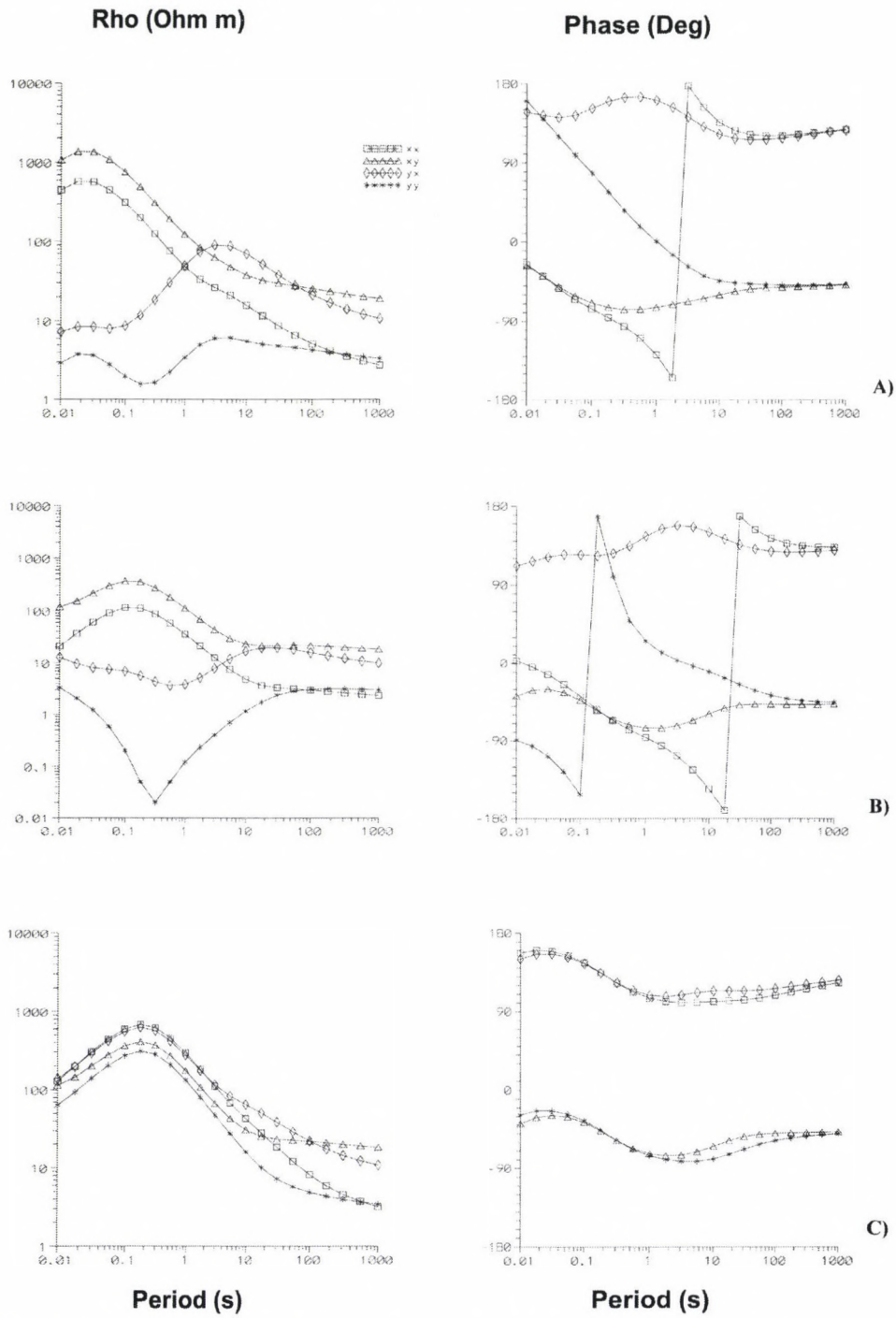


Fig. 5. Distorted apparent resistivities and phases data obtained from models A, B and C shown in Table II, in a reference system oriented in accordance with the deep anisotropy

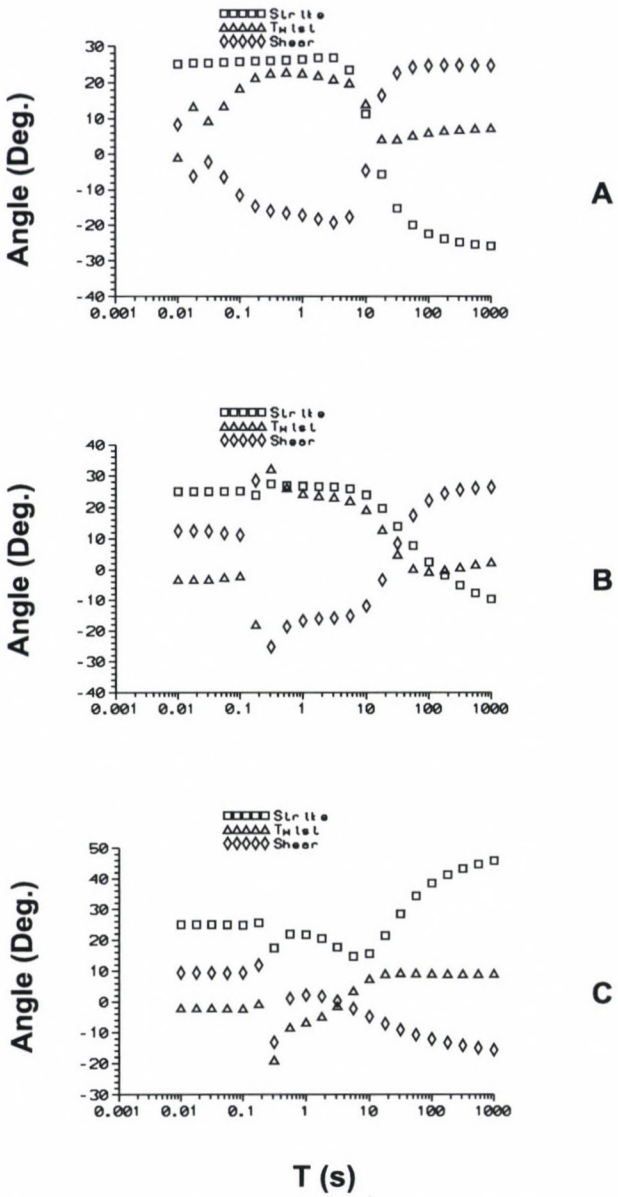


Fig. 6. Groom-Bailey's decomposition of the impedance tensors obtained from the 3D/(SA+DA) models shown in Table II. The decomposition parameters are represented as following: squares-strike; triangles-twist; diamond-shear

At periods comprised between 30 and 1000 s in the case A and, from 200 to 100 s in case B, the shear parameter is also constant but its value is approximately +25. Those results seem indicate that the shear is strongly influenced by anisotropies (compare with 3D/DA shear and twist values).

Table II. Conductivity models (SA+DA types) used in the calculation of synthetic results

	Case A				Case B				Case C			
Layer	σ_1 S/m	σ_2 S/m	h km	Θ_0 Deg	σ_1 S/m	σ_2 S/m	h km	Θ_0 Deg	σ_1 S/m	σ_2 S/m	h km	Θ_0 Deg
1	0.005	0.5	0.1	25	0.005	0.005	0.1	0	0.005	0.005	0.1	0
2	0.0001	0.0001	4.0	0	0.005	0.5	0.5	25	0.5	0.005	0.5	25
3	0.25	0.001	2.0	60	0.0001	0.0001	3.5	0	0.0001	0.0001	3.5	0
4	0.001	0.001	4.0	0	0.25	0.001	2.0	60	0.25	0.001	2.0	60
5	0.1	0.1	–	0	0.001	0.001	4.0	0	0.001	0.001	4.0	0
6					0.1	0.1	–	0	0.1	0.1	–	0

In the case C the shear parameter is only constant at the shortest periods not influenced by the shallow anisotropy. At long periods the shear decreases monotonically to values of order of $-20°$. In this case (compare the directions of high conductivity within anisotropic layers in cases A, B and C) the anisotropy has more influence at short periods than at long ones.

The effect of the twist is to rotate the regional electric field. In the two first cases (A and B) and at periods ranging from 0.2 to 8 s, the values of the twist parameter are approximately $+20°$. At long periods the twist is near to $+6°$. In case C the behaviour of the twist is similar at long periods. Those results can be interpreted as following: the influence of the shallow anisotropy in the rotation of the regional field is more important at short periods.

4. Conclusions

The aim of this paper was to compare the characteristics of the distorted impedance tensors obtained over a one-dimensional anisotropic earth with those obtained over a two-dimensional earth and to analyse the applicability of the current methods in the study of the dimensionality and directionality of anisotropic data.

The results, obtained using synthetic data, show that:

1. The distorted magnetotelluric data acquired over an one-dimensional earth containing only a deep anisotropic layer have a rotational behaviour similar to those from a two-dimensional earth and only using the vertical magnetic component is possible to distinguish between both earth models.

2. The distorted data acquired over a one-dimensional earth containing shallow and deep anisotropic layers have a much more complex behaviour dependent on the characteristics (depth, thickness and orientation relatively to the deep anisotropy) of the shallow anisotropy. In this case the study of the dimensionality and directionality of the data is not a trivial subject and the usual tools must be used carefully.

Acknowledgements

The authors are grateful to Josef Pek for many valuable discussions and constructive comments. Thanks are also due to L Szarka and an anonymous reviewer for critical comments. This work was supported by the Fundaao para a Ciencia e Tecnologia (FCT/MCT).

References

Ádám A 1996: *Acta Geod. Geoph. Hung.*, 31, 191–216.
Bahr K 1988: *Journ. Geophys.*, 62, 119–127.
Bostick F X, Smith H W 1962: *Proc. IRE*, 50, 2339–2346.
Cantwell T 1960: Detection and analysis of low frequency magnetotelluric signals, Ph.D. thesis, MIT.
Chakridi R, Chouteau M, Mareschal M 1992: *Geophys. Journal International*, 108, 917–929.
Groom R W, Bailey R C 1989: *J. Geophys. Res.*, 94, 1913–125.
Keller G V, Frischknecht F C 1966: Electrical methods in geophysical prospecting. Pergamon Press
Larsen J C 1977: *Acta Geod. Geoph. Mont. Hung.*, 12, 183–186.
Loewenthal D, Landisman M 1973: *Geophys. J. R. astr. Soc.*, 35, 195–214.
Mann J E 1965: *J. Geophys. Research*, 70, 2940–2942.
Pek J, Verner T 1997: *Geophys. J. Int.*, 128, 505–521.
Rasmussen T M 1988: *J. Geophys. Research*, 93, 78997–7907.
Reddy I K, Rankin D 1975: *Geophysics*, 40 (6), 1035–1045.
Sénéchal G, Rondenay S, Mareschall M, Guilbert J, Poupinet G 1996: *Geophysical Research Letters*, 23 (17), 2255–2258.
Vozoff K 1972: *Geophysics*, 37, 98–141.
Zhang P, Roberts R G, Pedersen L B 1986: *Geophysics*, 52, 267–278.

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 63–73 (2000)

ABOUT THE DEPTH OF INVESTIGATION OF DIFFERENT D.C. DIPOLE-DIPOLE ARRAYS

S SZALAI[1]

[Manuscript received April 8, 1999]

The depth of investigation characteristic (DIC) functions of the most commonly used dipole arrays are presented in this paper as a function of the characteristic angle between the transmitter and the receiver dipoles. The depth of investigation for each of the arrays except the parallel dipole-dipole array has a specific value, which is independent on the characteristic angle. From the point of view of the depth of investigation obtained, the dipole-dipole arrays form two groups: (1) the axial-like and (2) the equatorial-like arrays. The widely used statement by Roy and Apparao (1971) that the parallel dipole array with the special angle $\vartheta^* = 54°44'8"$ would have an infinitely large depth of investigation must be modified. After a small modification in the interpretation process it becomes unambiguous that this array does not see deeper than most of other dipole-dipole arrays.

Keywords: dept of investigation function; dipole-dipole arrays; direct current

Introduction

The depth of investigation characteristic function (DIC) for different dipole-dipole arrays was defined by Roy and Apparao (1971). They presented analytical calculations for an infinitely thin and resistive layer in a homogeneous halfspace. The depth of investigation was defined as the depth of the layer where the DIC response had its maximum.

Although these DIC functions have been widely used till now (see for example Knödel et al. (1997)), the DIC functions of different dipole arrays with the exception of the dipole axial and dipole equatorial arrays have not been studied yet.

Furthermore, a close insight into the DIC functions by Roy and Apparao (1971) shows some discrepancies in their interpretation. Namely, Roy and Apparao (1971) thought that the depth of investigation of parallel arrays in special situations, when the electric field above homogeneous halfspace is zero, is infinitely large. It is not correct as it is shown below.

In this paper at first the main steps of calculation of DIC functions are summarized then I show DIC functions for the traditional dipole-dipole arrays. They are presented in a detailed way (with a step of 10 degrees). In the end of the paper I suggest a slight modification in the interpretation method by Roy and Apparao (1971).

[1]Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H–9401 Sopron, POB 5, Hungary

1217-8977/2000/$ 5.00

The DIC function

Roy and Apparao (1971) derived this function in an analytical way by integrating the effect of small cubes. These cubes were handled as three dipoles which are perpendicular to each other, created by the external electric field. To get the response of a thin layer (sensitivity curves), it is enough to summarise the dipole components perpendicular to the xy plane, since the resulting electric field due to the horizontal elementary dipoles is zero.

The electric field E_z^R of an elementary vertical dipole due to the presence of a small size cube was given by Szalai and Szarka (1999) as follows.

$$E_z^R = \frac{p_T}{2\pi^2\varepsilon_0} \cdot \frac{ka^3}{r_{TC}^5 \cdot r_{RC}^5} \cdot G_{zC}^T \cdot G_{zC}^R \tag{1}$$

where

$$r_{TC}^2 = x_C^2 + y_C^2 + z_C^2 \tag{2a}$$

p_T is the dipole moment of the transmitter, k is the reflection coefficient and a the sidelength of the cube, respectively r_{TC} is the transmitter-cube distance and x_C, y_C and z_C are the coordinates of the cube in the coordinate system of the cube, see Fig. 1.

$$r_{RC}^2 = x_C^2 + (y_C - R)^2 + z_C^2 \tag{2b}$$

r_{RC} is the receiver-cube distance, R is the transmitter-receiver distance,

$$G_{zC}^T = 3x_T z_T \tag{2c}$$

with

$$x_T = x_C \cdot \cos\beta + y_C \cdot \sin\beta \tag{2d}$$

and

$$z_T = z_C \tag{2e}$$

and

$$G_{xC}^R = 3x_R z_R \tag{2f}$$

with

$$x_R = x_C \cdot \cos\gamma + (y_C - R) \cdot \sin\gamma \tag{2g}$$

and

$$z_R = z_C \tag{2h}$$

T and R indices describe the coordinates in the coordinate system of the transmitter and of the receiver respectively.

β and γ are shown in Fig. 1. β is the angle between the x axis of the coordinate system of the transmitter and the x axis of the cube; γ is the angle between the x axis of the coordinate system of the receiver and the x axis of the cube. Since the

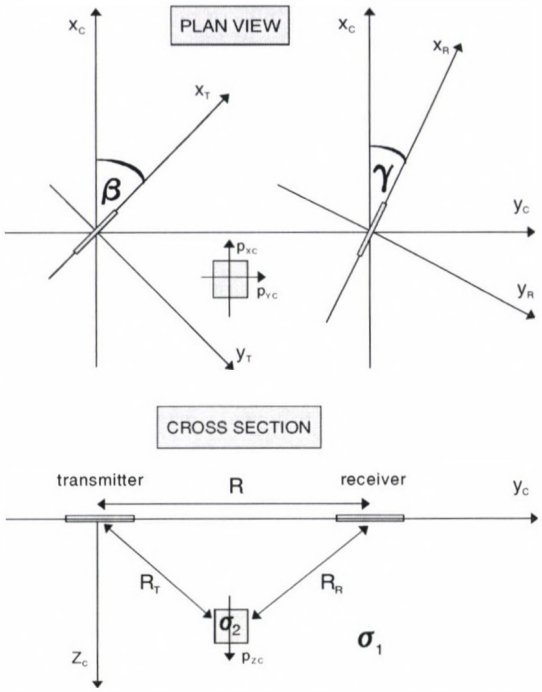


Fig. 1. The geometry of the problem to calculate the effect of a small size cube

component E_z^R is to be integrated for the whole x_C, y_C plane at a depth of z_C, the final formula for the DIC function is as follows.

$$\begin{aligned} \mathrm{DIC}(z_C) \quad = \quad & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_z^R dx_C dy_C = \frac{p_T}{2\pi^2\varepsilon_0}\cdot \\ & \cdot ka^3 z_C^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{x_C\cos\beta + y_C\sin\beta}{(x_C^2+y_C^2+z_C^2)^{5/2}}\cdot \\ & \cdot \frac{x_C\cos\gamma + (y_C-R)\sin\gamma}{[(x_C^2+(y_C-R)^2+z_C^2]^{5/2}}\, dx_C dy_C\,. \end{aligned} \tag{3}$$

The four fundamental types of dipole-dipole arrangements (D1, D2, D3 and D4) for which DIC functions were computed are shown in Fig. 2. The different arrays are simpler to be studied as a function of one single angle, ϑ. The transformation formulas from β and γ to ϑ are as follows:

D_1 (parallel dipole array):	$\gamma = \vartheta + 90°$
D_2 (perpendicular array):	$\gamma = \vartheta$
D_3 (radial array):	$\gamma = 90°$
D_4 (azimuthal array):	$\gamma = 0°$
and in all above configurations:	$\beta = \vartheta + 90°$ (see Fig. 2).

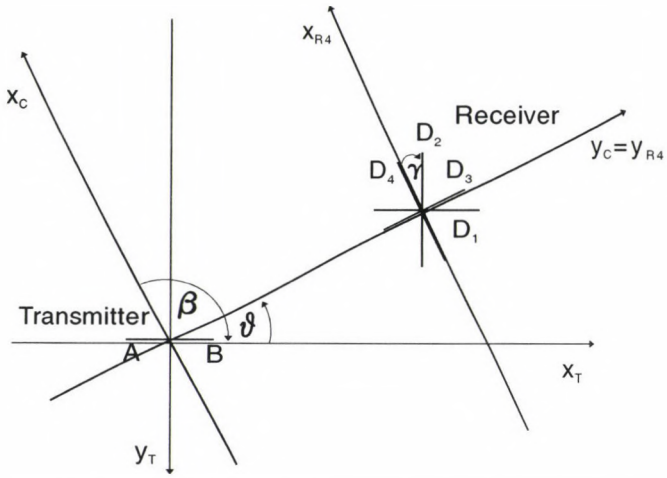


Fig. 2. Illustration for the transformation from β and γ to ϑ

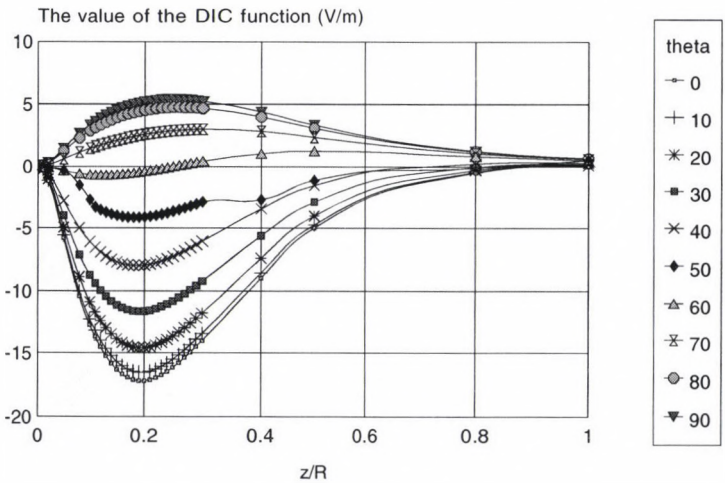


Fig. 3. DIC functions for parallel dipole arrays

An analysis of the DIC functions

At first DIC functions for the aforementioned four dipole arrays as the parallel D_1 (Fig. 3), the perpendicular (D_2, Fig. 4), the radial (D_3, Fig. 5) and the azimuthal dipole (D_4, Fig. 6) arrays with ϑ changing from 0° to 90° by a step of 10 degrees are shown.

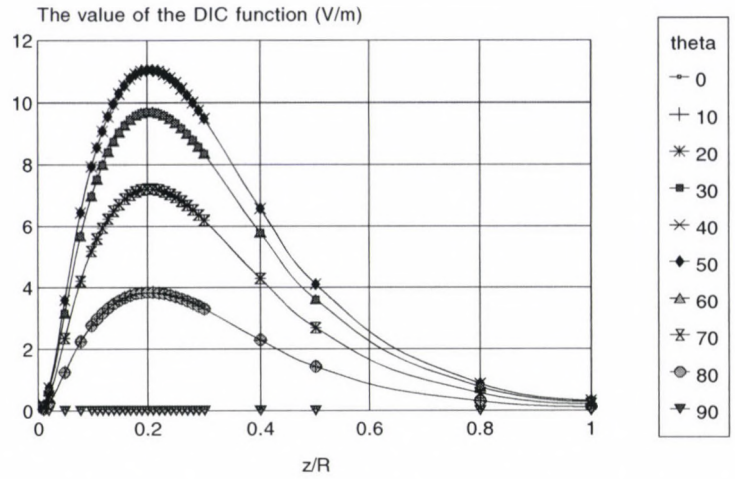


Fig. 4. DIC functions for perpendicular dipole arrays

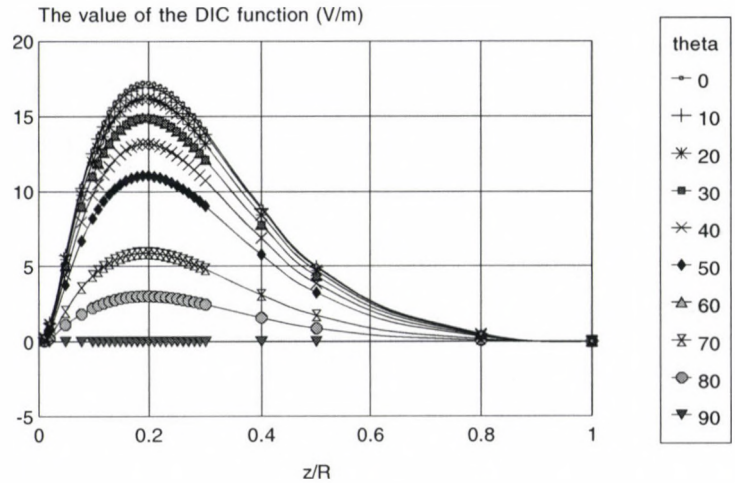


Fig. 5. DIC functions for radial dipole arrays

Asymptotes

All of the DIC curves have a zero value at a depth $z/R = 0$ and at an infinitely large depth. I note that for $z/R = 0$ this value is zero, because the effects of small volume elements will cancel each other. Individually they may cause even a very strong field. This is the reason why these electrode systems are very sensitive to near-surface heterogeneities (Roy and Apparao 1971).

Furthermore we should not forget that these curves are sensitivity curves, which means that T_2 (transverse resistance) or S_2 (longitudinal conductance) parameters are very small (not larger than 0.001). If they had a higher value, the curves

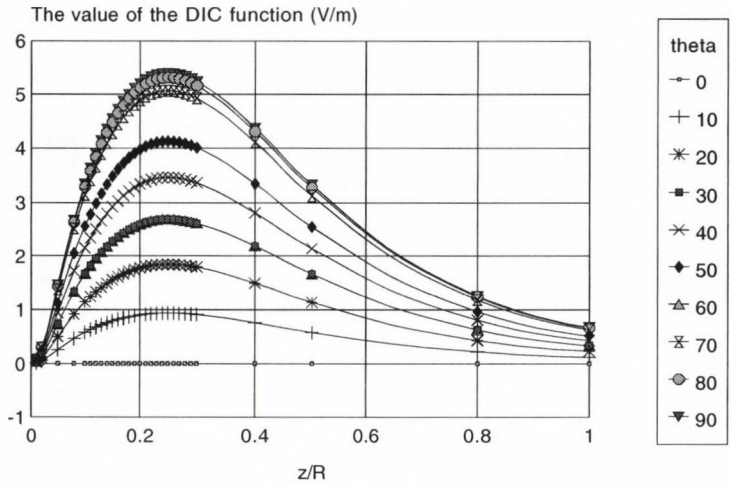


Fig. 6. DIC functions for azimuthal dipole arrays

would change strongly and the maximum of the curves would move towards the surface. The greater the T_2 (or the S_2) values are, the nearer have the curves their maximum values to the surface. If the thin layer is "really" resistive (or conductive), the maximal value should be close to the surface, since the currents are restricted to flow between the surface and the layer (or in the conducting layer). In our case the layer thickness is very small, as well as the T_2 and S_2 values, too.

The sign of the DIC functions

We observe that the DIC function in the $z/R \leq 1$ domain changes its sign only in case of the parallel array. (At larger depts the DIC function values are not interesting from the investigation point of view). The change of sign in case of parallel arrays is not a surprise, since the electric field around an x-directed horizontal electric dipole above homogeneous halfspace (Alpin et al. 1966) are as follows.

$$\begin{aligned} E_x^o &= \tfrac{p}{r^3}(3\cos^2\vartheta - 1) \\ E_y^o &= \tfrac{3p}{2r^3}\sin 2\vartheta \\ E_r^o &= \tfrac{2p}{r^3}\cos\vartheta \\ E_\vartheta^o &= \tfrac{p}{r^3}\sin\vartheta \end{aligned} \tag{4}$$

Here p is the dipole moment, r is the distance between the transmitter and receiver dipoles. It can be clearly seen that it is only the E_x^o component, which changes its sign above a homogeneous halfspace if $0 \leq \vartheta \leq \frac{\pi}{2}$.

This is the reason for the more complicated behaviour of the parallel array than that of the other ones.

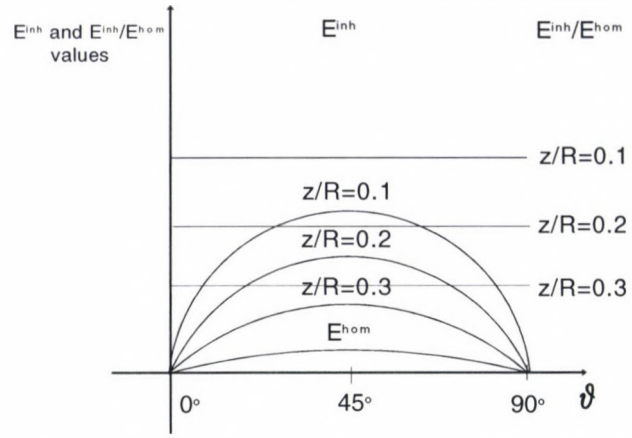


Fig. 7. Behaviour of the dipole perpendicular array as a function of ϑ. Scheme

DIC functions with varying ϑ

Another interesting feature is that in case of the DIC groups (with the exception of the parallel arrays again) the DIC values belonging to a constant depth of the investigation layer are approximately the same function of ϑ as in case of a the homogeneous halfspace. It means, if these values are normalized by the homogeneous field values (measured at the corresponding angle) the result is nearly a constant (Fig. 7). Consequently, the effect of a thin layer (if the effect of a thin layer means the signal above it normalized by the signal above homogeneous halfspace) at a fixed depth is the same for all arrays, independently of ϑ. At the same time, the value of this constant depends obviously on the depth of the thin layer. This value decreases with increasing depth if the depth is greater than the depth of the maximal value of the given DIC curve and it approximates zero at very large depths in case of all arrays.

This is not the case for the parallel array, which has a much more complicated feature from this point of view, too.

Although the normalization leads to interesting conclusion in this case, it may not be made always automatically as it will be shown below.

The depth of investigation

The depth of investigation was defined by Roy and Apparao (1971) as the depth where the DIC curve has a maximal value. In their paper they determine the depth of investigation only for the dipole equatorial array and for the dipole axial one and for all dipole arrays at the angle $\vartheta = 45°$.

As it was discussed earlier, with exception of the parallel array, the DIC curves are similar to each other at any ϑ. It means that the depth of investigation for these arrays does not depend on ϑ. The depth of investigation for the dipole perpendicular arrays is at $z/R = 0.2$, for the radial arrays it is at $z/R = 0.19$ and for the azimuthal array it is at $z/R = 0.25$.

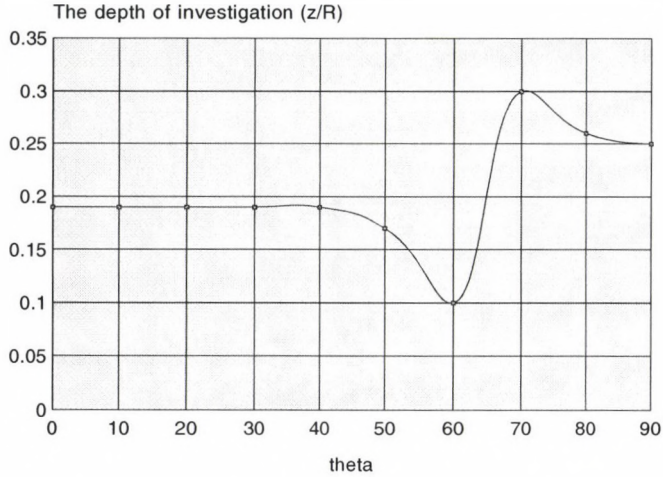


Fig. 8. The depth of investigation of parallel arrays

Table I. The depth of detection for the different arrays (z/R)

ϑ	parallel	perpendicular	radial	azimuthal
0°	|	|	|	|
10°	|	|	|	|
20°	0.195	|	|	|
30°	|	|	|	|
40°	|	0.20	0.19	0.25
50°	0.17	|	|	|
60°	0.10	|	|	|
70°	0.30	|	|	|
80°	0.26	|	|	|
90°	0.25	|	|	|

In case of the parallel arrays the depth of investigation remains nearly constant ($z/R = 0.195$) in a wide interval ($0 \leq \vartheta < 40°$), but between $\vartheta = 40°$ and $\vartheta = 80°$ it has abrupt changes as it can be seen in Fig. 8 and in Table I. $z/R = 0.3$ (at $\vartheta = 70°$) value is the maximum depth of investigation among all studied arrays and $z/R = 0.1$ (at $\vartheta = 60°$) the minimum depth of investigation among all these arrays. It is especially interesting that the distance between these two extreme values in terms of ϑ is only 10°.

As we have it seen, the depth of investigation of the radial arrays and that of the perpendicular arrays are approximately the same as that of the dipole axial array ($z/R \approx 0.195$), and the depth of investigation of the azimuthal arrays ($z/R = 0.25$) is equal to the dipole equatorial array.

On basis of their behaviour in 1D case, we can set all these arrays in two main groups: (1) the group of axial-like arrays and (2) the group equatorial-like arrays (Table II).

Table II. The groups of dipole arrays on the basis of their detection depth in 1D case

	Axial-like arrays	Equatorial-like arrays	Other arrays which do not fit
	– parallel array if $0° \leq \vartheta \leq 40°$ – radial array – perpendicular array	– parallel array if $\vartheta > 80°$ – azimuthal array	– parallel array if $40° < \vartheta \leqq 80°$
Depth of detection (z/R)	0.19–0.20	0.25	0.1–0.3

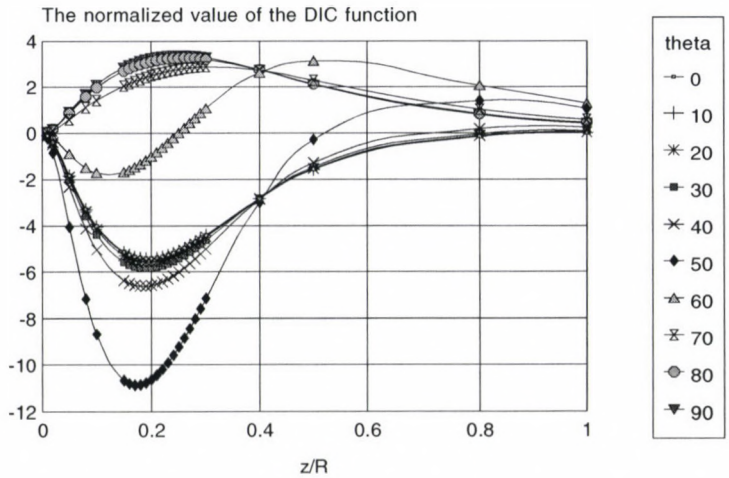


Fig. 9. The value of the DIC function normalized by the value above homogeneous halfspace

At small ϑ values the parallel arrays can be found in the family of axial arrays. At high ϑ values the parallel arrays are already in the family of the equatorial arrays. Their transition is not monotonous. It has under- and overshootings, as it is shown in Fig. 8.

A modification in the interpretation by Roy and Apparao (1971)

I have found a coherent system of investigation depths belonging to different dipole-dipole systems. I should note that there is an important difference between the representation of the DIC curves by Roy and Apparao (1971) and that by this paper. Roy and Apparao (1971) normalised the secondary signals by the homogeneous halfspace value as it can be seen in Fig. 9, too. Their interpretation was based on this representation and this led to the conclusion that: "... for $\vartheta = 54°44'8"$, the parallel dipole would have an infinitely large depth of investigation, and should therefore see deep targets even with very small spacing." (Roy and Apparao 1971).

Some other authors concluded to the same statement. E.g. Fröhlich (1967) did

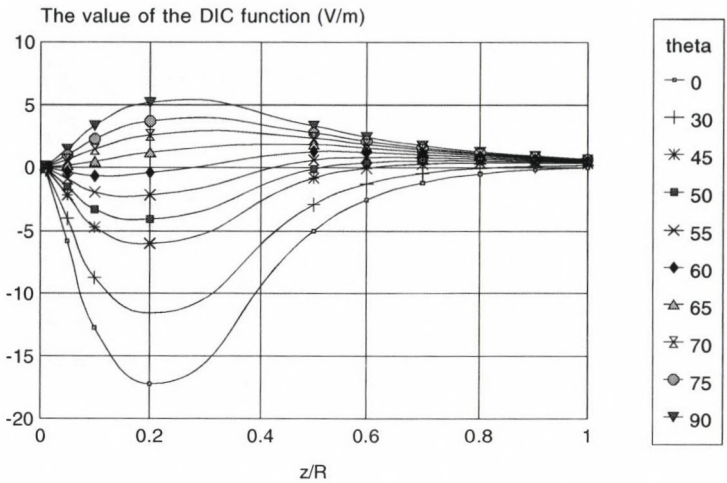


Fig. 10. DIC functions for parallel dipole arrays

not advise the use of this array because "though the depth penetration would be very great, the slightest error in measuring J would require another interpretation curve" (Fröhlich 1967).

I show how this statement comes from the an erroneous interpretation of the DIC functions.

The parameter, which is presented in most cases:

$$\frac{E^{\text{total}}}{E^{\text{hom}}} = \frac{E^{\text{hom}} + E^{\text{inh}}}{E^{\text{hom}}} = 1 + \frac{E^{\text{inh}}}{E^{\text{hom}}} . \tag{5}$$

We should ask: what is the information coming from the thin layer itself? This is the E^{inh}, the field which originates from the thin layer. For what reason we need normalize it by any value? If we want to compare the response above an inhomogeneous area (which contains a thin layer) with the homogeneous halfspace response, we can take simply the difference between them:

$$E^{\text{total}} - E^{\text{hom}} = (E^{\text{hom}} + E^{\text{inh}}) - E^{\text{hom}} = E^{\text{inh}} . \tag{6}$$

The result is E^{inh}, which contains the information about the layer. It is not necessary to divide it by anything, since it leads only to misunderstanding. This was the case by Roy and Apparao (1971): they normalized the response by the zero value belonging to the homogeneous halfspace.

From Fig. 10 it can be seen, that the depth of investigaton of this array is not greater, even smaller ($z/R \sim 0.2$) than that by using most of other dipole arrays. Consequently we cannot claim that this array would have an infinitely great depth of investigation as it was stated by Roy and Apparao (1971) if we are consequent.

Conclusion

The DIC functions were computed for traditional dipole-dipole arrays, as for the dipole parallel, dipole perpendicular, dipole radial and dipole azimuthal arrays. It was found that the depth of investigation — which is the depth where the DIC function has the maximal value — has a quasi constant specific value in case of all of these arrangements except the parallel arrays. Based on the value of this constant, two dipole-dipole array families can be distinguished. The perpendicular and the radial arrays and the parallel array in the $\vartheta \leqq 40°$ domain are the so called axial-like arrays having the same depth of investigation value. The azimuthal array and the parallel array in the $80° < \vartheta \leq 90°$ domain are the equatorial-like arrays. The depth of investigation by using the parallel dipole array changes very quickly and in a wide interval ($z/R = 0.1 - 0.3$), if $40° < \vartheta \leqq 80°$.

It is a second feature that the normalized effect of a thin layer is independent of ϑ in case of all arrays with exception of the parallel array. Furthermore, the parallel arrays change sign even in case of a homogeneous halfspace.

Finally, it was shown that the so-called null arrangement ($\vartheta = 54°44'8"$ parallel array) does not have an infinitely large depth of investigation. Its depth of investigation is even not greater than that of the most of the other dipole arrays. The source of this misinterpretation was a meaningless normalization by a value close to zero.

Acknowledgements

The author thanks László Szarka and Antal Ádám for their constructive comments and Mrs M Fleischhacker for the careful typewriting of the the manuscript.

References

Alpin L M, Berdichevskii M N, Vedrintsev G A, Zagarmistr A M 1966: Dipole methods for measuring Earth conductivity. Consultants Bureau, New York

Fröhlich R K 1967: *Geoexploration*, 5, 195–203.

Knödel K, Krummel H, Lange G 1997: Geophysik. (Handbuch zur Erkundung des Untergroundes von Deponien und Altlasten). Band 3, Springer Verlag, Berlin, Heidelberg, New York, etc.

Roy A, Apparao A 1971: *Geophysics*, 36, 943–959.

Szalai S, Szarka L 1999: An approximative analytical approach for computing geoelectric dipole-dipole response due to a small burried cube. Submitted to Geophysical Prospecting

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 75–80 (2000)

BRITTLE-DUCTILE TRANSITION OF ANISOTROPIC ROCKS DURING THREE-POINT BENDING TEST

K LEDNICZKY[1] and B VÁSÁRHELYI[1]

[Manuscript received November 16, 1998]

Firstly Kármán (1910, 1911), after Griggs (1936) later many other researchers (e.g. Handin and Hager 1957, Paterson 1958, Heard 1960, Byerlee 1968, Edmond and Paterson 1972, Rutter 1972, Gowd and Rummel 1980, Scott and Nielsen 1991, etc.) observed the influence of confining pressure on the behaviour of rocks which were brittle at zero and low confining pressure. At high confining pressure, however, the same rocks may be ductile. That means they may fault or otherwise deform without loss of compressive strength. The amount of ductile deformation and the strength increase progressively with increasing confining pressure until fully ductile deformation occurred with apparent work-hardening.

The aim of this paper is to show the influence of confining pressure during a three-point bending test. An anisotropic gneiss was investigated in 5 different directions of the foliation under 4 different confining pressures. The function of the tensile stress vs. hydrostatic pressure was determined, as well.

Keywords: bending test; brittle-ductile transition; confining pressure; rock anisotropy

1. Introduction

It is widely known that ductility increases with increasing confining pressure, and that the transition from the brittle to ductile state occurs at a certain confining pressure. This phenomena was investigated in conventional triaxial test by many researchers (Orowan 1960, Byerlee 1968, Mogi 1974). They found that in case of the most rocks brittle-ductile transition pressure can be predicted by Orowan's frictional hypothesis, according to which, when the frictional resistance is higher then the shearing strength of the rock, stress drop cannot occur and rocks show a ductile behaviour.

The goal of this research was to examine the mechanical behaviour, specially the transition from brittle to ductile behaviour of rocks with initial crack during three-point the bending. The other aim of this research was to determine the influence of the anisotropy. Therefore anisotropic samples were deformed 0°; 30°; 45°; 60°and 90°degrees from the foliation at confining pressures of 0; 10; 30 and 60 MPa.

[1]Department of Engineering Geology, Technical University of Budapest, Hungary

1217-8977/2000/$ 5.00

Table I. Mechanical properties of the investigated gneiss (Vásárhelyi 1995)

Physico-mechanical properties	Index	Unit	Result
Modulus of elasticity	E1	GPa	41.69
	E2 = E3	GPa	70.05
Poisson's ratio	$\nu_{12} = \nu_{13}$	–	0.166
	$\nu_{21} = \nu_{31}$	–	0.262
	$\nu_{23} = \nu_{32}$	–	0.247
Compressive strength	ν_1	MPa	238
	$\nu_2 = \nu_3$	MPa	267

2. Specimen preparation and test equipment

The tests were carried out on an anisotropic gneiss. This metamorphic rock consists of small grains; during its formation the minerals deform in one direction as can be seen with the unaided eye. Table I shows the mechanical properties of the rock. (Direction 1 is perpendicular to the plane of foliation, directions 2 and 3 are in the plane of foliation and they are perpendicular to each other.)

Extensive specimen preparation is necessary for performing confined three-point bending tests. This includes initial specimen machining, pre-cracking, final machining, instrumenting and jacketing (Vásárhelyi 1997). For the best result the specimens were cut from one block. The experimental measurements were performed at the rock mechanics laboratories of the Technical University of Lorraine in Nancy, France.

Forty notched specimens of gneiss were dry-machined. Each specimen was 25 mm high (h); 15 mm wide (w); and the distance between the two fixed points (L) was always 100 mm. All tests were performed on specimens of equal size with nearly equal crack length a (5 mm long) in the middle of the bottom of the specimen. Fig. 1 shows the dimensions of the samples.

To perform tests under confining pressure, the specimens had to be jacketed under vacuum with a special, flexible material that would cover all the machined surfaces including the crack, but would not enter the crack. All specimens were air-dried and tested at room temperature.

A pressure vessel was specifically designed for confined pressure tests (Fig. 2). The test chamber was 15 cm in diameter and 25 cm high with a working pressure of 100 MPa. The confining pressure of 0, 10, 30 and 60 MPa was supplied and controlled by a servo-controlled intensifier. Light hydraulic oil was used as the pressure medium. The load was measured with a proving ring equipped with an extensometer. This instrument was located between the loading piston and the specimen.

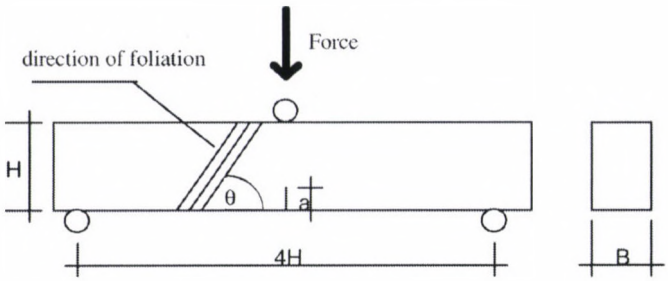


Fig. 1. Schematic sample used for the three-point bending tests

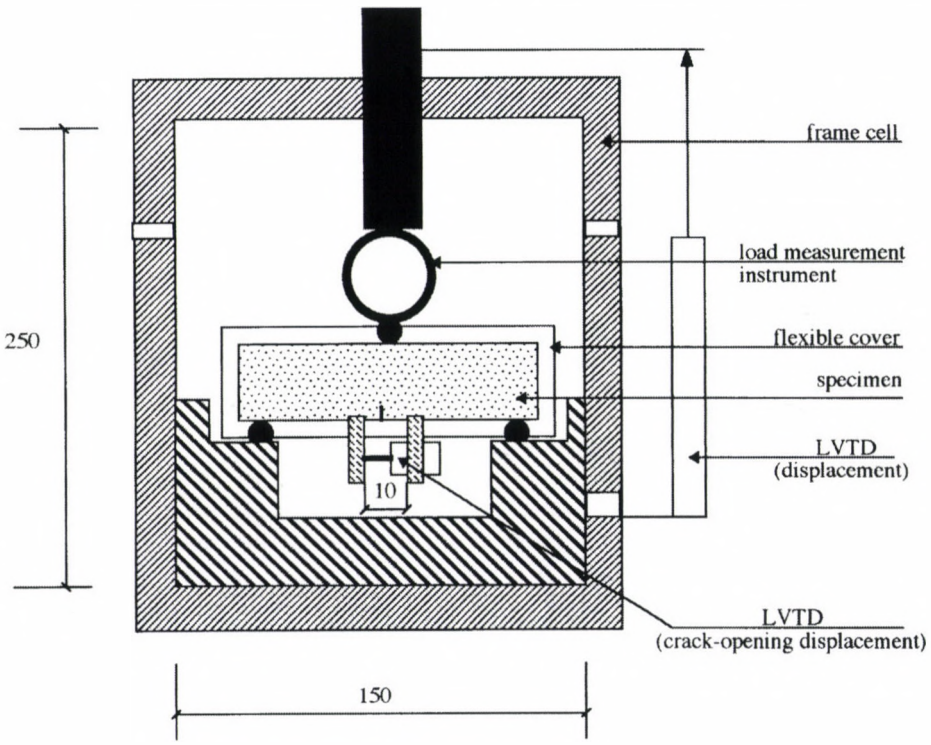


Fig. 2. Test configuration (dimensions in mm)

The load was increased by increasing the displacement at 0.3 μm/s. Vertical displacement and the crack-opening displacement were measured with linear-variable differential transformers (LVTD).

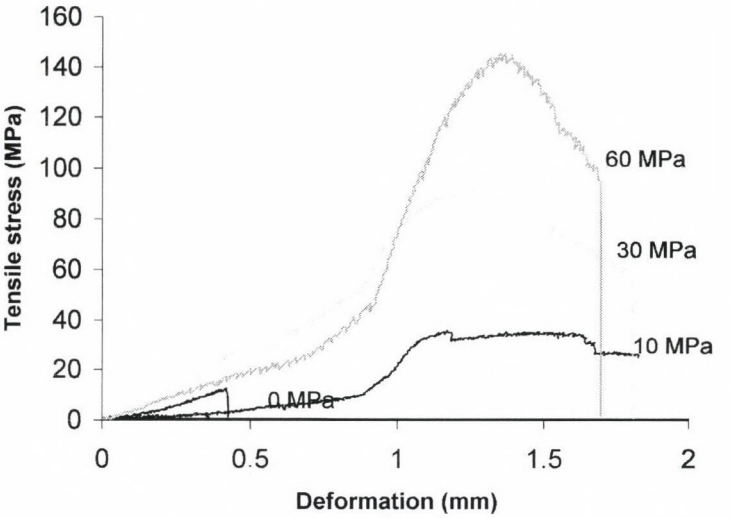


Fig. 3. Stress-displacement curves for the three-point bending tests. Numbers at the end of curves are the values of hydrostatic pressure

3. Experimental results

Four characteristic stress-displacement curves of tests (for 0; 10; 30 and 60 MPa hydrostatic pressure) are presented on Fig. 3. These results are for samples deformed 30 degrees from the foliation. The numbers at the curves are the hydrostatic pressures. There are no significant differences in the curves for samples deformed in different foliation orientations.

At 0 MPa confining pressure the gneiss failed in a brittle way. The amount of ductile deformation and the strength increased with increasing confining pressure (10 MPa) until plastic deformation occurred with apparent work-hardening (30, 60 MPa). It means that the brittle material became ductile with higher confining pressure not only for triaxial tests but for this tensile-bending test, too. The mode of failure of the samples did not change with the increasing hydrostatic pressure; linear fracture formed in all cases (there was no influence of the direction of the anisotropy) starting from the top of the pre-crack and stopping at the point of the applied force.

The connection between the peak-stress at the top of the pre-crack and the hydrostatic pressure was turned out to be linear. The following equation was carried out:

$$\sigma = a + bP \tag{1}$$

where a and b are material constants, P is the hydrostatic pressure in MPa and σ is the tensile stress in MPa at the top of the pre-crack.

According to the test results there are not large differences among the slopes (2.081–2.231) of the lines. The average value and the variance are 2.145 and 0.062, respectively. The direction of anisotropy does not influence the slope of the line but slightly influences the value of a. A possible explanation for the fact that stress

Table II. The material constants of the anisotropic gneiss in five different directions with the variance (according to Eq. 1)

Direction	a	b	Variance
0°	17.22	2.123	5.04
30°	20.33	2.081	3.52
45°	20.78	2.205	1.94
60°	14.28	2.231	1.03
90°	16.65	2.085	3.60

values at 30°and 45°are about 15 percent higher then in other cases is that in this two cases the compressional caused by the confining pressure presses the layers to each other making them stronger while in other cases the tensile stress caused by the bending separates the layers from each other (Vásárhelyi 1995, 1997). Table II consists of the results for the different directions.

4. Conclusion

Forty three-point bending tests were carried out on an anisotropic gneiss in five directions of the foliation (0; 30; 45; 60 and 90 degree) and under four differences hydrostatic pressures (0; 10; 30 and 60 MPa). There was no significant difference between the results for samples deformed in the different directions. The brittle rock became ductile at higher confining pressure. The maximal tensile (peak) stress increased linearly with the hydrostatic pressure.

Acknowledgements

The authors thank both for French Government for supporting this research and for Prof. F Homand. Their acknowledgement both for the Th Cholnoky Foundation and the Hungarian Grant (OTKA F022620) for financial support while this paper was written.

References

Byerlee J D 1968: *J. Geophys. Res.*, 73, (14), 4741–4750.
Edmond J M, Paterson M S 1972: *Int. J. Rock Mech. Min. Sci.*, 9, 161–182.
Gowd T N, Rummel F 1980: *Int. J. Rock Mech. Min. Sci.*, 17, 225–229.
Griggs D T 1936: *J. Geol.*, 44, 541–577.
Handin J, Hager R V 1957: *Bull. Am. Assoc. Petrol. Geologist*, 41, 1–50.
Heard H C 1960: *Rock Deformation, Geol. Soc. Am.*, 79, 193–226.
Kármán T 1910: *Magyar Mérnök- és Építészegylet Közlönye*, 10, 212–226.

Kármán T von 1911: *Verhandl. Deut. Ingr.*, 55, 1749–1757.
Mogi K 1974: *Tectonophysic*, 21, 273–285.
Orawan E 1960: *Rock Deformation, Geol. Soc. Am.*, 79, 325–345.
Paterson M S 1958: *Bull. Geol. Soc. Am.*, 69, 465–476.
Rutter R H 1972: *Int. J. Rock Mech. Min. Sci.*, 9, 183–189.
Scott Th E, Nielsen K C 1991: *J. Geophys. Res.*, 96, (B1), 405–414.
Vásárhelyi B 1995: Etude de l'influence de la pression de confinement et de l'orientation de la foliation sur la propagation des fissures dans un gneiss, Rapport de Stage, INPL-Nancy.
Vásárhelyi B 1997: *Rock Mech. Rock. Eng.*, 30, 59–64.

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 81–85 (2000)

WILHELM FOERSTER AND THE DEVELOPMENT OF SOLAR-TERRESTRIAL PHYSICS

W SCHRÖDER[1]

[Manuscript received March 14, 1998]

Extracts from William Foerster's papers relating to geomagnetism, solar and auroral physics are presented here so as to give an idea about his contribution to the solar-terrestrial physics during the years 1871 to 1906.

Keywords: Foerster W; geomagnetism; solar-terrestrial physics

Topics belonging in our times to geophysics had been studied in the 19th century in the framework of astronomy and geography. Therefore it is not surprising that studies of geomagnetism, auroras, physics of the upper atmosphere etc. were often published in astronomical, geographical and even in physical journals. The cause is simply that geophysics had not been constituted as an independent discipline and correspondingly there were no special geophysical journals. Significant repeated impetuses came especially from astronomy. Wilhelm Foerster, an eminent organiser of science contributed significantly to the development of new methods of research which helped geophysics.

A short curriculum

Wilhelm Julius Foerster was born on December 16, 1832. Having finished school education, he started in 1850 to study mathematics and astronomy in Berlin. He continued studies between 1852 and 1854 in Bonn where his tutor was among others Argelander. He was appointed in 1855 second assistant of the Berlin Astronomical Observatory headed by Encke.

Foerster made study-tours already during his university years in England, France, Austria and Switzerland and established there scientific connections. His thesis dealt with the determination of the polar height in Bonn using the then not yet tested transit instrument in the first vertical plane. At the Berlin Astronomical Observatory he made observations with the twelve feet Frauenhofer-refractor and was especially interested in comets and minor planets. He succeeded together with Lesser in discovering Erato in 1860. Additionally he dealt with problems of astrometry.

He qualified as university lecturer (privat docent) at the Berlin University in 1858, and in 1860 he took over Karl Bruhns' position who accepted an invitation to Leipzig. He dealt substantially with the history of astronomy and his science popularising activity dates from this interval, too. In 1863 he became an associate professor and in 1865 he followed Encke as director of the Observatory. The seat

[1]Hechelstrasse 8, D-28777 Bremen-Rönnebeck, Germany

1217-8977/2000/$ 5.00

of an astronomer in the Academy of Sciences which was vacant after Encke's death was filled by Auwers.

In the years to come Foerster developed wide-ranging activity in science organisation both in national and in international framework. He was not only a leading personality of the Astronomical Society which he had helped to found but also in the introduction of the metric system into Germany, to establish the international latitude service and the Central Bureau for Astronomic Telegrams. He was interested in technical problems, too, and this interest resulted in an early connection with Ernst Abbe. He played important role in the establishment of the Jena Glass Factory (Schott) and of the Imperial Physical and Technical Institute.

Foerster coupled the care for popularising work always to his scientific and science organising activities. We mention here the foundation of the well-known Berlin institution for adult education "Urania" and the "Society of Friends of Astronomy and Cosmic Physics", as well as support to the known populariser of astronomy, Bruno H Bürgel.

There were important discoveries in the field of solar-terrestrial physics during Wilhelm Foerster's life. The Berlin Astronomical Observatory, as well as the later established Potsdam Astrophysical Observatory contributed important results to the solution of many individual problems. He died on January 18, 1921 after a long life full with activity.

Foerster's geophysical interests

Herrmann (1975, 1982) and Oleak (1977) dealt previously with different aspects of Foerster's science organisational activity. In the present connection Foerster's memorandum from the year 1871 is of great importance entitled "Memorandum about the establishment of a solar observatory, i.e. of an institute to observe all processes on the Sun and in its surroundings, together with the complete and regular measurements of all its direct and indirect influences on terrestrial conditions, especially with recording of observations of geomagnetism, earth currents, atmospheric electricity, as well as heat of the air and of the soil and of the atmospheric pressure" and therefore it is quoted here with emphasis.

The interest in solar and geophysical problems had some tradition at the Berlin Astronomical Observatory and Foerster joined this tradition with his plans. Leonhard Euler listed auroral observations as one of the regular duties of the observatory in his program for its establishment. Johann Gottfried Galle, assistant of the institute between 1835 and 1851 contributed to the activities of the Göttinger Magnetischer Verein (led by Gauss and Weber) (Wiederkehr 1964) and carried out studies for the determination of the height of auroras. Therefore it is not surprising that due to this tradition Foerster got early in contact with these problems. Solar-terrestrial physics reached significant results at that time. Hermann Fritz proved in 1862 in Zurich the connection between auroras and solar activity by presenting the parallelism between the frequency of occurrence of sunspots and auroras. Spectroscopic studies by A J Ångström excited great interest. He found the dominant spectral line of the auroras as a yellowish green emission line with a wavelength of

557.7 nm which seemingly could not be brought in connection with any terrestrial material. Later studies led to the discovery of two more red lines with wavelengths 630.0 and 636.6 nm.

Foerster reported about his own auroral studies already in 1870 and 1871 in the Society for Earth Science in Berlin. He allowed that he is "not to be considered as a specialist in this field" (Foerster 1878, p. 105), nevertheless these phenomena have close connections with astrological problems, and therefore his contributions are acceptable. He remarked about the nature of auroras that they are essentially electric discharges which appear both in the lowest and highest air strata. The geographical and geometric distribution and development "is governed by the directing force of geomagnetism similarly to an arbitrary magnet movable which arranges conductors of electric currents and discharges in its vicinity in certain directions which are given with geometric exactness by the relative position of these currents to the poles of the magnet. Thus in few words: 'material of electric light and geomagnetic architecture of the phenomenon' " (Foerster 1878, pp. 105-106).

A very strong aurora was observed in 1852 which was observed even in India, too. Foerster studied this event and wrote "they namely auroras appear in their specific zone quite regularly and their occurrence frequency does not show any dominant influence from cosmic processes and periodicities as they belong to the economy of the inevitable terrestrial compensation" (Foerster 1878, p. 110). "The great terrestrial lights (auroras) which develop in the highest strata of the atmosphere have a clear connection with the intensity and occurrence frequency of certain compensation processes on the solar surface and in its surroundings (ibid., p. 110).

Foerster's notices show that the Berlin observatory conducted regular spectroscopic studies of the night sky and of the green auroral line. Else this was the same project which was reiterated some years later in Göttingen by the first professor in geophysics of the world, by Emil Wiechert. Foerster wrote: "The spectroscope — among others in the hands of Dr Tietjen from the astronomical observatory here — delivered significant contributions to the explanation of the complex of these phenomena. Spectroscopic studies led in recent years to the discovery that on many clear nights when no aurora was visible, more or less significant traces of the same light-lines can be identified which appear with an increased intensity during visually observable auroral phenomena as characteristic peaks of the optical phenomenon" (ibid., p. 112).

Foerster remarked concerning the nature of aurora that unipolar induction mainly supplies the electricity to the compensation taking place in the aurora, he emphasised, nevertheless, that the complex of cosmic-solar influences have to be considered together with it. Similar considerations, first of all concerning the solar causes for electric phenomena of the upper atmosphere, were presented by Siemens and Sirks.

Foerster was interested in the First International Polar Year (1882-1883), in the forerunner of the International Geophysical Year, too. At his time there was a need for world- wide meteorological and geomagnetic observations. Data were especially lacking from the Arctic and Antarctic regions. Neumayer, the founder of the Maritime Observatory Hamburg suggested this international effort. Foerster remarked about it: "The whole effort, to a certain extent the first complete attempt

to deal — at least in a scientific sense — with the Earth, as with a site belonging to the whole mankind is worthy of general participation. We hope that good luck will escort all the efforts, especially the people carrying out observations in polar stations who have to offer high personal devotion" (Foerster 1884).

Foerster participated in planning and elaboration of the research conception of the Göttingen Geomagnetic Observatory. During E Schering's directorship there were different considerations about it which led after his death to the separation of the geomagnetic division from the Astronomical Observatory, later in 1898, to the establishment of a Geophysical Institute headed by E Wiechert.

A further field of study was earth current observations and study of the disturbances in the telegraphic service. At a lecture in the "Electrotechnical Society", Foerster referred to the coming sunspot maximum and emphasised the importance of corresponding activities. Foerster hoped that subsequent studies of the disturbances in the telegraphic service with his participation would help to prepare suitable preventive measures (as e.g. for transcontinental cable connections).

During Foerster's directorate not only Tietjen carried out spectroscopic studies of the night sky, but Otto Jesse, the discoverer of noctilucent clouds, made several experiments to determine the height of the auroras. Galle's and Jesse's height measurements indicated that auroras appear in great heights within the terrestrial atmosphere. Together with the later measurements of noctilucent clouds these proved the existence of the atmosphere up to a height of about 100 km.

Physics of the upper atmosphere

The eruption of the volcano Krakatoa in the Sunda-Street in the year 1883 was not only a gigantic catastrophe, but also it was a first rank astronomical event, too. Subsequent eminent dusk phenomena observed everywhere in the world awoke not only interest of the scientific community; amateur naturalists observed them with the same surprise as specialists. Foerster should have been very much hit by the event, as he wrote: "The year 1883 was surely in a certain extent milestone toward a new activity of the Berlin Astronomical Observatory in the field of meteoritic-astronomic observations, namely by the colossal phenomena which had started by the eruption of the volcano Krakatoa in the Sunda-Street in August 1883" (Foerster 1991).

In the years to come, the Berlin Astronomical Observatory developed an unprecedented research program on the upper atmosphere which led to a basic understanding of the noctilucent clouds. This program included the monitoring of these clouds from several stations (among others from Berlin-Steglitz, Nauen, Rathenow, Frankfurt/Oder and Warnemünde) and the consequent use of photography. Using observations made through several years, some basically new ideas could be developed e.g. on the inner variations of the circulation in the upper atmosphere and on the periodicity of the appearance of noctilucent clouds. The Berlin atmosphere program was a forerunner and pioneer of modern aeronomy projects (Schröder 1982).

Finally it should be remarked that Foerster took pains to include wide groups of the population into scientific research. In addition to the foundation of the men-

tioned Urania and of the Society of Friends of Astronomy and Cosmic Research, he was also active in the publicistic field. The next quotation shows how he imagined the inclusion of wide circles into research: "The participation in such cosmic research, even in the simplest form, brings about an incomparable satisfaction which means perhaps more to laymen and to friends of research than to specialists as with the latter the simple joy and delight becomes paralysed in the intensive stress of the work".

References

Foerster W 1878: *Z. Ges. Erkunde, Berlin*, 8,105–110.

Foerster W 1884: Sammlung populärer astronomischer Mitteilungen. Part II. Dümmler Verlag, Berlin

Foerster W 1906: Von der Erdatmosphäre zum Himmelsraum. Hillger-Verlag, Berlin and Leipzig

Foerster W 1911: Lebenserinnerungen und Lebenshoffnungen. Reimer Verlag, Berlin

Herrmann D B 1975: Geschichte der Astronomie von Herschel bis Hertzsprung. Verlag der Wissenschaften, Berlin

Herrmann D B 1982: Karl Friedrich Zöllner. B G Teubner, Leipzig

Oleak H 1977: Wilhelm Foersters Verdienste um die Wissenschaftsorganisation. In: Sternzeiten II. Akademie-Verlag, Berlin, 35–39.

Schröder W 1982: Disziplingeschichte als wissentschaftliche Selbstreflexion der historischen Wissenschaftsforschung. Eine Darstellung unter Heranziehung von Fallstudien der Wissenschaftsgeschichte der Geophysik. Verlag P Lang, Bern and Frankfurt a. M.

Wiederkehr K-H 1964: *Nachr. Akad. Wiss. Göttingen, II. math.-phys. Kl.*, No. 14

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 87–93 (2000)

ALFRED WEGENER AND THE PHYSICS OF GEOPHYSICAL PHENOMENA IN THE UPPER ATMOSPHERE

W SCHRÖDER[1]

[Manuscript received August 7, 1998]

Alfred Wegener's papers related to noctilucent clouds, aurora (including the so-called geocoronium/geocorona), meteors, and other geophysical phenomena in the Earth's upper atmosphere are presented here so as to give an idea of his contribution to this field made during the years 1900 to 1930.

Keywords: aurora; noctilucent clouds; upper atmospheric physics; Wegener Alfred

1. Introduction

The 120th anniversary of Alfred Wegener's (1880–1930) birth gives us occasion to commemorate his comprehensive research activity. Within it, continental drift not surprisingly occupies a significant position. Nevertheless, when Wegener's scientific achievements are dealt with, it is not to be forgotten that traces of his activity are felt in other areas of geosciences, too. That is why this paper deals with Wegener's contributions to different problems of the physics of the upper atmosphere, as a reconstruction of his work would be uncompleted without consideration of these contributions.

2. Wegener and the upper atmosphere

Alfred Wegener started his activity when an epoch of significant development in physics of the upper atmosphere had just ended. Rudolf Wolf's (1816–1893), and especially Hermann Fritz' (1839–1893) research on solar-terrestrial physics led to notable results in this field (Schröder 1980). Simultaneously the Berlin Project to monitor the night sky was carried out, as initiated by Wilhelm Foerster (1832–1921) which aimed at a monitoring of the green auroral line. The discovery of the noctilucent clouds in the year 1885 and the Berlin Research Project of the upper atmosphere initiated by Otto Jesse (1838–1901) brought about precise height determinations of the uppermost layers of the atmosphere and thus the development of the special field of geophysics called "aeronomy" was prepared. Further free balloon flights in Berlin during the years 1891 to 1900 (Assmann 1845-1918, Berson 1859-1942 and Süring 1866–1950) should also be mentioned which contributed to the basis of aerology and led among others in 1902 to the discovery of the stratosphere by Assmann and Teisserence de Bort (1855–1913). It is by no means surprising that

[1]Hechelstrasse 8, D-28777 Bremen-Rönnebeck, Germany

1217-8977/2000/$ 5.00

these developments in the geophysics and meteorology increased the general interest in "Cosmic Physics" (Hellmann 1903)[2] and Wegener was also influenced by them. He was interested in such problems as early as during his time as privat-docent. His wife, Else Wegener wrote about this: "Recently it interested me (A Wegener) the problem that "Cosmic Physics" has more legitimacy as a topic at the universities than a speciality like Meteorology. In textbooks of physics, Cosmic Physics is treated in significant details as example. At the universities this is not taken into account as no time remained there for such excursions. It follows the necessity to present there Cosmic Physics in special lectures as supplement to experimental physics" (Wegener E 1960).

Alfred Wegener obtained habilitation at the Marburg University in 1909 in Practical Astronomy, Meteorology and Cosmic Physics and took over the management of the small astronomical observatory there. It is understandable why Wegener told the previously cited sentences on the importance of Cosmic Physics: Wegener had studied meteorology but he did not find the appropriate application in this field. At the Marburg University Wegener lectured on Astronomy, he could, however, only attract a small number of students. This is connected very likely with the fact that his topics (aerology and thermodynamics of the atmosphere) were little known and were only slowly introduced into meteorology.

During his time in Marburg, he got acquainted with the Heidelberg astronomer Max Wolf (1863–1932), too. Wegener asked him already in 1909 to lend him slides for his lectures and for the establishment of a collection of pictures in astronomy[3]. It is to be mentioned that Wegener discussed several times his meteor studies with Wolf.

The "Project for a Systematic Observation of Atmospheric Electricity" carried out by Wiechert in Göttingen and recent determinations of auroral heights initiated by Wiechert and carried out by Robitsch (1917) and K Wegener were of decisive importance for the following development of Wegener's activity. The connections between Wiechert and Wegener are interesting from the point of view of history of science, too, they discussed among others the problems of air seismics (Wiechert 1926, Wegener 1925)[4].

Wegener's interest in atmospheric electricity and in similar atmospheric phenomena appeared in the plans of the inland ice expedition to Greenland in the years 1930 and 1931[5], too. Meteorological projects included studies of glacial anticyclones, radiation measurements, observation and photographic study of air mirroring, measurement of the polarisation of the sky light and halo observations.

[2]G Hellmann (1854–1939) who is the author of significant contributions to the history of meteorology remarked that since the publication of Alexander von Humboldt's "Kosmos", "'Cosmic Physics' has meant all applications of physical laws to problems of the sky and earth sciences in a closed complete system"

[3]Correspondence between Max Wolf and Alfred Wegener; I thank Professor Wolf, son of Max Wolf for the permission to use these letters

[4]Wiechert and Wegener exchanged ideas on this topic and Wegener sent his paper before publication to Wiechert for reviewing.

[5]Plan of a German inland-expedition to Gronland 1930/31 and a pre-expedition 1929 headed by Professor Alfred Wegener, manuscript, 30 pp

Table I.

Year	Topic studied
1905	Development of cosmic ideas in philosophy
1906	Studies of air surges
1910	Layers in the atmosphere; nature of the uppermost atmospheric layers
1911	Studies of the nature of the uppermost atmospheric layers; publication of the book "Thermodynamics of the Atmosphere"
1912	Research of the uppermost atmospheric layers
1915	Atmospheric lunar tides; colour changes of meteors
1917	Detonating meteor of April 3, 1916
1918	Colour change of big meteors
1919	Air resistivity against meteors; upward collapse theory of lunar craters
1920	Upward collapse theory of lunar craters
1921	Origin of lunar craters
1924	Second edition of "Thermodynamics of the Atmosphere"
1925	Temperature of the uppermost layers of the atmosphere
1926	Temperature of the uppermost layers of the atmosphere: discussion of this topic with Lindemann and Dobson
1927	Initial and final heights of big meteors
1928	Acoustics of the atmosphere
1935	Lectures on the thermodynamics of the atmosphere, edited by Kurd Wegener

The wide scope of his studies is summarised in Table I.

Considering the wide range of activities, including a big meteorological part, Wegener's relation to meteorology gets clear when he wrote: "Meteorology, including meteorological optics and dragon flights, is the field in which I work best and with greatest joy" (Wegener 1907).

2.1 Noctilucent clouds

The story "of noctilucent clouds (can be) paradigmatically understood for the history of the cosmic- physical research", as Treder (1975) put it as a correct interpretation of the situation.

The discovery of the noctilucent clouds in 1885 offered an unexpected possibility to get indirect information on the upper atmosphere in a way never expected before. Schiaparelli, the Italian astronomer expressed as late as in 1871 his concern in his book on "Shooting Stars" that mankind would never be able to reach these heights; correspondingly no exact information would became available about this region of the atmosphere.

Meanwhile in the scope of the Berlin Research Project on the Physics of the Upper Atmosphere a comprehensive photographic monitoring of the upper atmosphere was carried out, thus Jesse (1896) could present quite exact height determinations.

As noctilucent clouds appeared soon after the eruption of the volcano Krakatau in the Sunda Street, initially a causal connection was supposed between the two phenomena. Nevertheless, it was soon observed that the frequency of appearance of the noctilucent clouds did not decrease even after the fall-out of the volcanic material was recorded. Wegener early supposed the existence of a layer boundary

in the height where the "Krakatoa-clouds" appeared, namely as they ended there their rise. In his book "Thermodynamics of the Atmosphere" (1911a) he treated noctilucent clouds in many details.

Wegener correctly reminded to the fact that in the year when his book was published (namely in 1911) the fall-out of volcanic material from great heights ceased; therefore volcanic material could not be causally connected to the appearance of noctilucent clouds. Wegener supposed, however, that in connection with the Krakatau-eruption tremendous quantities of water vapour was shot into the upper atmosphere. As this vapour was supposed to contain a corresponding supply of free oxygen, therefore it could break through the isothermal layers of the stratosphere upward into the hydrogen sphere, and thus a nearly horizontal spread became possible in the upper layers of the atmosphere. As the quantity of oxygen only slowly decreased in the high region of the atmosphere, a possible source of the origin of noctilucent clouds was given.

Nevertheless, Wegener suspected that noctilucent clouds consisted of local condensation phenomena in these extended layers of water vapour; thus they would not differ essentially from "ordinary" cirrus-clouds.

Concerning the life span of noctilucent cloud fields Wegener surmised that they have to be of short duration. This is no more correct as the space-time extension of noctilucent clouds is quite big and it can be as long as several hours (Schröder 1974). In connection with the origin of noctilucent clouds, Wegener expressed a further interesting idea: he argued namely that the adiabatic rise of the layer boundary had to result in their appearance (Wegener 1911a). Even if recent studies of the noctilucent clouds, as e.g. by Chapman and Kendall (1965) are taken into account, this early idea by Wegener is remarkable.

Wegener discussed noctilucent clouds again in a book published in 1935 after his death, in "Lectures on the Thermodynamics of the Atmosphere". He repeated here his idea that they should be local condensation phenomena in extended layers of water vapour.

Wegener discussed the wave-like phenomena observed in noctilucent clouds, too. This discussion is connected to his earlier research of air waves. He raised the question whether e.g. cyclones can be supposed to be Helmholtzian air surges (Wegener 1921). Bjerknes proved that this putting of the question was false (Ertel 1938). It was also found that Wegener's discussion of the wavelengths of the Helmholtzian surges which were computed by him much bigger than they really were is not correct. Already Haurwitz (1931) supposed that the cause of discrepancy between theory and observation lies the omission of the compressibility. Nevertheless, Wegener's treatment of the air surges was an interesting analogy to the formation of waves and surges in noctilucent clouds.

Nevertheless, Wegener's contribution has remained valuable as he supposed — similarly to Jesse before him — the noctilucent clouds to be physical manifestations of certain processes in the upper atmosphere and correspondingly he included them into the study of the upper atmosphere. Moreover, his "ice hypothesis" is a possibility for the explanation of the noctilucent clouds, as this idea of his proved to be a fruitful starting-point for further discussions (Schröder 1975).

2.2 Aurora

In order to show the importance of Wegener's contribution to the research of auroras, some facts are to be listed here. On the occasion of a prominent solar-terrestrial phenomenon in 1859, Foerster (1878) remarked that this implied "a striking case of the influence of condition of geomagnetism, too, by an extremely sudden and intensive eruption of hot gases from the solar surface".

This remark by Foerster deserves attention as it shows that he had to be very early aware of the complex connection Sun-Earth. It is moreover important from the point of view of science history that Foerster's institute, the Berlin Astronomical Observatory carried out from 1879 on systematic spectroscopic studies of the night sky. Foerster (1878) remarked that "in many clear nights when no aurora can be seen, more or less obvious traces of the same light-lines in the whole sky can be found which are present with a higher intensity in auroral phenomena, and are thus visible for the naked eye as characteristic form of the optical phenomenon".

Considering that the first publication by Ångström on auroral spectra was published in 1869, the time distance to the Berlin Project is surprisingly short. Ångström discovered during his observations the so-called auroral line, 5567 A.

In connection with these studies concerning the nature of the aurora which had been considered only from the mid-17th century on as manifestation of a physical process, Wegener's studies even gain importance (Schröder 1979). Wegener remarked that the line λ:5577 was especially strong in the case of the highest auroras. In order to explain the green auroral line he supposed that it is due to the existence of the gas geocoronium (from geocorona = uppermost layers of the terrestrial atmosphere). Wegener (1911a) added that by this supposition it would be proven that this line corresponds to the spectrum of an unknown gas which lies in a wide extension above the hydrogen sphere.

Wegener concluded further ideas from the existence of the geocoronium supposed by him; geocoronium should be visible during solar eclipses and should have such a low inertia that comets could cross the solar corona without observable resistance. Similarly meteors can cross the geocoronium zone of the terrestrial atmosphere without observable braking. It followed for the structure of the terrestrial atmosphere that no hydrogen sphere exists in the strict sense of the word, what would explain why hydrogen lines are never the brightest in auroral spectra. If one takes into account following Wegener that geocoronium disappears only below 60 km, it can be explained why auroras appeared seldom or even never without the main line. Wegener remarked further that the fact that hydrogen lines are never the only visible ones in the auroral spectrum, hints at the relative thinness of the layer with this gas: further the fact that hydrogen lines belong always to the weaker ones shows that the maximum content is considerably below 100 percent (Wegener 1911b).

Even if the appearance of a series of unknown spectral lines led to the supposition of a light, unknown gaseous element — geocoronium — and this conception became outworn by Vegard's (1931), Harang's and others studies, Wegener's idea to explain the unknown auroral spectral lines with the supposition of the hypothetical geocoronium, was at least heuristically a remarkable contribution in the direction of a profound physical explanation of the upper atmosphere.

Wegener's contributions both to the noctilucent clouds as indicators of processes in the upper atmosphere and to auroras are significant in context of the history of science as early experiments to explain the terrestrial atmosphere from a profound physical point of view (Ertel 1938).

2.3 Further studies

It is shown in Table I that Wegener collected and arranged meteor observations, too. The prominent fire-ball from April 3, 1916 should be especially mentioned for which Wegener (1917) collected so numerous single data that he could reconstruct its flight path and due to it, the meteor was found.

Studies which did not belong directly to physics of the upper atmosphere, as e.g. studies of wind- and water-hoses, halo phenomena and generally meteorological optics are parts of Wegener's activity, too.

Wegener's studies on meteors are to be considered in connection with his ideas on the origin of lunar craters. He carried out experiments with concrete powder to find facts for the genesis of these craters. Wegener explained the majority of the lunar craters by the impact of meteors or of similar bodies. His ideas — being not very different from our present knowledge — were forgotten for a time similarly to his continental drift theory and are not mentioned in recent studies.

3. Concluding remarks

Wegener's contribution belong to a series of studies which tried early to explain atmospheric phenomena with known laws of physics. For the physics of the upper atmosphere, as well as for aeronomy his contributions are building stones of our modern knowledge about a comprehensive physical interpretation of the upper atmosphere. The increase of knowledge in solar-terrestrial physics (Wolf, Fritz, Ångström), in physics of the upper atmosphere (v. Bezold, Kiessling, Jesse, Foerster, Birkeland, Störmer) as well as in meteorology, especially due to Norwegian meteorologists around Bjerknes, is the frame within which Alfred Wegener followed the same goal, namely the physical explanation of the observed and recorded phenomena. Wegener belonged therefore to researchers who contributed to the development of geo- and cosmoscience and whose heritage should be conserved.

Acknowledgements

I thank Professor Treder for the initiative to this study. Thanks are due to libraries in Göttingen, Marburg, München, Heidelberg, Graz and Oslo for help and permission to use their material.

References

Chapman S, Kendall P C 1965: *Q.J.R. Meteorol. Soc.*, 91, 115.
Ertel H 1938: Methoden und Probleme der dynamischen Meteorologie, Springer, Berlin
Foerster W 1878: *Z. Ges. f. Erdkunde zu Berlin*, 8, 105.
Haurwitz B 1931: Veröff. Geophys. Inst. Univ. Leipzig, II. Serie, Heft 1
Hellmann G 1903: *Meteorol. Z.*, 20, 527.
Jesse O 1896: *Astron. Nachrichten*, 149, 101.
Robitzsch M 1917: Die Struktur des Polarlichtes während des Winters 1912/13, Vieweg, Braunschweig
Schröder W 1974: *Meteorol. Rundschau*, 27, 61.
Schröder W 1975: Entwicklungsphasen der Erforschung der Leuchtenden Nachtwolken, Akademie-Verlag, Berlin
Schröder W 1979: *Transact. American Geophys. Union*, 60, 1035.
Schröder W 1980: Wetter und Leben, 32, 26.
Treder H J 1975: Vorwort zu: Schröder W: Entwicklungsphasen der Erforschung der Leuchtenden Nachtwolken, Akademie-Verlag, Berlin
Vegard I 1931: *Gerlands Beitr. Geophys.*, 32, 288.
Wegener A 1911a: Vorl. Mitt. S.-B. d. Ges. z. Beförder. d. ges. Naturwiss zu Marburg, Nr. 1 and Gerl. Beitr. Geophys, (1912), 11, 194 and Thermodynamik der Atmosphäre, Leipzig
Wegener A 1911b: *Aberhalden, Fortschr. d. naturwiss. Forschung*, 3, 1.
Wegener A 1917: *Schriften d. Ges. z. Beförder. d. ges. Naturwiss. zu Marburg*, 14, 1.
Wegener A 1921: *Meteorol. Z.*, 38, 300.
Wegener A 1925: *Meteorol. Z.*, 42.
Wegener E 1960: Alfred Wegener, Tagebücher, Briefe, Erinnerungen. Brockhaus Verlag, Wiesbaden
Wegener A 1980: Tagebuch der Danmark, 1. Februar 1907, in: Schwarzbach M, Alfred Wegener und die Drift der Kontinente, Wiss. Verlagsanstalt, Stuttgart, p. 62
Wiechert E 1925: *Nachr. Ges. d. Wiss. Göttingen, Math.-phys. Kl.*, 40 pp.
Wiechert E 1926: *Meteorol. Z.*, 43, 81

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 95–98 (2000)

AURORA AND NOCTILUCENT CLOUDS

W SCHRÖDER[1]

[Manuscript received October 2, 1998]

The problem of coupling of auroras with noctilucent clouds (abbreviated NLC) at mid-latitudes is discussed. Two cases of the simultaneous occurrence of auroras and NLC (July 1964) are described. No unusual changes were observed in NLC after onset of aurora.

Keywords: aurora; dynamic processes; mesopause; noctilucent clouds; temperature profile

1. Introduction

Most geophysical phenomena are dependent in some measure on geographic/geomagnetic latitude, concerning both the frequency of local effects and the effect of world-wide disturbances. Many problems of the geomagnetic activity are divided into problems of: auroral zone and polar cap, mid- latitudes and equatorial zone (with a significant contribution of the ring current). Significant differences are to supposed concerning the effects of the phenomena in the different zones; these differences, however, do not refer to the physical mechanism and its corresponding mathematical treatment.

It follows from Störmer's theory the impulse of particles being necessary to reach a certain point on the Earth's surface. This magnetic rigidity is expressed (Pfotzer 1956) as

$$P = \frac{K}{R^2} \cos^4 \Phi \tag{1}$$

where P is the impulse for an elementary charge, then $K = 6,025.10^8/Z$ (GeV/c) [km^2] with Z, the number of elementary charges. of a particle, R the distance of the point from the magnetic centre and Φ is its geomagnetic latitude referred to this centre.

The transformation of the geographic co-ordinates to the geomagnetic ones is carried out using the following equations:

$$\sin \Phi = \cos \varphi \cos \varphi_0 \cos(\lambda - \lambda_0) + \sin \varphi \sin \varphi_0 \tag{2}$$

$$\cos \Lambda = \frac{\cos \varphi \sin \varphi_0 \cos(\lambda - \lambda_0) - \sin \varphi \cos \varphi_0}{\cos \Phi} \tag{3}$$

where φ and λ are the geographic latitude and longitude of the actual point.

[1]Hechelstrasse 8, D-28777 Bremen-Rönnebeck, Germany

1217-8977/2000/$ 5.00

2. Observations

There are several cases of the simultaneous appearance of NLC and auroras (Pfotzer 1956, Villmann 1962, Byrne 1964, Haurwitz and Fogle 1966). Paton and Villmann published photos, too, where the simultaneous appearance of both phenomena can be observed.

The problem of a connection between geomagnetic activity (geomagnetic storms) and NLC was especially studied by Villmann (1962). His results, however, especially concerning the appearance of ssc, have a preliminary character.

Paton reported that colours changed during a NLC-phenomenon. It follows from the American records that "the general effects observed were a reduction of the NLC intensity and extent and a transformation of their well-ordered structure and formless veils. On two occasions, the NLC vanished completely within one hour after onset of the aurora" (Fogle 1965, p. 306).

Two observations made in Rönnebeck ($\varphi = 53.2°$ N) in 1963 have shown that when aurora and NLC appeared simultaneously, that during the later no change in the former was observable.

Concerning visual observations, the following remark is to be taken into account: as compared to NLC, aurora is a more conspicuous phenomenon. The observer's attentiveness may change; that means that the level of the threshold of stimulation may oscillate. As a consequence it follows that the actual stimulation can escape observation. If and how effectively such an effect exists when morphologically different phenomena, such as NLC and aurora are observed and therefore changes in the luminosity are erroneously observed, needs experimental confirmation (concerning the problem of attentiveness see e.g. Hofstätter ed. 1967, pp. 33-39).

As it was remarked in the introduction it is necessary to take into account in the interpretation of such observations the different geographical and geomagnetic, resp. latitudes of the site of the observation. It is therefore important that the observations in America were carried out in the auroral zone and the English observation was also at a higher latitude than hat of Rönnebeck. Further the different levels of the geomagnetic disturbances accompanying a certain event are stronger in the auroral zone that at mid-latitudes.

It is known from the auroral statistics that auroras appear at mid-latitudes (as e.g. at $\varphi = 54°$) appear much less frequently during solar minimum conditions. Moreover, important auroras appear only connected to strong solar flares, as it was observed e.g. during the International Geophysical Year, too.

3. Temperature variations at the mesopause

Temperature measurements in auroras were only made up to now at higher latitudes. The result was that temperature was in auroras of the type A higher then else. Other measurements indicated that the temperatures are significantly different (about 3500 K) during nights with aurora from those observed during nights without auroras ($\lambda = 6300$ OI $= 1200$ K).

NLC appear exclusively in the height range 80 to 85 km. As the present problem

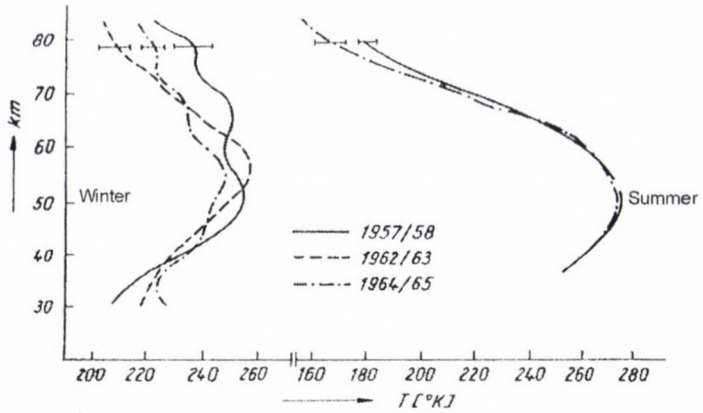


Fig. 1. Changes of the average temperature profile at Fort Churchill (59° N) during solar maximum and solar minimum (after Groves 1968)

refers exclusively to this low altitude, (integrated) spectral measurements may be problematic. The whole question is not yet solved.

Some studies were published on the distribution of temperature at the mesopause and higher as depending on the solar-geomagnetic activity. Truttse (1969) reported on the response of the higher atmosphere on geomagnetic storms accompanied by auroras. According to him, there may be connections between the density of the atmosphere and solar-geomagnetic phenomena. Nordberg (mentioned by Haurwitz and Fogle 1966) carried out temperature measurements in the winter mesosphere in the years 1957/58 and 1962/64, He found that temperatures were significantly higher during the former period than later.

According to auroral observations, pulsating auroras reach a lowest height of about 90 km (Cresswell 1969). A local heating of the mesosphere in situ is hardly possible, it is much more likely that the developing hydromagnetic waves cause an increase of the temperature (Fig. 1).

The temperature distribution in different years was analysed by Groves (1968). According to him, the data from Fort Churchill indicate a temperature of about 237 K for the height of 80 km in the years 1957/58; the average temperature was in the years 1962/63, however, 208 K, then for the years 1964/65 temperature raised again to about 220 K. In this context Groves referred to meteor observations, too; the Perseid observations confirmed a significant increase of the frequency during solar minimum.

London (1965, p.199) wrote about the heating of the mesosphere during the appearance of auroras:

"1. Particles coming in and losing their kinetic energy - this reaction would be a delayed one. 2. The UV radiation 1500 to 3000 A) might affect ozone concentrations at about 80 km. Heating at this level is very sensitive to ozone concentration. This would produce an immediate reaction".

Leovy (1965, p. 39) discussed photomechanical processes possible in heights of 80 to 120 km and hinted at the following differences:

"These are: (1) absorption of solar radiation by ozone, (2) absorption of solar radiation by molecular oxygen, (3) fluctuations of optical depth for molecular oxygen absorption and (4) heat released by oxygen recombination".

4. Discussion

The present observations do not unambiguously settle the problem of the connection between appearance of auroras and heating of the mesosphere. If this would be true and if the appearance of NLC would depend on the temperature profile, then morphological changes would be observable in them. American observations hint at such a connection.

No reports were available from mid-latitudes on this connection, and the two Rönnebeck observations from the year 1963 did not confirm variations. This fact confirms the previously mentioned difference that the possible changes in NLC-s depend on the geographic (geomagnetic) latitude of the site of the observation, on the level of the solar-geomagnetic activity and on the type (e.g. p2) of the aurora. Nevertheless, the probability of expressly dynamic variations of NLC-s is significantly lower than e.g. for latitudes around 65° N.

Nevertheless the number of reports of simultaneous appearance of NLC and auroras is very low; the necessity of a simultaneous monitoring both events is the more pressing, if both phenomena are visible.

References

Akasofu S L, Chapman S, Meinel A B 1966: The Aurora, in: Handbuch der Physik, Vol. 49/1, Flügge, Heidelberg/New York

Byrne F D 1969: *Meteorol. Mag.*, 93, 121.

Cresswell G R 1969: *J. Atm. Terr. Phys.*, 31, 179.

Groves G V 1968: *Space Res.*, 8, 741.

Haurwitz B, Fogle B 1966: *Space Sci. Rev.*, 6, 279.

Leovy C 1965: *NCAR Techn. Note*, TN-8, 39.

London J 1965: *NCAR Techn. Note*, TN-8, 199.

Paton J 1964: *Meteorol. Mag.*, 93, 161.

Pfotzer G 1956: Mitt. Max Planck Inst. Physik der Stratosph., No. 9 Psychologie, ed. Hofstätter, P.R., Frankfurt/M, (1967), pp. 33-39.

Truttse Y L 1969: Lecture, Twelfth Plenary Meeting COSPAR, Prague

Villmann C I 1962: Trudy soveshchaniya po serebrystim oblakam, Tallin, pp. 151-155.

Acta Geod. Geoph. Hung., Vol. 35(1), pp. 99–103 (2000)

Book reviews

M BURŠA, J KOSTELECKY: *Space Geodesy and Space Geodynamics.* Published by the Ministry of Defence, Czech Republic, 1999. 459 pp.

This book is the latest among the great number of monographs published on the use of space techniques for geodynamics in the last three decades. The book presents the foundations of geodetic space methods and the most recent results. Because the limited extent of the book the actual problems and task of space geodesy are treated in sufficient detail. The outline is presented on the principles of space geodynamics which is closely related to space geodesy. The authors have tried to emphasize the parts of space geodesy which have not been dealt with in more details in the books published earlier, such as the genesis of geopotential models up to now. On the contrary, the observation techniques, such as GPS (Global Positioning System), are treated to a limited extent because sufficiently specialized literature are available.

The text of the book is arranged into seven chapters and three supplements. In the first chapter the definition of the subjects as well as the short outline of the historical evolution of the research are summarized as pre-satellite and satellite era. In the second chapter the coordinate systems are introduced. The third chapter is devoted to the equations of motion of a satellites and to their solution. In the fourth chapter the theory of gravitational and gravity potential at points of the Earths surface and the outer space is given. In the fifth chapter the motion of an artificial satellites in the gravitational field of the Earth, Moon and Sun system is evaluated. The sixth chapter discusses the methods and achievements being solved by the means of space geodetic techniques. In the seventh chapter the comprehensive description of a principles of space geodynamics are given.

The summary of the most recent results, the numerous tables containing the relevant parameters of geodynamics and the outline of the actual problems make this book excellent for scientists and students as well.

L Bányai

C C J M TIBERIUS: *Recursive data processing for kinematic GPS surveying.* NCG KNAW Netherlands Geodetic Commission, Publication on Geodesy, 45, 1998

This book contains the results of PhD studies carried out at the Faculty of Civil Engineering and Geosciences of the Delft University of Technology, Department of Mathematical Geodesy and Positioning, under the supervision of Prof. P Teunissen.

Based on a through theoretical investigation, a new procedure for the recursive data processing for kinematic GPS surveying is derived and analysed in different practical applications.

The book is divided into six chapters so that the first five can be handled as independent texts, as well, supplied with their own summaries, references and appendices.

The functional and stochastic models of the estimation principle together with integer estimation and hypothesis testing are given in the first Chapter as the review of mathe-

matical geodesy. In Chapter two the concept of recursive data processing is summarised introducing the estimation, quality control, precision and reliability as well as the mechanisation aspects of the Square Root Information Filter technique (SRIF). The concept and the different mathematical models of GPS surveying are discussed and summarised in the third Chapter. The recursive data processing for kinematic GPS applications is implemented in Chapter four that is based on the SRIF technique using the Least-squares AMBiguity Decorrelation Adjustment method (LAMBDA) and the Detection, Identification and Adaptation testing procedure (DIA) developed at the "Delft School". In the fifth Chapter three different data sets are analysed to deeply investigate the quality of the geometric information obtained by the developed kinematic GPS procedure and different receiver types. In the last Chapter the results are summarised.

In conclusion this book is very useful for those who are interested in the theoretical and practical background of GPS data processing especially in the kinematic applications.

L Bányai

P J DE JONGE: *A processing strategy for the application of the GPS in networks.* NCG KNAW Netherlands Geodetic Commission, Publication on Geodesy, 46, 1998

This book is a result of PhD studies conducted by the group Mathematical Geodesy and Positioning of the Faculty of Civil Engineering and Geosciences of the Delft University of Technology.

The processing strategy of the GPS observations is derived using the undifferenced approach, which is a unique concept comparing to the world-wide used double-differenced concept. The rank defects of the undifferenced approach are treated by the properly chosen S-basis. An algorithm for resolving the rank defect due to the GPS carrier phase ambiguities and the S-basis providing the results identical with the double-differenced approach are derived as well.

The book is built up of six chapters and four appendices.

After a short introduction a functional model and the error sources of the GPS observables are summarised. In Chapter three the undifferenced model is introduced and discussed in details. The fourth Chapter is devoted to the application of multiple GPS observable types, where the different linear combinations are used to handle mainly the ionospheric delays. The integer ambiguity estimation strategy called as LAMBDA (Least-squares AMBiguity Decorrelation Adjustment) method is described in Chapter five. Finally, in the last Chapter the LAMBDA and the method used in the GAMIT software of the Massachusetts Institute of Technology are applied and compared using an observations from a regional network.

In conclusion this book is very useful for the software designer and those who are interested in the details of high precision relative static GPS data processing.

L Bányai

A MORECKI, G BIANCHI, M WOJTYRA eds: *Theory and Practice of Robots and Manipulators.* Proceedings of the Twelfth CISM-IFTOMM Symposium. CISM Courses and Lectures, No. 405. Springer-Verlag, Wien, New York, 1998, pp. 461, 295 illustrations, ISBN 3-211-83143-6

The book contains lectures about the recent results of robotics. The main topics are the following:

Chapter I: mechanics, experimental tests and computer simulation of the mechanical structure of robots.

Chapter II: kinematics and motion control of robots.

Chapter III: overview about programming and software architecture of advanced robots.

Chapter IV: synthesis and design of robot hands and manipulators.

Chapter V: some examples for the application of robots.

Chapter VI: theory and design of robots with legged locomotion.

Chapter VII: some special applications in biomechanics.

Some parts of the book, especially the theoretical ones can be used at the development of measuring robots in engineering geodesy, too.

Gy Mentes

W SCHRÖDER: *Emil Wiechert – Physiker – Geophysiker – Wissenschaftsorganisator.* History Commission of the German Geophysical Society and Interdivisional Commission on History of IAGA (European Section) - Mitteilungen des Arbeitskreises Geschichte der Geophysik, 19. Jahrg., Heft 1–2, Science Edition, IDC History and History Commission DGG (2000), 216 pp, 20 USD

This book contains reprints of selected papers of the author together with several original papers (in English, on Wiechert and the foundation of geophysics, and on Wiecherts importance for the development of geophysics as an exact science) and documents connected with Wiechert (photos of the staff of his institute and of his masters and friends, as well as buildings and instruments connected with him). The reprints deal mostly with problems of the then "new" physics, especially with scientists like Lorentz, Sommerfeld, Einstein, as well as with the Scientific Society in Göttingen. Wiechert was the first professor of geophysics of the world, he worked especially in the field of seismology, organized the International Seismological Cooperation. Nevertheless, he also took interest in other geophysical phenomena, as e.g. atmospheric electricity, aurora etc. The book presents him as a renowned scientist who was hesitant to accept relativity at once, defended the ether theory for a rather long time, but nevertheless was always ready to help other physicists in their efforts. Thus he also promoted the organization of scientific expeditions from Germany to Northern polar regions, to Spitzbergen, to Iceland and also the establishment of geophysical stations in Germany and in the former German colonies, first of all in Samoa, the Apia observatory which played an important role in the development of several branches of geophysics.

For the Hungarian reader, repeated references to contemporary and later Hungarian scientists, especially to Eötvös and Réthly are of special interest. Eötvös and Wiechert could be acquainted with each other, as they spent both some time in Königsberg, and Einstein asked for Eötvös opinion about him, too.

This collection is an interesting reading for all those interested in the history of physics, especially of geophysics.

J Verő

W SCHRÖDER: *Aurora in Time (Das Polarlicht).* Science Edition (IDCH - IAGA), Reprint of das Phänomen des Polarlichts, 2000, 172 pp, including a color plate, 20 USD

The book consists of two parts: the first is a short 11-page summary of the second part in English, the second is a reprint of a longer history of aurora and auroral research, containing many early representations of auroras, scientists from the history of auroral research, historic photos etc.

The author discusses the development of our knowledge about auroras in four main phases. The first lasted till 1716, till the end of the so-called Maunder minimum of very low solar and auroral activities. In the first part of this period aurora was not identified as an independent phenomenon, it belonged to a group of miraculous celestial phenomena which were supposed to hint at imminent disasters - wars. In Late Medieval times, auroras were described in one-page prints distributed like newspapers today – the book contains copies of several such prints from German speaking countries.

The second phase of auroral research started with the very strong aurora on March 17, 1716 which seized in Germany people by fear and which was discussed by Wolff and Wagner in details, in order to reduce this fear. Nevertheless, these lectures and publications started a new age in auroral research and were soon followed by observations of geomagnetic activity during auroral events (Hjorter and Celsius).

The third, strictly scientific period started with the discovery of the connection with the 11-year solar cycle, then followed the spectroscopic study leading to the identification of the atmospheric phenomena during auroras.

The last phase includes active auroral research in the magnetosphere in close connection with geomagnetic phenomena, started during the International Geophysical Year (1957/58) and with the launch of the first artificial satellite.

The author presents a clear and richly illustrated description of all these phases and concludes about the historical aspects of auroral research, too, concerning the changes of the human attitudes against auroras. It is a most interesting lecture for everybody interested in auroras and in history of science in general.

J Verő

W SCHRÖDER ed.: *Physics and Geophysics (A compilation with special historical case studies).* History Commission of the German Geophysical Society, Mitteilungen des Arbeitskreises Geschichte der Geophysik der DGG, 15. Jahrg, (1999), Heft 1–3, Science Edition/DGG, 344 pp, 30 USD

In the introductory paper, Why Research into the History of Geosciences?, the Editor explains his aim with the publication of the series of publications. He emphasizes two aspects: the first is the history itself, namely the development of ideas during the centuries. The second is a specific one: old data about the occurrence frequency and other parameters of the phenomena contribute to the explanation of the mechanisms which led to these events, as e.g. aurora.

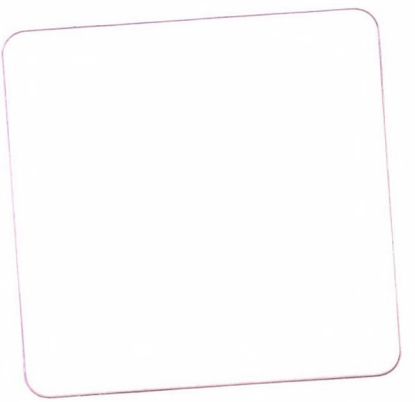
Following this introduction, a 80 pages long paper by Schröder discusses several aspects of the history of science, especially of geosciences, and contributes to the understanding of interconnections between general development of science and certain branches of geophysics, too, giving a philosophical treatment of the events and discoveries. Time histories of publications, expeditions, geomagnetic observatories etc. are shown to be in close connection with social-economical development, and also with current ideas about science.

In the next section of the book, Schröder, Hörz and Treder discuss several physical ideas, discoveries and disputes connected with geophysics, where names like Leibniz, Helmholtz, Ertel, Descartes, Newton, Mach, Eibnstein, Planck, Eddington, Laue, Gauss, Neumayer, Schrödinger, Föppl appear. Special attention is paid to phenomena like auroras – the green auroral line and noctilucent clouds.

In the last part, the Gregori's discuss the prehistory of geophysics as deduced from ancient navigational and orientation methods and variational principles in geomagnetism.

This book is again a most valuable addition to the history of physics, especially of geophysics, and offers a lot of interesting discussions and ideas for all readers.

J Verő



INSTRUCTIONS TO AUTHORS

Manuscripts should be sent to the Editorial Office (address see above). Articles not published or submitted for publication elsewhere are only accepted.

Manuscripts should be preferably on disc (LATEX *.tex files or Word for Windows documents or text-only ASCII files) with a written copy, or via e-mail (actagg@ggki.hu). Files will be accepted, after previous agreement, via *anonymous ftp*, too. Typewritten manuscripts (in duplicate) are also accepted.

Manuscripts should include the following components which should be presented in the order listed (tables and illustrations should be separated in case of manuscripts on disc, too).

1. Title, name(s) of the author(s), affiliation, dateline, abstract, keywords
2. Text, acknowledgements
3. References
4. Footnotes
5. Legends
6. Tables
7. Illustrations

1. The *affiliation* should be as concise as possible and should include the complete mailing address of the authors. The *date of receipt* will be supplied by the editors. The *abstract* should not exceed 250 words and should clearly and simply summarize the most important methods and results. 5-10 significant expressions describing the content are used as *keywords*. Authors may recommend these keywords.

2. The *text* should be in English and as short and clear as possible. In case of typewritten manuscripts, please note the following:

- avoid possible confusion between o, O (letters) and 0 (zero), l (letter) and 1 (one), ν (Greek nu) and u, v (letters), etc.
- explain ambiguous and uncommon symbols by making marginal notes in pencil
- formulas are to be numbered consecutively with the number in parentheses to the right of the formula.
  Text references to the equations may then be made by the number in parenthesis. The word equation in this context is to be abbreviated to Eq. and Eqs in the plural
- the International System of Units (SI) should be used.

3. *References* are accepted only in the Harvard system. Citations in the text should be as:

... (Bomford 1971) ... or Bomford (1971) ...
... (Brosche and Sündermann 1976) ...
... (Gibbs et al. 1976b) ...

The list of references should contain names and initials of all authors (the abbreviation et al. is not accepted here); for journal articles, year of publication, title of the journal in abbreviated form, volume number, first and last page. For books or chapters in books, the title is followed by the publisher and place of publication.

All items must appear both in the text and in the list of references.

4. *Footnotes* should be typed on separate sheets.

5. *Legends* should be short and clear. The place of tables and figures should be indicated in the text on the margin.

6. *Tables* should be numbered serially with Roman numerals. Vertical lines are not used.

7. All the *illustrations* should contain the figure number and author's name in pencil on the reverse. The most important point with figures is clearness. Photographs and half-tone illustrations should be sharp and well contrasted. Colour photographs will be accepted, but the extra cost of reproduction in colour must be borne by the authors. Encapsulated postscript (EPS) files will be accepted on disk or via e-mail, too.

Only original papers will be published and a copy of the Publishing Agreement will be sent to the authors of papers accepted for publication. Manuscripts will be processed only after receiving the signed copy of the agreement. Information is sent to the first author if no other wish is expressed.

Acta Geodaetica et Geophysica Hungarica

VOLUME 35, NUMBER 2, 2000

Akadémiai Kiadó, Budapest

AGG 35 (2) 105–253 (2000) HU ISSN 1217-8977

ACTA GEODAETICA ET GEOPHYSICA HUNGARICA

A Quarterly Journal of the Hungarian Academy of Sciences

Acta Geodaetica et Geophysica Hungarica (AGG) publishes original reports and reviews on geodesy and geophysics in English.

AGG is published in yearly volumes of four issues by

AKADÉMIAI KIADÓ
Prielle K. u. 4, H-1117 Budapest, Hungary
http://www.akkrt.hu

Manuscripts and editorial correspondence should be addressed to

AGG Editorial Office
Geodetical and Geophysical Research Institute
P.O. Box 5, H-9401 Sopron, Hungary
Phone: (36-99) 314-290
Fax: (36-99) 313-267
E-mail: actagg@ggki.hu
http://www.ggki.hu/agg/

Subscription information

Orders should be addressed to

AKADÉMIAI KIADÓ
P.O. Box 245, H-1519 Budapest, Hungary
Fax: (36-1) 464-8221
E-mail: kiss.s@akkrt.hu

Subscription price for Volume 35 (2000) in 4 issues US$ 180.00, including normal postage, airmail delivery US$ 20.00.

Printed in Hungary
PXP Ltd., Budapest

AGeod 35 (2000) 2

CONTENTS

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 105–132 (2000)

ON SPECTRAL METHODS IN SCHUMANN RESONANCE DATA PROCESSING*

J VERŐ[1], J SZENDRŐI[1], G SÁTORI[1], B ZIEGER[1]

[Manuscript received September 10, 1998]

Our experience in the processing of time series stems from the processing of magnetotelluric and geomagnetic pulsation data where conditions and aims are different from the processing of Schumann resonances. Nevertheless, several points are common in each of these fields. The major problem when processing time series is mostly the correct selection of the time/frequency resolutions. If one of them is increased, the other decreases. In the case of the Fourier transform, both resolutions are determined – in terms of a dynamic spectrum – by the time dimension of the time/frequency box within which the actual computations are made. In case of convolution filtering (especially if both components of complex vectors are computed), the selection is more versatile. The step in frequency can be freely selected, the independence, however, of the filtered series is only ensured if filters have sufficient length in time what means a corresponding time resolution. If time resolution is to be increased, then the frequency step most be increased (frequency resolution decreased) to get independent time series. Convolution filtering has the advantage that disturbed sections can be easily cut from the filtered series without disturbing other sections, thus the reduction of noise is more effective. Moreover "momentary" spectra can be found in any moment or section. Additionally, an impulsive event can also be resolved independently of possible other impulses in arbitrary position outside of the length of the filter. The same method of convolution filtering can also be used for the complex demodulation of time series, resulting in high precision frequency determination.

Keywords: dynamic spectrum; Schumann resonance; time series analysis

1. Introduction

The first problem in time series processing emerged in our Institute in connection with the processing of magnetotelluric data. Here frequency dependent, complex linear transfer functions are to be determined from the four (five) recorded electromagnetic components, between two electric and two (three) magnetic components. The usual method for the computation is to find the complex spectrum — in the possible widest range of frequencies — together with cross factors (coherence functions), select sections of the records where certain coherence conditions are fulfilled, then proceed with rather simple formulas. In spite of the apparent simplicity of this procedure, several problems emerge which are more or less common for all similar computations. Some of these problems are the following:

*Paper presented at the Schumann-resonance symposium, Sopron, Hungary, September 10, 1998

[1]Geodetic and Geophysical Research Institute of the Hung. Acad. Sci., H–9401 Sopron, POB 5, Hungary

1217-8977/2000/$ 5.00

1. There are three possible causes for low coherence. The first is an inherent property of the incoming electromagnetic waves which are supposed to be vertically incident plane waves. In certain cases this condition is not fulfilled, therefore the coherence remains low for longer time intervals, especially during geomagnetic storms. The second cause of low coherence is specific for magnetotellurics, namely strongly inhomogeneous geologic structure. In contrast, the third is common for all time series processing, namely noise, more exactly low signal/noise ratio.

2. As the scatter of the computed apparent resistivities may be sometimes rather high, some kind of weighting/qualification of the final data is unavoidable.

3. The transfer functions magnetic $\rightarrow$ electric and electric $\rightarrow$ magnetic are theoretically inverses (reciprocals) of each other, but due to noise (and inhomogeneous geologic structure) their computed product differs from the unit tensor.

The most frequently used method for the computation of the spectrum is Fast Fourier Transform (FFT). In the MT case, however, it proved to be very difficult (at least in the sixties-early seventies) to effectively use it. The major problem was the elimination of impulsive noise, easily recognisable on the analogue records. If too many such impulses were cut off, the FFT results were difficult to handle. That is why we switched on convolution filtering (Verő 1972). If the coherence condition is set in the program to 0, all vectors are accepted. (Convolution filtering is the same as wavelet analysis with a bandpass filter in a modern terminology.)

The equations for the convolution filter used for the computation of the spectra are:

$$\begin{aligned} F(t) &= \frac{1}{2\pi t}\left(\sin\frac{2\pi t}{p_1} - \sin\frac{2\pi t}{p_2}\right)\cdot\left(\cos\frac{2\pi t}{T}+1\right) \\ F(0) &= \frac{2d}{p_1} - \frac{2d}{p_2} \\ G(t) &= \frac{1}{2\pi t}\left(\cos\frac{2\pi t}{p_1} - \cos\frac{2\pi t}{p_2}\right)\cdot\left(\cos\frac{2\pi t}{T}+1\right) \\ G(0) &= 0 \end{aligned} \tag{1}$$

where t is time (of the filter), F is the in-phase, G the out-of-phase component of the filtered vector, d is the sampling rate, p_1 and p_2 the period limits of the box-car type filter, T the length of truncation (with a Hanning-window).

The result of the convolution consists of an in-phase (computed with $F(t)$) and of an out-of-phase (computed with $G(t)$) amplitude which give together the complete vector with phase situation. In the original version of the program the step between adjacent filters corresponds to the integer part (expressed in units of the sampling rate) of the period, but it may be shorter, even may correspond to the sampling rate. The original choice meant an earlier significant saving of computer time.

There are, naturally, several more sophisticated methods for the spectral decomposition of time series, but they are mostly too pointed to be used in automatic systems.

2. Parameters in spectra and their interdependence

The following considerations can be best explained considering the case of dynamic spectra. In dynamic spectra, the horizontal axis corresponds to time, the vertical to (logarithm of) frequency/period, and amplitudes are indicated in form of isolines, shading, colours etc. For the determination of the spectrum, a rectangular box is to be used in this kind of representation. In the case of FFT, the dimensions of this box are determined by a single parameter, namely by the length of the time interval processed. The possible coverage in the frequency/period direction is completely determined by the length of the time window, as the longest period in the computation corresponds to the length of the box, the shortest one is determined by the sampling rate within the time window (Nyqvist period). The distance between the "lines" parallel with the time axis, i.e. the frequency/period resolution is changing within the box, being roughly proportional to $1/n$, n being the serial number of the harmonics considered. Thus "lines" are much denser at shorter periods/higher frequencies than at longer periods/lower frequencies. Thus if e.g. due to an impulse a section of the records must be cut out, then the distribution of the lines changes, too (or the lacking data are to be substituted somehow).

In the case of the convolution filtering, the situation is quite different. Here the distance between adjacent filters can be freely chosen, e.g. corresponding to a certain step in $\log T$ (or $\log f$). Naturally, in a strict sense of the word, the choice of the distance between components is optimum in case of FFT; if the step is different, the lines are not independent (at long periods) or each line includes several harmonic components (at short periods). This theoretical limitation has to be correctly considered in each utilisation of the method, but in a number of cases, it does not mean a serious problem.

The differences in the resolutions are illustrated in Fig. 1, where the distribution of the location of lines is given in a dynamic spectrum for an FFT and for a few convolution-type filter series. In the case of the FFT spectrum, the location of the lines is strictly given by the length of the time window and the same situation is found for a filter series with equal spacing between adjacent filters in frequency. In this case the density of lines is very high at short periods and very low at long periods (Fig. 1a). For a filter series with equal spacing in period, the two rows in Fig. 1a get inverted figure, else the two distributions are the same, only the high density is at low frequencies and the low density is at high frequencies (Fig. 1b). In the case of a filter series with equal spacing in percents (of the central frequency), the distribution is more uniform, therefore the construction of the dynamic spectrum is easier than in the two previous cases (Fig. 1c). Naturally, other possibilities have also their merits, as e.g. in dynamic spectra from FFT, the harmonic structure can be most easily recognised if the vertical scale corresponds to frequencies as in Fig. 1a top right: in that case the lines are equidistant.

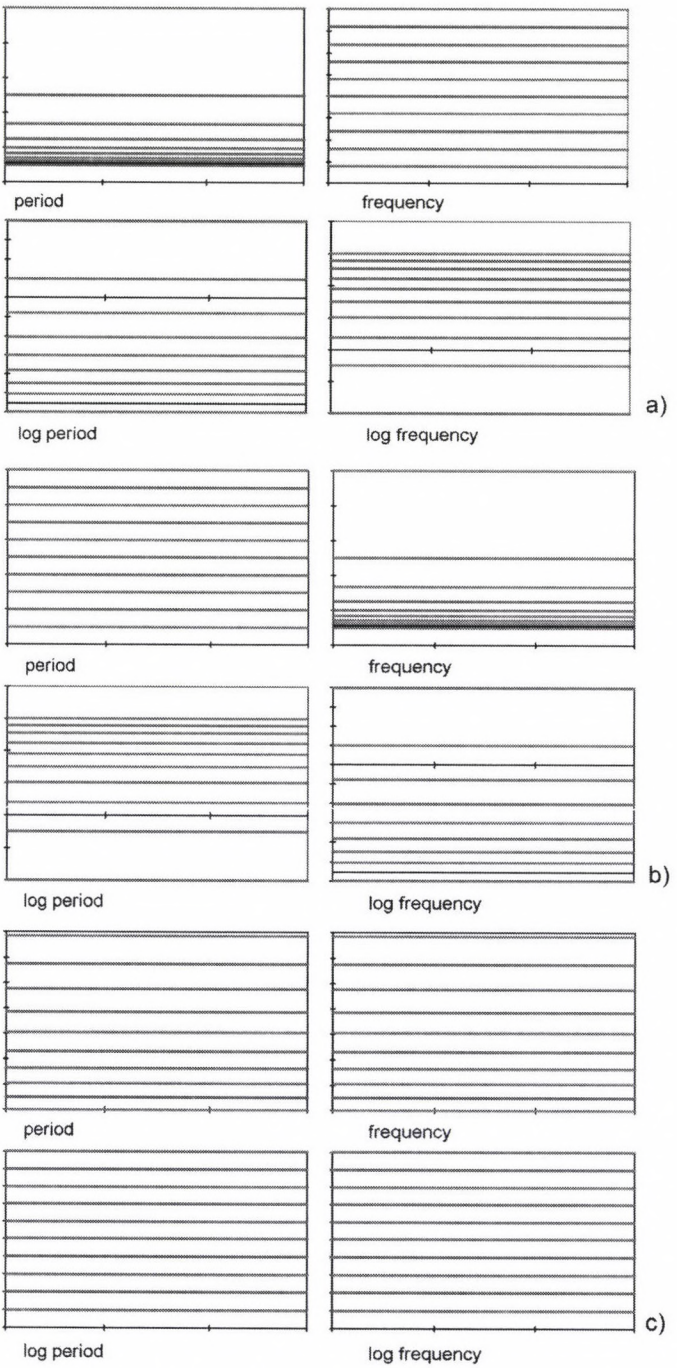


Fig. 1. Distribution of the location of lines in dynamic spectra computed with a) FFT, or filters equally spaced in frequency, b) filters equally spaced in period, c) filters equally spaced in $\log T$ or $\log(1/T)$ (equal spacing in percents of the central frequency/period)

In the case of the convolution filters given by Eq. (1), the time resolution of the filter is controlled by T and the frequency/period resolution by the distance of p_1 and p_2 for a given sampling rate, d. The two resolutions are theoretically independent, in the practice, however, they are not. Namely the length of the filter is most advantageously chosen if the end of the Hanning-window coincides with the first minima of the filter envelop on the left and right sides of $t = 0$. The position of this minimum depends on its side on p_1 and p_2, more exactly, on the ratio of the two limits. In the case of "normal" dynamic spectra (our nomenclature), they are 10 percent less/more than the central period p (in the subsequent filter, the central period changes by 10 percent, too, thus the whole range of periods is covered by the filter series with double overlapping). Correspondingly, $T = 5p$ and this is the half length of the record from which values are used for the determination of a filtered value.

If we want a higher frequency/period resolution, the distance between $p - p_1$ and p_2 must be less, e.g. 2 percent. In this case, however, a higher value must be chosen for T (else the "theoretical" limits p_1 and p_2 would be missed too much). An appropriate choice would be on the same basis as previously $T = 25p$.

Thus, if the frequency/period resolution is increased n-times, then the time resolution, being proportional with T, has to decrease, at least approximately n-times, too.

2.1 Period/frequency

The two methods which are compared here have evidently their own merits and setbacks. Concerning periods determined with FFT, Figs 2 and 3 show an extreme case to call attention to effects eventually occurring in a certain degree in realistic data series, too. In the analogue artificial record represented in Fig. 2, two impulses of identical form (period 100 s and both of the same amplitude), but in opposite directions occur with a time distance between the end of the first impulse and the beginning of the second one corresponding to an integer multiple of the period, in Fig. 2 to 100 s. In such a case the "period" of the impulse appears as minimum in the spectrum (Fig. 3) and a series of maxima appear at periods differing quite significantly from the "realistic" one. This situation may occur in the case of impulsive events, it is less dangerous in the case of more or less continuous activities. In the case of filtering, the result depends strongly on the distance between the two impulses; if it is short (Fig. 4), then two maxima appear in the spectrum as in FFT spectra, but in the case of longer separation, the single maximum is in the right position (Fig. 5), while the separation influence FFT spectra less.

In general, the application of FFT gets more and more difficult the more variable the frequency is, as it cannot follow the time history of the signal. This fact is reflected e.g. in the use of matched filtering when dispersive signals, thus whistlers are processed. In the case of slowly varying frequencies the problems are naturally less severe, but it remains difficult to judge how strongly the results are distorted.

Convolution filters with properly chosen period ranges trace automatically, within the pre-defined range, the changes of the period or if a filter series is used, even changes covering the period ranges of several filters can be traced. Based on the

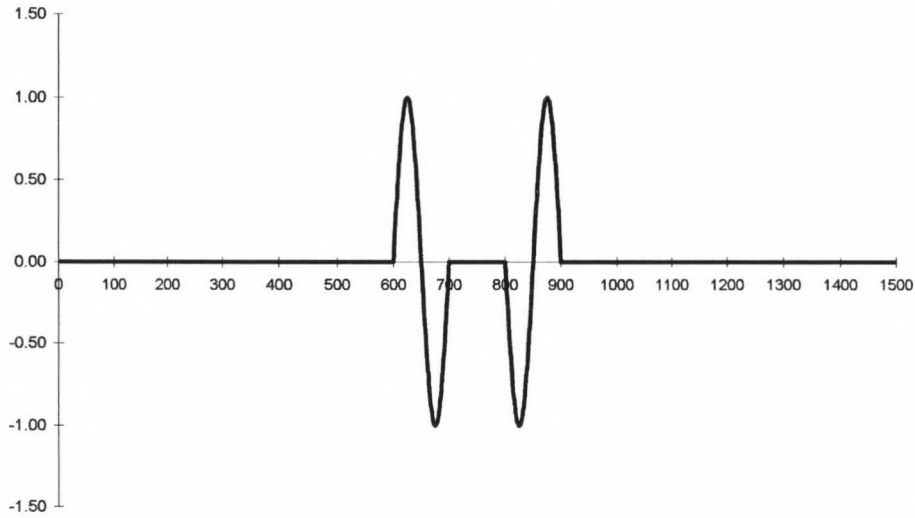


Fig. 2. Synthetic record with two 100 s impulses in anti-phase, time interval between the impulses 100 s

vectors resulting from the filtering, complex demodulation yields the possible most exact periods based on the difference of the actual rotation of the field vector and of the rotation expected for a vector with exactly the central period of the filter. This method for the determination of the momentary periods of the Schumann resonances is used at the Nagycenk observatory.

The determination of the period(s) of a signal implies other problems, too. Here a special problem connected to more or less impulsive events should be mentioned. Namely in our experience dynamic spectra are most informative if the resolutions in the directions of time and frequency/period are the same with the chosen scales. In such a case, circular structures are often seen in dynamic spectra of geomagnetic pulsations. A horizontal row of such circular spots can be interpreted in two different ways: either impulses are present on a background of more or less strong activity, or two or more adjacent spectral lines interfere. In the latter case the length of spots (most advantageously expressed as the number of waves between maxima or minima) indicates the difference between the interfering lines. If the resolution in the direction of periods is increased (the nominal bandwidth of the filter is decreased), then the supposed interference structure can be resolved into a number of lines with a difference of periods between them corresponding to the previously mentioned value. In this case, however, the time resolution decreases, as the length of the filter has to be increased to reach the necessary frequency/period resolution. That means that it cannot be decided on the basis of the dynamic spectra if the impulsive- or the interference-type representation corresponds to the actual physical situation.

The line splitting discussed is quite often present in certain methods, e.g. in MEM (Maximum Entropy) spectra — especially in the case of low noise — the splitting appears very often, it is, however, clearly an artefact of the method.

The situation outlined here was recently mentioned by Oliver et al. (1998) in a different context, but with the same conclusion.

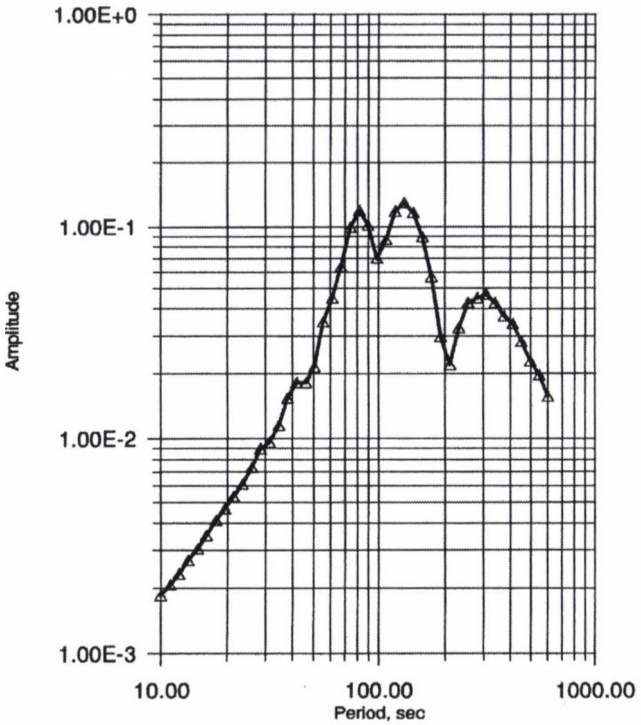


Fig. 3. Fourier spectrum of the synthetic record in Fig. 2

2.2 Amplitudes

The determination of the amplitude of a certain spectral component is by no means a simple problem, even if the period is already known. Namely time series are nearly always infected by noise, the spectrum of the noise is seldom known, and therefore in automatic processing systems, an unknown amount of noise with unknown spectra is to be considered. The situation is simpler if some parameter of the noise is at least vaguely known, thus e.g. if the amplitude of impulsive noise is higher than the average amplitude of the signal, then by choosing an appropriate upper amplitude limit for the acceptance of the (filtered) vectors a significant part of the noise can be eliminated. (Similarly a lower limit may also be useful e.g. in the complex demodulation method to eliminate signals/noise with too low amplitude, therefore of low reliability of the periods and amplitudes.) In the case of FFT that means that the boxes within which the computations are made are pre-tailored by the sequence of impulsive noise. In the case of convolution filtering, noisy sections can be easily cut out from the dynamic spectra and even momentary spectra can be found as profiles in the direction of the frequency axis if at least short noise-free sections exist in the record.

Other methods for the elimination of artificial (the term "artificial" means in the following all signals which are not due to investigated phenomenon, as e.g. to Schumann resonance, not only really artificial noise) noise may also be used, in the

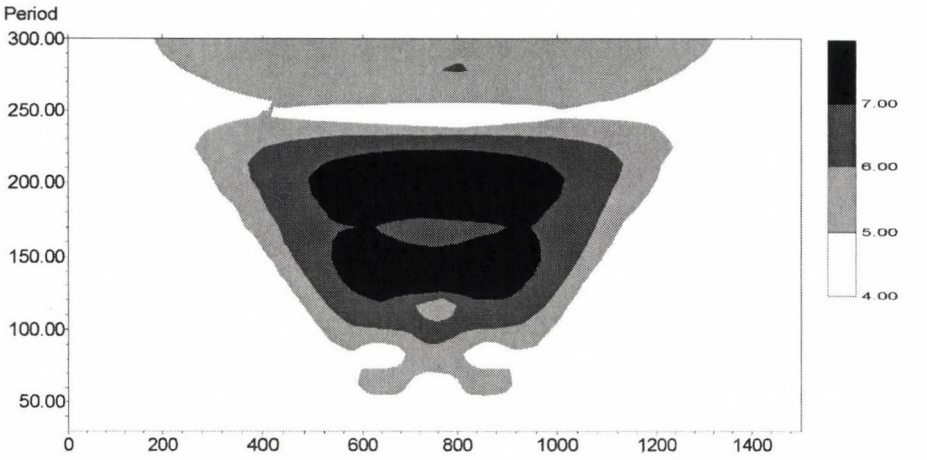


Fig. 4. Dynamic spectrum from convolution filtering of the synthetic record in Fig. 2

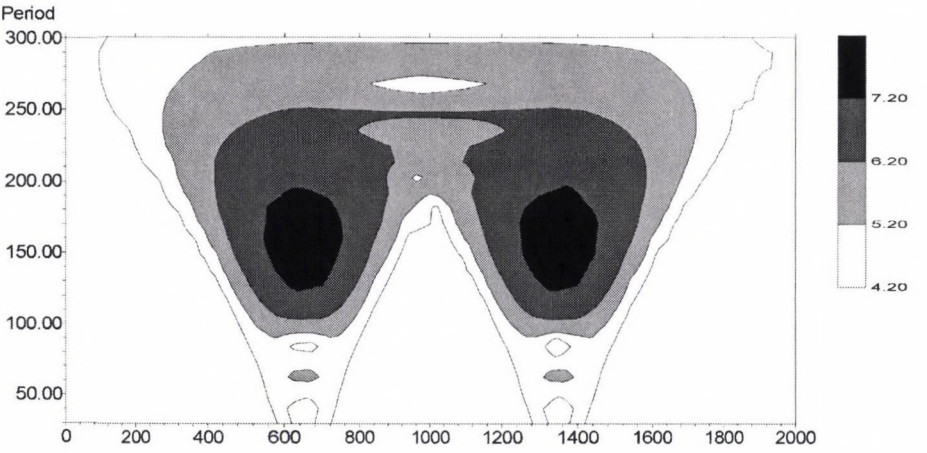


Fig. 5. Dynamic spectrum of a synthetic record being similar to Fig. 2, but with longer interval (600 s) between the two impulses

original MT program e.g. the coherence limit eliminates sections with noise in only one of the electromagnetic components, as e.g. magnetic signals due to movement of a car.

2.3 Phase

Among parameters of spectral component, the phase information can be extracted with most difficulties. Without going into details, only the necessity of rigorously identical periods for the two components to be compared are to be mentioned what is no constraint in the case of the convolution filtering in the sense that if two series are compared for phase, a continuous phase shift between two components indicates some difference in the observed periods (identical periods in any two components, however, are ensured as a constraint in FFT). Time difference between data points should also be negligible or they must be taken into account.

3. Examples from the processing of Schumann resonance data

In the following, we apply the previously discussed points to three records made in the Nagycenk observatory. All the three have a length of 36 s and refer to the vertical electric component.

3.1 Event 1, July 18, 1995, 15:16:13 LT, Nagycenk, WEST2

This record (Fig. 6) is nearly noise-free, a stronger impulsive event appeared around 18 s, and long period (about 2.5 Hz) waves were present between 21 and 23 s. The first three harmonics of the Schumann resonance can be clearly identified at 8, 14 and 21 Hz in the FFT spectrum computed in 2.56 s long time windows (Fig. 7). The first two have similar amplitudes, the third is less clear, its absolute power is less by only 10 percent, but with respect to the background it has a peak about half of that of the first two harmonics.

The first series of convolution filters with 10 percent difference between the central periods of the adjacent filters (thus the bandwidth corresponds to twice this value) yielded the spectrum presented in Fig. 8, the length of the filters corresponds to 8 complete cycles. Here the first two harmonics are clearly present, the third has only slight indication.

Increasing the frequency resolution of the filters (Fig. 9, distance of the central period of adjacent filters is now 5 percent, length of the filters 20 cycles) there are more details in the spectrum, the third harmonic gets much more evident, but also spurious peaks appear, especially between the first two harmonics.

A repeated increase of the frequency resolution (Fig. 10, 2 percent, filters 40 cycles long) results in even more peaks, the spectrum gets rather rugged, nevertheless, the actual periods of the three harmonics (7.93, 13.8 and 19.7 Hz) can be easily found.

A further decrease of the time resolution (Fig. 11, filters 80 cycles long) results in a serious splitting of the peaks and the determination of the frequencies belonging to them becomes dubious.

This event and its processing should indicate the possibilities of the convolution filtering. The dynamic spectrum with the first series of filters (Fig. 12) enables us to trace the time history of the activity of the first two harmonics. Both are impulsive, spectra are decisively influenced by a series of strong impulses around 1,

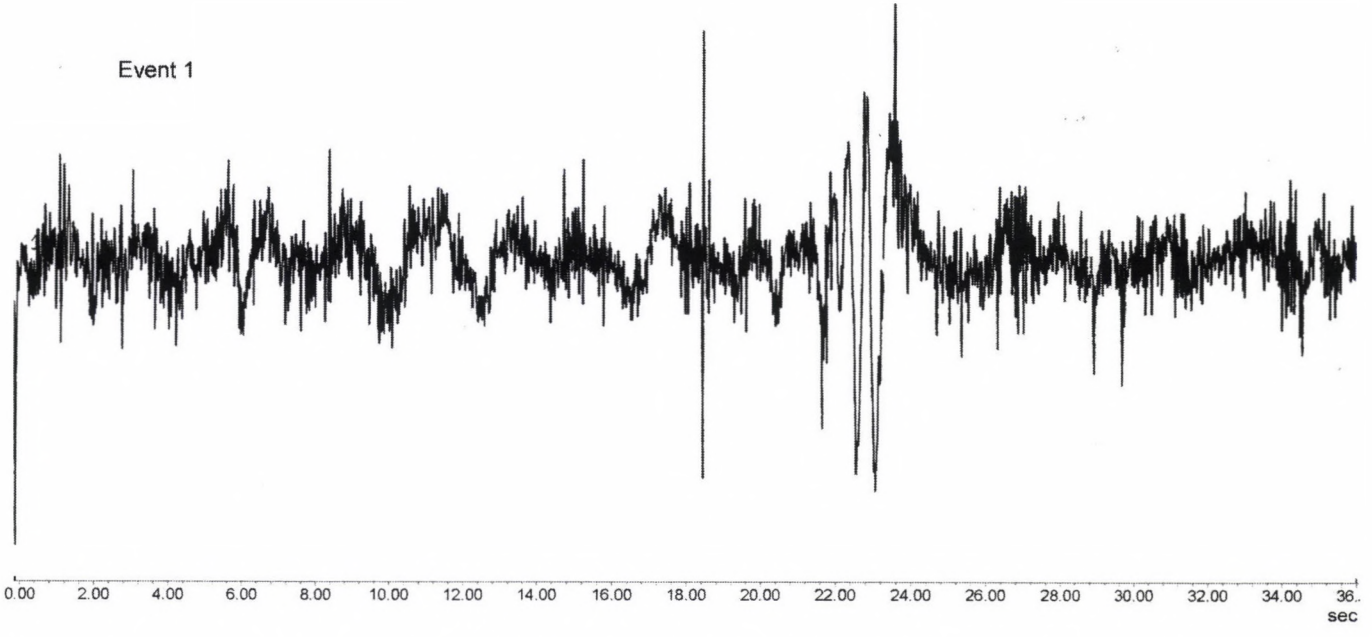


Fig. 6. Analogue record of Event 1, July 18, 1995, 15:16:13 LT, Nagycenk (WEST2). Sampling frequency 100 Hz, length 36 s, component E_z

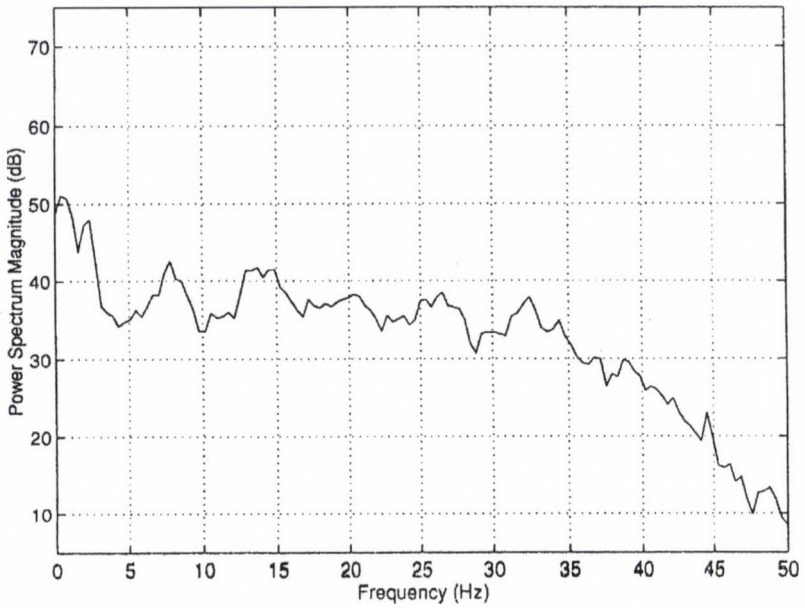


Fig. 7. Spectrum of Event 1, computed using FFT in 256 data long sections

4, 9, 22, 24, 25, 27, 31, 35 s. A part of them (especially at 1 and 24 s) are about equally strong in both harmonics, others are stronger in the second harmonic. As the latter group is more numerous, the peak in the overall spectrum corresponding to the second harmonic is also higher. Supposedly artificial impulses are present at the very beginning of the event and also at 18 s: these do not have the characteristic spectrum of the resonances.

In the dynamic spectrum with a period resolution of 5 percent (Fig. 13) the spots are stretched in the direction of the time axis, especially in the case of the first harmonic, where the temporal length of the filter is longer. Using this series of filters, the spots are most circular, according to the mentioned experimental criterion this dynamic spectrum can be coined as optimum.

In the case of 2 percent frequency resolution (Fig. 14) the spots are horizontally elongated, some of them becomes connected with each other, thus they cannot be separated, and even the artificial impulses can be less easily identified.

The dynamic spectrum with lowest time resolution (Fig. 15) is hardly informative, spots corresponding to impulses are smeared and the time history of the events cannot be reconstructed.

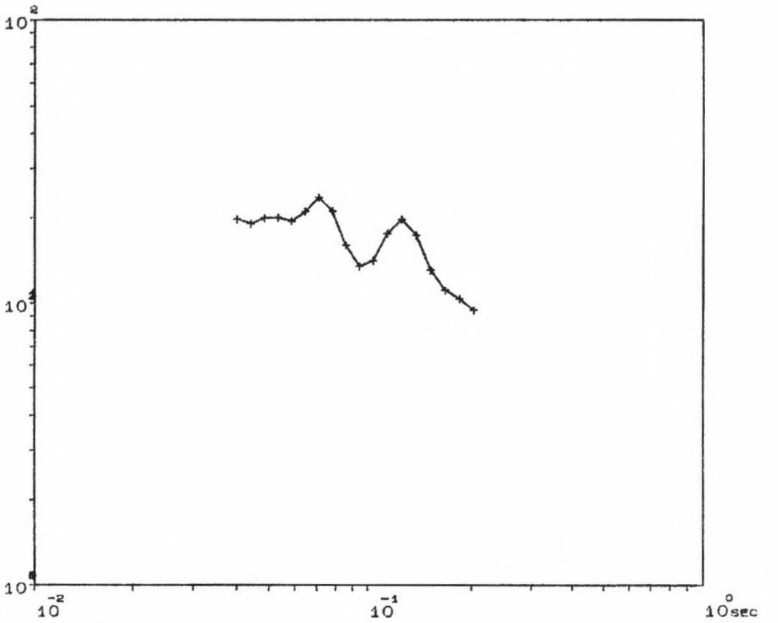


Fig. 8. Spectrum of Event 1, computed with convolution filters, having 10 percent spacing between them, and the nominal bandwidth is the same 10 percent on both sides of the central frequency. Length of the filters 8 complete cycles

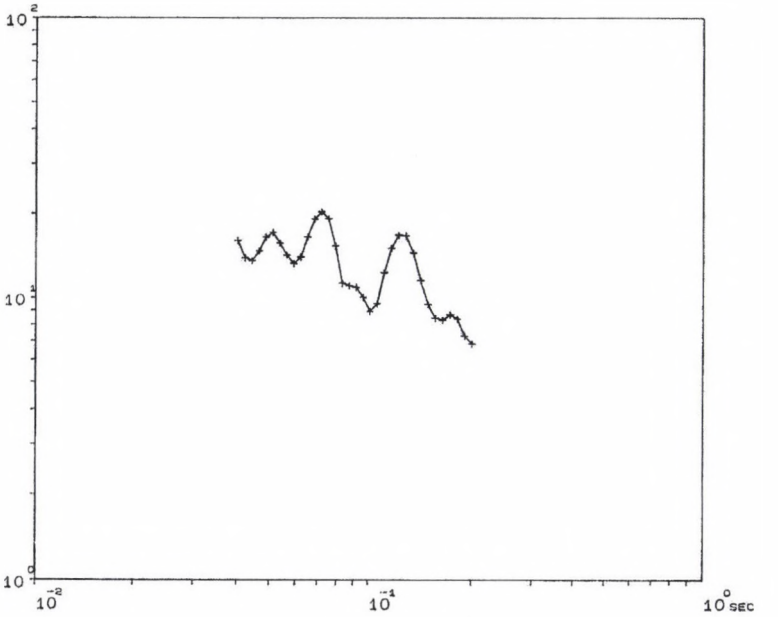


Fig. 9. Spectrum of Event 1 computed with convolution filters like in Fig. 8, but the spacings of the filters and the bandwidths are here 5 percent, with 20 cycles long filters

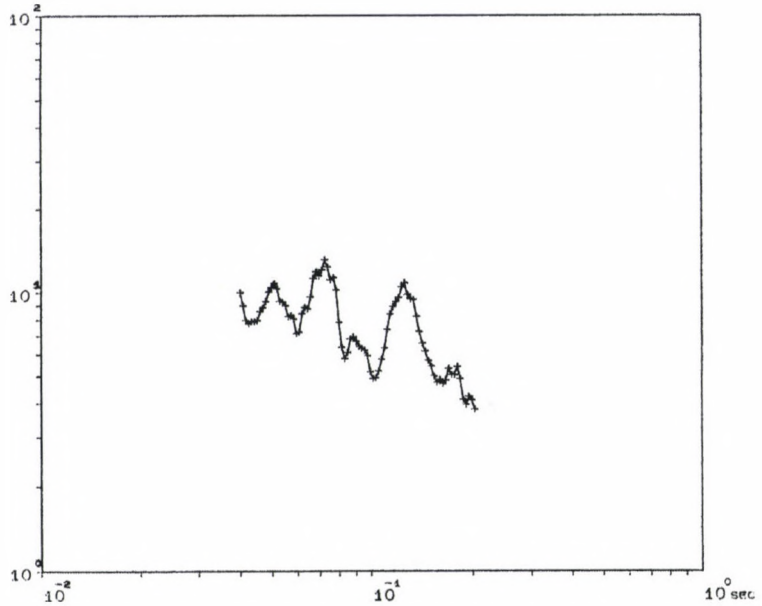


Fig. 10. Spectrum of Event 1 computed with convolution filters like in Fig. 8, but the spacings of the filters and the bandwidths are here 2 percent, with 40 cycles long filters

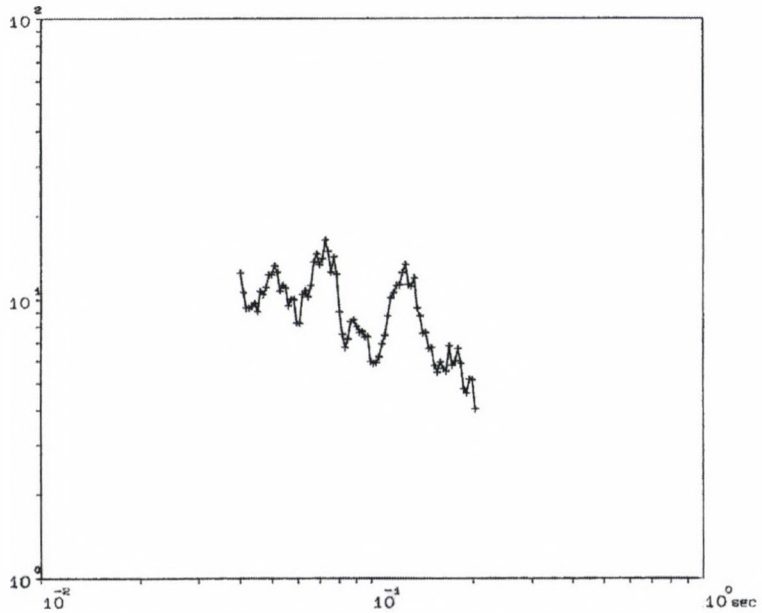


Fig. 11. Spectrum of Event 1 computed with convolution filters like in Fig. 8, with spacings of the filters and bandwidths like in Fig. 10, 2 percent, but with 80 cycles long filters

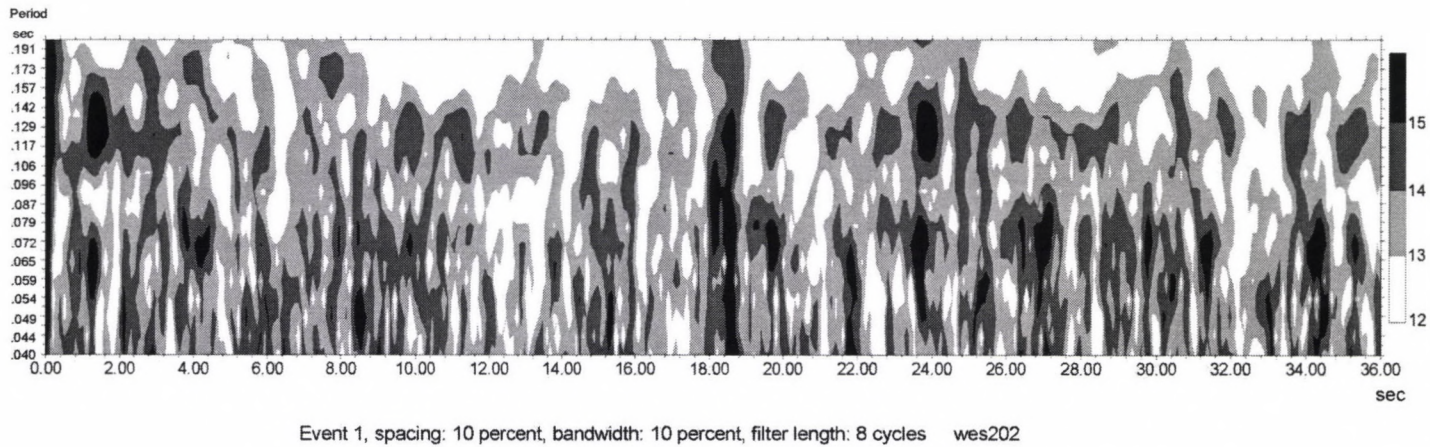


Fig. 12. Dynamic spectrum of Event 1 computed with convolution filters. The filters are the same as in Fig. 8

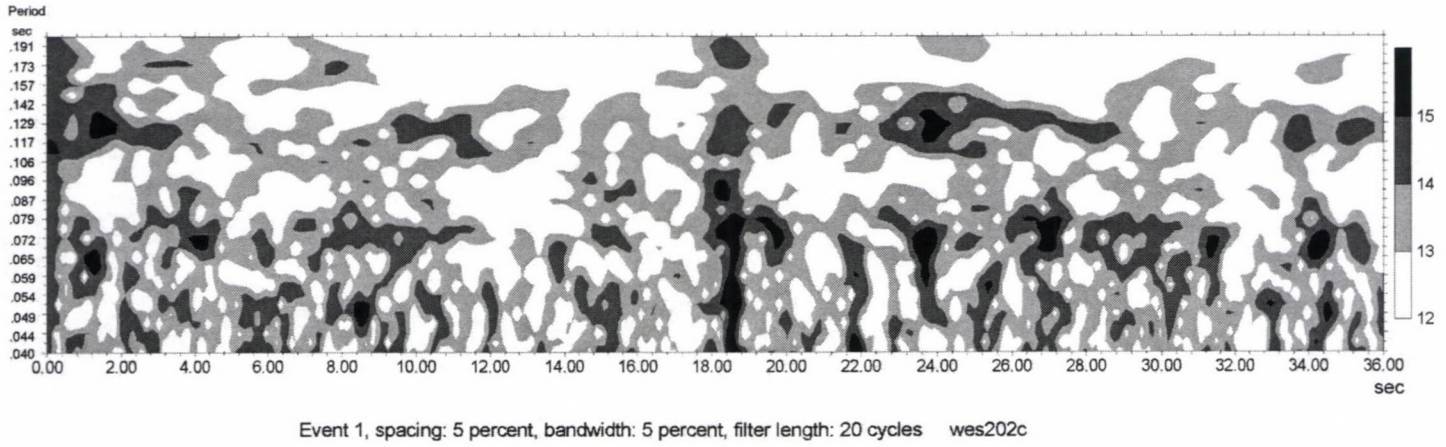


Fig. 13. Dynamic spectrum of Event 1 computed with convolution filters. The filters are the same as in Fig. 9

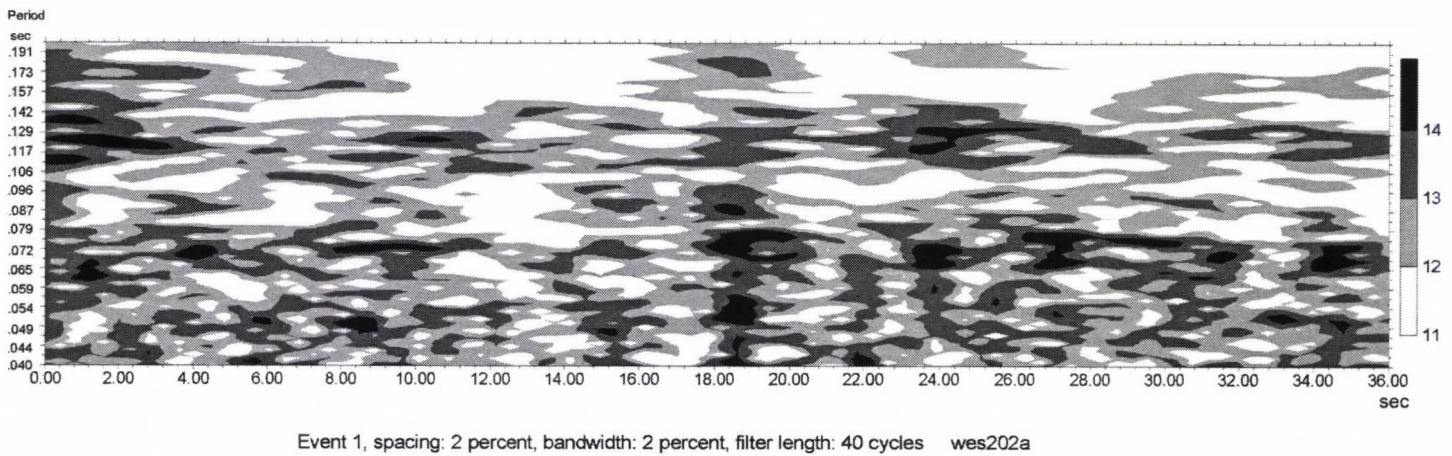


Fig. 14. Dynamic spectrum of Event 1 computed with convolution filters. The filters are the same as in Fig. 10

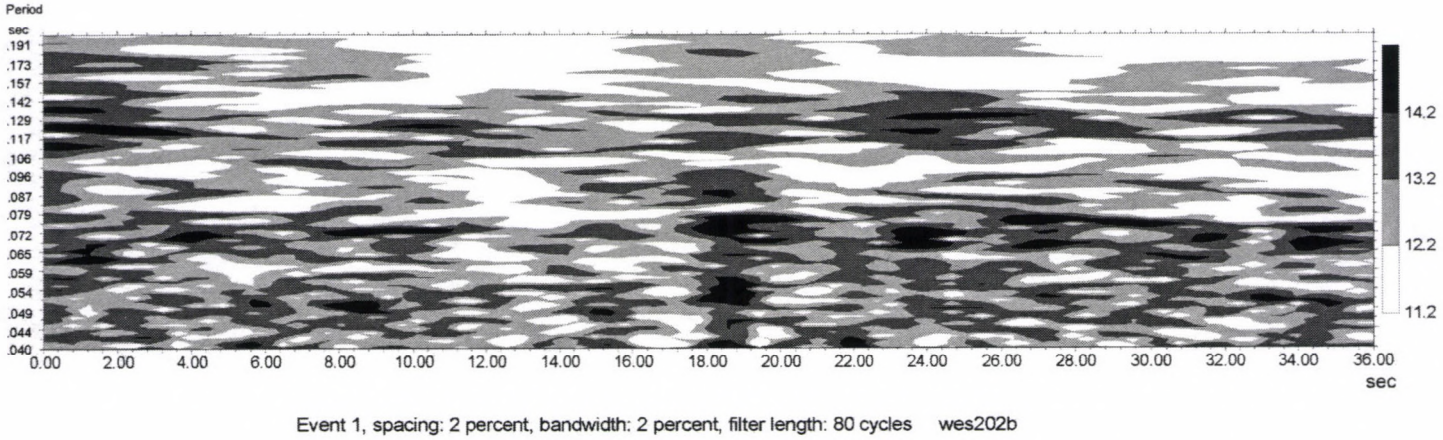


Fig. 15. Dynamic spectrum of Event 1 computed with convolution filters. The filters are the same as in Fig. 11

3.2 Event 2, July 18, 1995, 15:44:57 LT, Nagycenk, WEST16

This event (Fig. 16) includes shortly after its beginning (at about 2 s) a very strong artificial impulse which determines the spectrum of the whole event. The FFT spectrum of the event (Fig. 17) differs in several points from that of the first event (Fig. 7). The Schumann resonances cannot be identified and the slope of the part at 0 to 20 Hz differs strongly, too: in Fig. 7, this part is roughly horizontal (no slope), here the level decreases by 10 to 15 dB. This decrease is surely due to the effect of the impulse at 2 s.

If the impulse is cut out, then the FFT spectrum changes dramatically (Fig. 18): three Schumann resonant frequencies become evident, the slope gets similar to that of Event 1. As in that case the impulse was of rather long duration (about 1 s), the long period/low frequency part of the spectrum is most strongly contaminated by it; nevertheless, details of the spectrum at short periods/high frequencies differ significantly, too, as here not only the impulse, but the signal has much smaller amplitudes.

In the case of the spectrum from convolution filtering, the situation is naturally quite similar: with the impulse at 2 s, the slope is opposite to that without the impulse (Figs 19 and 20). Nevertheless, there is a significant difference in the spectrum with the impulse against that from FFT: the second and third harmonics of the Schumann resonance are present in Fig. 19, too. This difference is due to the more robust character of the filtering, it is less sensitive to outliers, as the impulse can be treated as such, too. Here again the third harmonic gets more accentuated if the bandwidth of the filter is reduced to 5 percent.

The dynamic spectrum with 10 percent bandwidth (Fig. 21) is quite similar to that of Event 1, with the addition of the impulse at 2 s. The rather high activity around 30 s is due to Schumann resonances, as it has the characteristic structure with three identifiable frequencies.

It is informative to compare the spectra for this event obtained using the two methods. Figure 22 shows this comparison for the filter series with 5 percent distance between the filters. Concerning the frequency of the spectral peaks, the two methods give practically identical values. Concerning amplitudes, the initial (long period, low frequency) sections are similar, too, with increasing frequency the spectrum from filtering gives greater and greater amplitudes. The cause for this difference is evidently the different response of the two methods for impulsive (randomly distributed) signals. In the case of the FFT spectrum, the energy/amplitude in one 2.56 s long box can be "destroyed" if there are several impulses within the same box, see Figs 2 and 3, too). Thus the amplitudes may become less

a) if the corresponding frequency is due to noise — this is a favourable property of the method,

b) and if several impulsive signals of identical frequency, but of random distribution are present — this can be an unfavourable property of the method, as it suppresses frequencies being present in the sample.

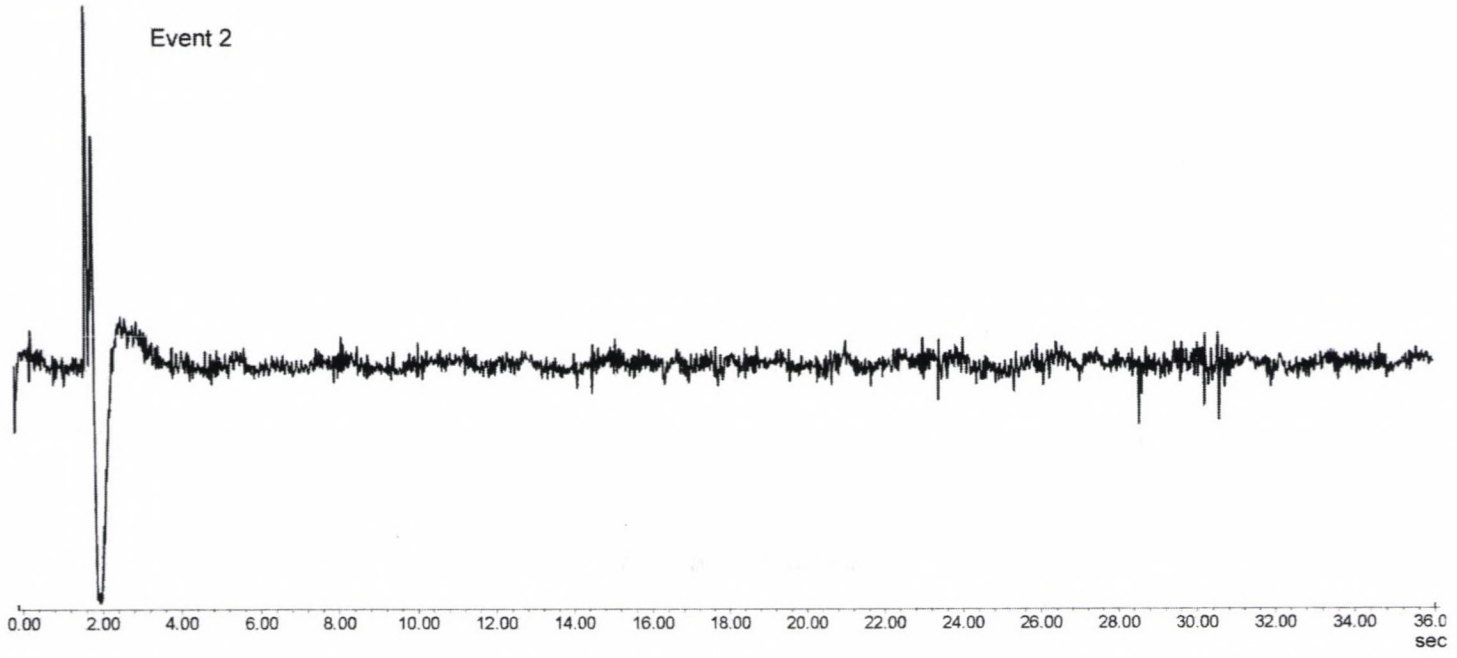


Fig. 16. Analogue record of Event 2, July 18, 1995, 15:44:57 LT, Nagycenk (WEST16), sampling and length as in Fig. 6

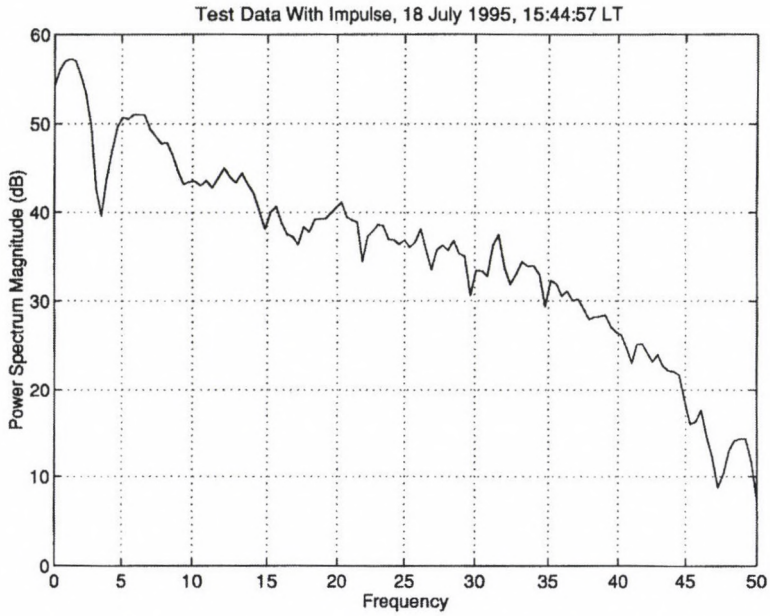


Fig. 17. Spectrum of Event 2, computed using FFT in 256 data long sections

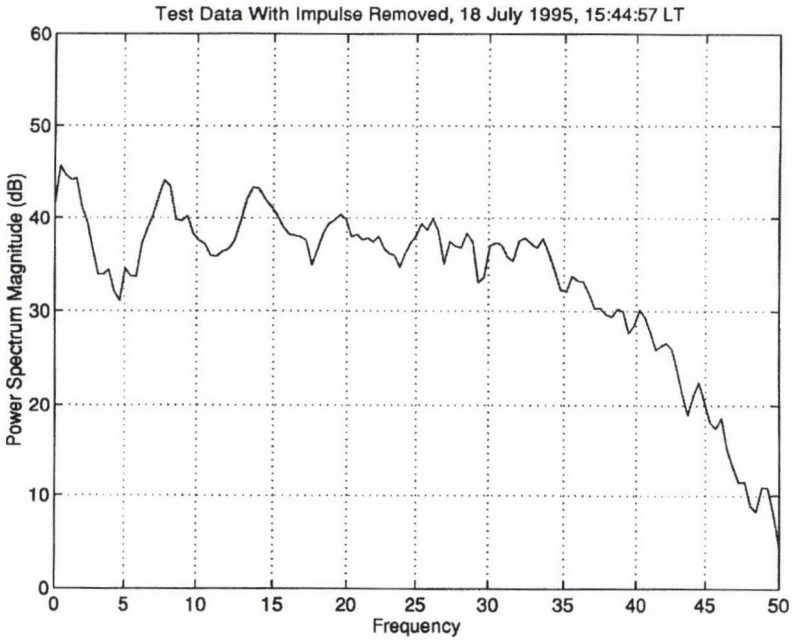


Fig. 18. Spectrum of Event 2 without the strong impulse at 2 s, computed using FFT in 256 data long sections

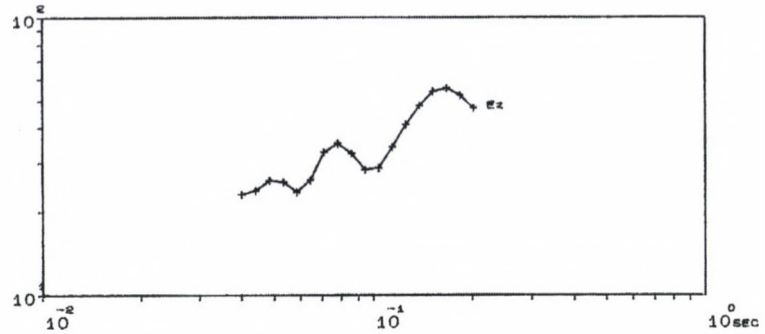


Fig. 19. Spectrum of Event 2, computed with convolution filters like in Fig. 8

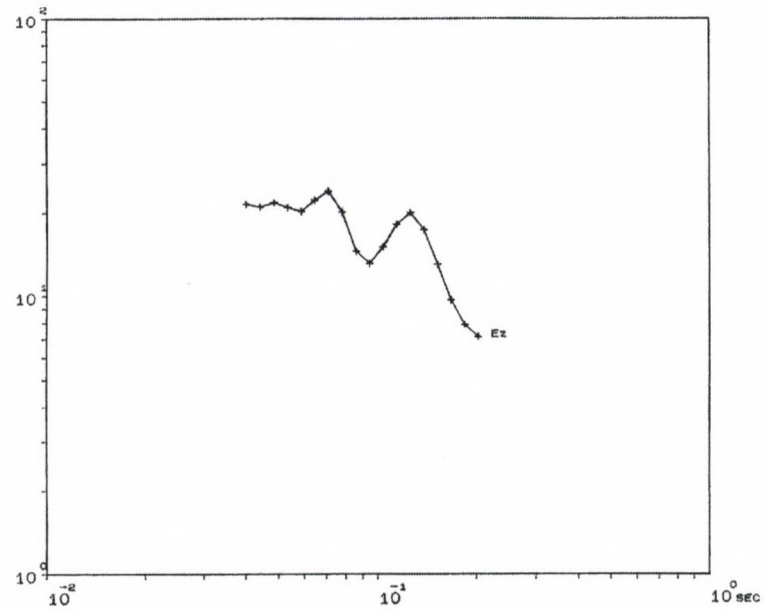


Fig. 20. Spectrum of Event 2 without the strong impulse at 2 s, computed with convolution filters like in Fig. 8

In the case of spectra computed from filtering, the phase situation of the signals — e.g. due to random appearance — does not influence the result, "energies" found by the filter in different positions are simply summed up, irrespective of phase situation. Thus, in both previously mentioned cases amplitudes can be greater from convolution filtering than from FFT. As in the given sample impulses of frequencies corresponding to (at least the second and third) Schumann resonances appear several times in the same 2.56 s long time window, and noise has comparable amplitudes to the signal(s), the higher amplitude level from the filtering can be explained.

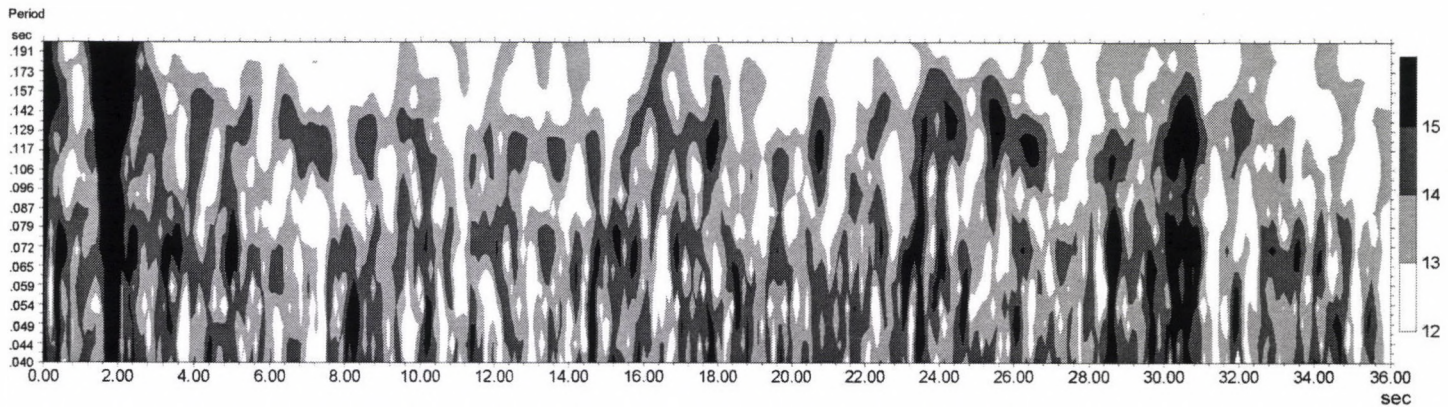


Fig. 21. Dynamic spectrum of Event 2 computed with convolution filters. The filters are the same as in Fig. 8

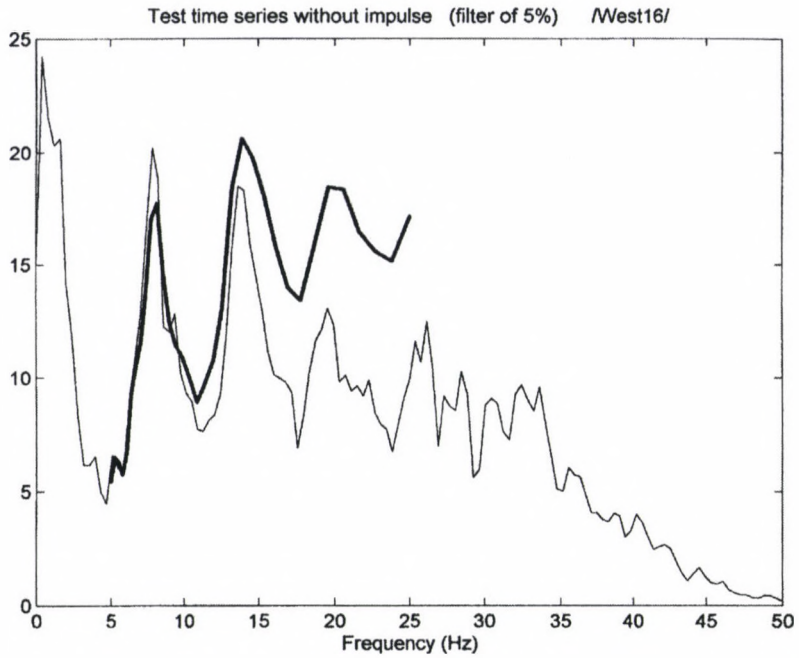


Fig. 22. Comparison of spectra of Event 2 without the strong impulse at 2 s, obtained by FFT (as in Fig. 18) and by convolution filtering (spacing and bandwidth 5 percent, filter length 20 cycles)

3.3 Event 3, July 18, 1995, 15:29:45 LT, Nagycenk, WEST9

This event (Fig. 23) is characterised by a series of artificial impulses, both long period and short period ones. The only longer section free of impulses is between 5 and 17 s. The corresponding FFT spectrum (Fig. 24) is similar to that of Event 2, it has a strong slope, perhaps even steeper than that of Event 2 due to more impulses. No Schumann resonance frequency can be identified.

The spectrum from the convolution filtering (10 percent spacing of the filters, Fig. 25) is also similar to that of Event 2, here the second and third harmonics can be perhaps found. The corresponding spectrum from the short noise-free section (Fig. 26) has opposite slope, too, the harmonics are with the 10 percent bandwidth, however, not very clear. An increased frequency resolution (Fig. 27) does not improve significantly the situation, indicating that this time interval is perhaps too short for an unambiguous determination of the frequencies.

The dynamic spectrum (Fig. 28) shows that the activity of the first harmonic is very poor in the noise-free interval, that of the second somewhat stronger. The very high amplitudes shift the automatically chosen limits of the different shadings upward, therefore the activity seems to be even lower than it is in realty, nevertheless, there is only a single impulse around 9 s which has the characteristic structure of the Schumann resonance.

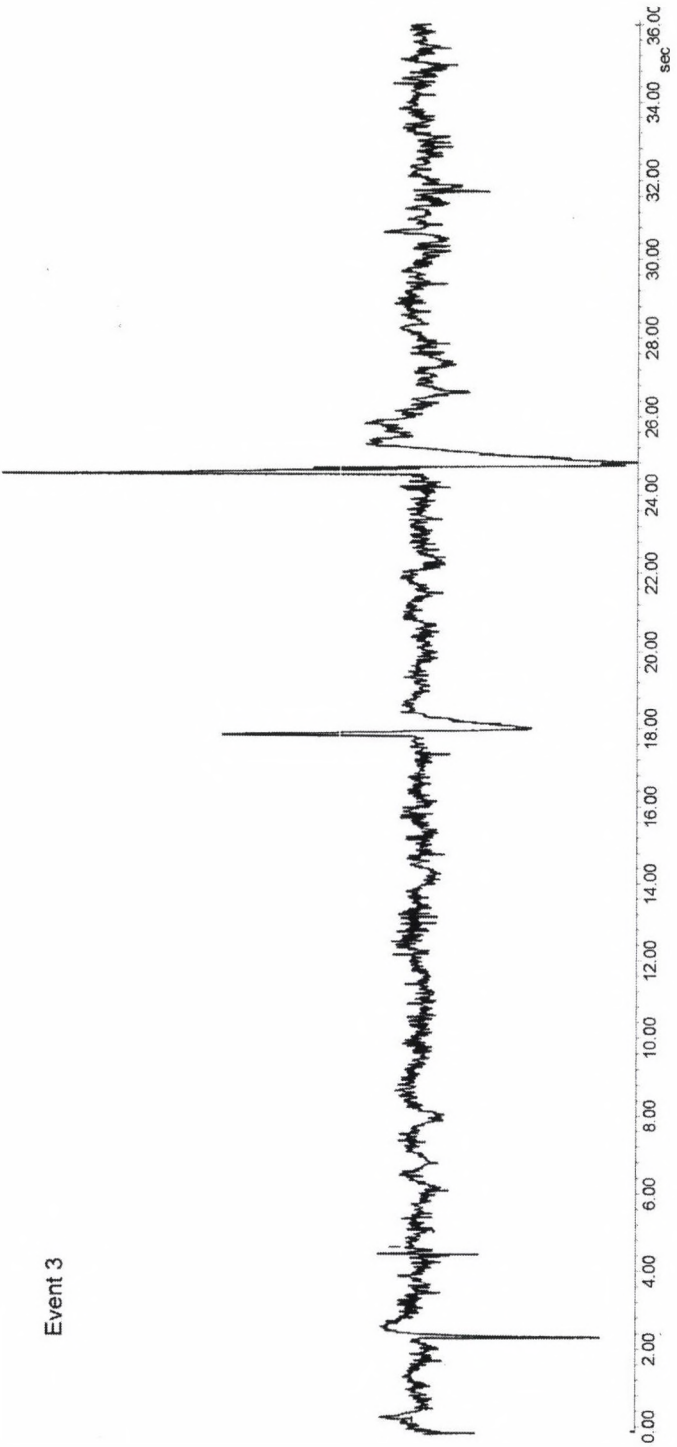


Fig. 23. Analogue record of Event 3, July 18, 1995, 15:29:45 LT, Nagycenk (WEST9), sampling as in Fig. 6

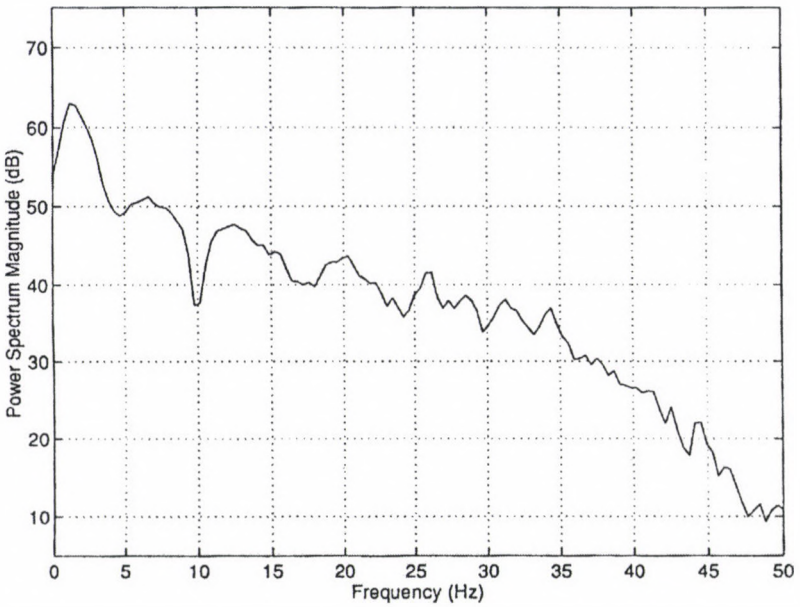


Fig. 24. Spectrum of Event 3, computed using FFT in 256 data long sections

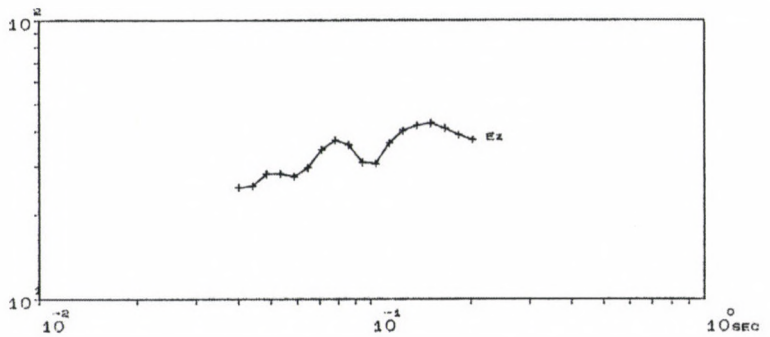


Fig. 25. Spectrum of Event 3, computed with convolution filters like in Fig. 8

4. Conclusions

We compared two simple methods for the (automatic) determination of spectra, namely FFT and convolution filtering (wavelet analysis with box-car filters). Theoretically both methods yield correct values for the amplitude/power of the components, nevertheless, practically there are several limitations for both methods.

In our opinion the greatest advantage of the convolution filtering is its versatility, namely that time and frequency resolutions can be selected more freely than in the case of FFT. In FFT, the length (in time) of the record sets a limit to the frequency resolution, too, as harmonics of the complete length are computed, cor-

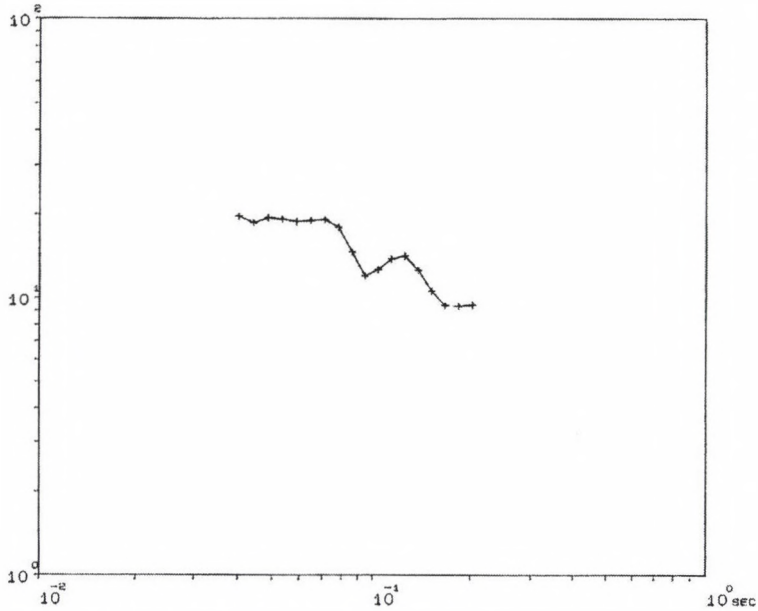


Fig. 26. Spectrum of the short noise-free section of Event 3, computed with convolution filters like in Fig. 8

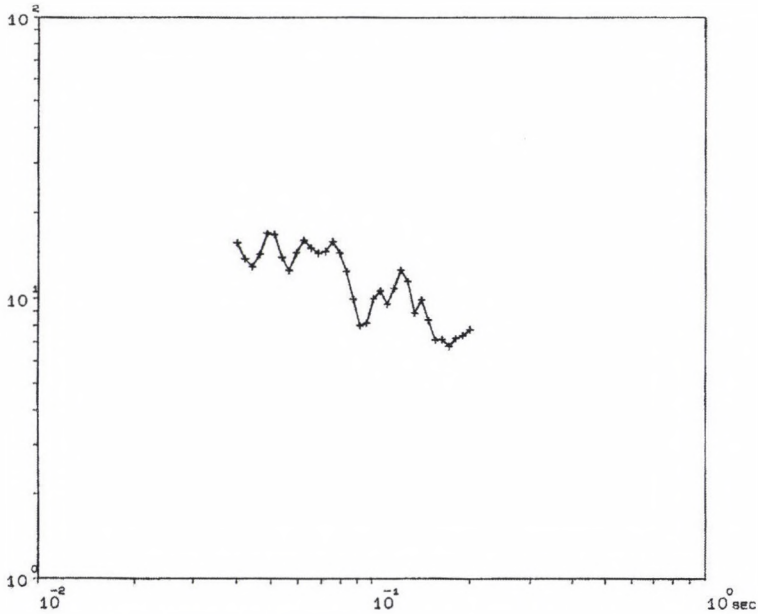


Fig. 27. Spectrum of the short noise-free section of Event 3 computed with convolution filters like in Fig. 26, but the spacings of the filters and the bandwidths are here 5 percent, with 20 cycles long filters

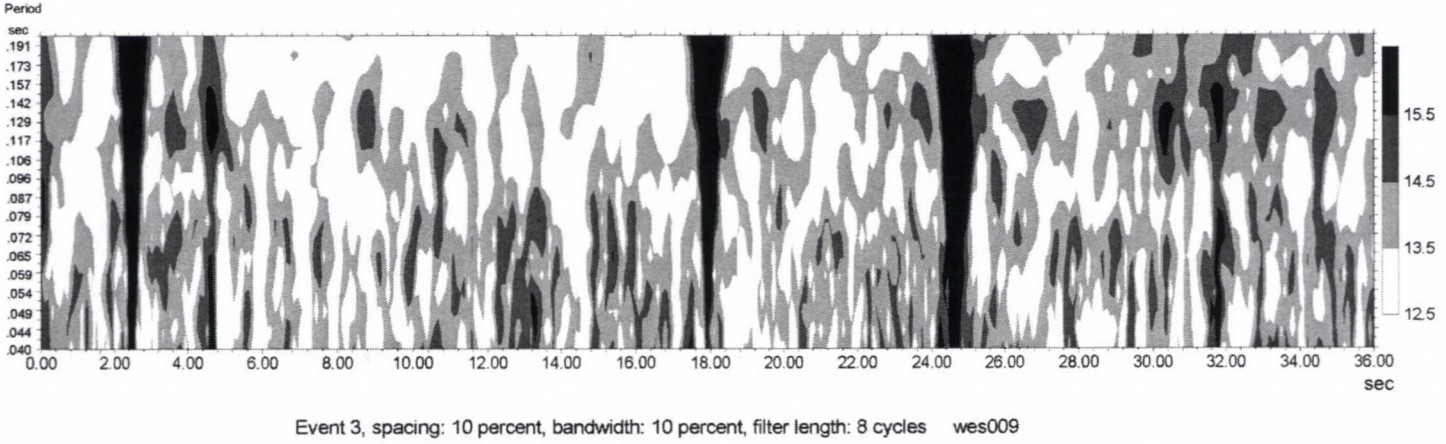


Fig. 28. Dynamic spectrum of Event 3 from convolution filtering, filters like in Fig. 8

respondingly the distance between adjacent lines decreases quickly with increasing frequency/decreasing periods. In convolution filtering, filters with equally spaced central frequencies can be used, independently of the length of the record analysed, in bargain for some compromises regarding dependence/independence of the adjacent values. In addition, the resolution in one direction can only be increased at the expense of the other resolution. It is to be emphasised that the original record can be unambiguously restituted from FFT spectra, while from spectra computed by convolution filtering this is not the case.

Another point is that impulsive events can be harmful for FFT spectra; namely two impulses in a random distance may result in spectra being very different from the spectrum of a single impulse. If they happen to be in antiphase (and of equal amplitude), the characteristic period of the impulse disappears. Thus, when using FFT for the processing of impulsive events, greatest care should be paid to this circumstance. In the case of convolution filtering, impulses (at least in greater time distance from each other) can be resolved into interfering components, but here again the limits of reliability should be observed.

Amplitude/energy values from both spectra differ if noise is present, or — in the FFT spectrum — several impulsive events occur in the same time window. Both effects decrease FFT amplitudes with respect to the filtered ones, thus e.g. amplitude ratios of the amplitudes of the Schumann resonance harmonics can be hardly compared from the two methods. If amplitudes are to be compared, the physical process, the background of the signals is to be considered when deciding for one or other method.

Acknowledgement

This research was supported by the Hungarian State Grant T023111.

References

Oliver R, Ballester J L, Baudin F 1998: *Nature*, 394, 552–553.

Sátori G, Szendrői J, Verő J 1996: *J. Atm. Terr. Phys.*, 58, 1475–1481.

Verő J 1972: *Acta Geod. Geoph. Mont. Hung.*, 7, 333–351.

Verő J, Lühr H, Vellante M, Best I, Strestik J, Cz. Miletits J, Holló L, Szendrői J, Zieger B 1998: *Ann. Geophysicae*, 16, 34–48.

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 133–141 (2000)

ATMOSPHERIC ELECTRIC PARAMETERS AROUND DAYS WITH INCREASED NUMBER OF CONDENSATION NUCLEI

F MÄRCZ[1]

[Manuscript received August 27, 1999]

Air conductivity and atmospheric electric potential gradient measured at Świder observatory (Poland) between 1965 and 1995 have been investigated around days with increased number of condensation nuclei (NCN). Data determined for the evening hours have been analysed by the superposed epoch method. Two groups of days with increased NCN were selected by distinguishing two different degrees of the increase (group 1: from 50% to 75%, group 2: >75%, as related to the actual monthly average of NCN). These days have been used as key days for independent superposed epoch analyses.

The analyses have shown a decreased air conductivity around the days with increased NCN. A rather significant decrease appears on the key days and its significance depends on the degree of the NCN increase. The latter also seems to influence the duration of the air conductivity decrease after the key days. Using atmospheric electric potential gradient data around the selected key days, two more analyses were carried out. The results show that the potential gradient also responds to the increased NCN: it enhances both in the case of the moderately (group 1) and the highly (group 2) increased NCN. This is in accordance with the decreased air conductivity. A significant enhancement of the potential gradient appears on the key days, however, this diminishes more slowly after the key days than the opposite effect in air conductivity. The potential gradient apparently rises even before the key days, especially in the case of the stronger NCN events.

Keywords: air conductivity; atmospheric electric potential gradient; number of condensation nuclei

1. Introduction

The specific electrical conductivity of air both for positive and negative ions depends on the number of ions and their mobilities. Ions can be distinguished according to their sizes: as small, intermediate and large ions. Actually, small ions play the most significant role in air conductivity because of their relatively large number and larger mobility in comparison with that of the two other kinds of ions as described in several general works on atmospheric electricity (e.g. Chalmers 1957 and Israël 1957). The atmosphere also contains uncharged particles of the size similar to that of the large ions which are called condensation nuclei (or Aitken nuclei). The actual air conductivity can highly be influenced by the number of condensation nuclei. Additionally, air conductivity determined at the ground is

[1]Geodetic and Geophysical Research Institute of the Hung. Acad. Sci., H–9401 Sopron, POB 5, Hungary

1217-8977/2000/$ 5.00

an atmospheric electric parameter which depends on further local, environmental factors.

Based on measurements at different sites, it is known that conductivity in the near surface atmosphere shows a quite regular diurnal variation on fair weather days. The conductivity is generally higher at night than in day-time (showing a maximum around dawn), as e.g. demonstrated by data for Kew by Chalmers (1957). The latter work yields also a graph on the annual (seasonal) variation of air conductivity at Kew with higher values in summer and lower ones in winter. As argued by Chalmers (1957), pollution i.e. an increased number of condensation nuclei should be responsible for the minimum of conductivity in winter. This hints at the fact that the air conductivity is effectively influenced by changes in the number of condensation nuclei at the annual time scale. An example given for another time scale (Reiter 1992) shows that an increase of the number density of Aitken nuclei, due to the smoke from a distant forest fire, induced sudden changes in the air conductivity and the atmospheric electric field, as well as fluctuations appeared even in the air-earth current for some hours. Consequently, changes in condensation nuclei affect the atmospheric electric parameters on a rather short time scale, too.

Taking into account the previously mentioned examples (and further ones for different time scales), the present study especially intends to detect the response of atmospheric electric parameters to events of strongly increased number of condensation nuclei based on the long data series of the Świder observatory (Poland). Changes in the air conductivity and the atmospheric electric potential gradient will be shown on a time scale of several days by synchronizing the actual data around the occurrence of the selected condensation nuclei events.

2. Data and investigation method

In the Geophysical Observatory at Świder, the recording of atmospheric electric parameters and the observation of condensation nuclei started several decades ago. Since then, rather continuous data series are at disposal which were published in yearly reports (for the years between 1957 and 1965 in Prace Observatorium Geofizycznego im. S. Kalinowskiego w Świdrze and since 1966 in Publications of the Institute of Geophysics, Polish Academy of Sciences). In the present study, the number of condensation nuclei, air conductivity and potential gradient data of the period from 1965 to 1995 will be analysed. (This interval has been chosen as it covered three complete solar cycles as well as the year 1965 coincided with the very beginning of standard atmospheric electric measurements at Nagycenk.)

The number of condensation nuclei has been determined at Świder three times daily: early morning, before noon and in the evening. Data of the evening observations (generally from 1810 to 1830 UT) seemed to be favourable for selecting days with increased number of condensation nuclei. Whereas, it is supposed that horizontal and vertical air motions (e.g. due to convection) should less frequently disturb the measurements in the evening than in day-time. Going over the data of the yearly reports, two groups of days were distinguished by using two arbitrary selection criteria. For selecting a day for the group 1, an increase between 50% and

75%, related to the actual monthly mean, was required in the number of condensation nuclei, and the days with an increase of >75% were ranked into the group 2. These days have been used as key days for two independent superposed epoch analyses covering an interval five days before and after the key days. (The number of the investigated epochs was 444 for group 1, and 631 for group 2.) In order to show the behaviour of air conductivity around the key days, evening averages of the air conductivity values (measured between 1700 and 2000 UT) have been determined for each day of the selected epochs. Finally, the mean departures of these averages from the actual monthly means (given in percents) should properly demonstrate the changes in air conductivity associated with events of moderately (group 1) and strongly (group 2) increased number of condensation nuclei.

The atmospheric electric potential gradient has been analysed in a similar way around the days with increased number of condensation nuclei. As in the case of air conductivity, evening averages for the interval 1700–2000 UT were determined and investigated. Nevertheless, a preliminary selection of the potential gradient data was carried out by applying proper criteria. Negative values and those surpassing the actual monthly mean by 100% or more were regarded as disturbed ones, thus they have been discarded from the analysis. Moreover, a potential gradient value missing on the key day or one day after it (for the mentioned or even other reason) prohibited the inclusion of the whole epoch in the analysis. Due to these combined restrictions for data selection, the number of investigated epochs diminished in comparison with those used for air conductivity, however, the remaining ones (for group 1: 284, for group 2: 386) are certainly suitable for deriving reliable results. Again, the mean departures of the evening averages from the actual monthly means will be determined for showing the changes in the potential gradient around days with the moderately (group 1) and the strongly (group 2) increased number of condensation nuclei.

3. Results

3.1 Air conductivity around days with increased NCN

Inspecting the evening data of condensation nuclei which were at disposal from 1965 to 1995, altogether 1075 cases have been found when the number of condensation nuclei (NCN) surpassed the actual monthly mean by at least 50%. The total number of cases has been divided into two groups. In the first group, there were 444 cases with an NCN increase between 50% and 75% (above the actual monthly mean). The 631 remaining cases with an NCN increase of >75% have been ranked among the second group. Figure 1 shows the NCN departures (in percents) from the actual monthly means around the selected events (actually in the interval of 5 days before and after them). The mean increase determined for the strong events surpasses the value of 120%, while it is about the half of that in the case of the moderate events. A small NCN increase also appears one day before and after the events; this is true for both groups. Moreover, a rather weak surplus can be seen at day +2 in the case of strong events.

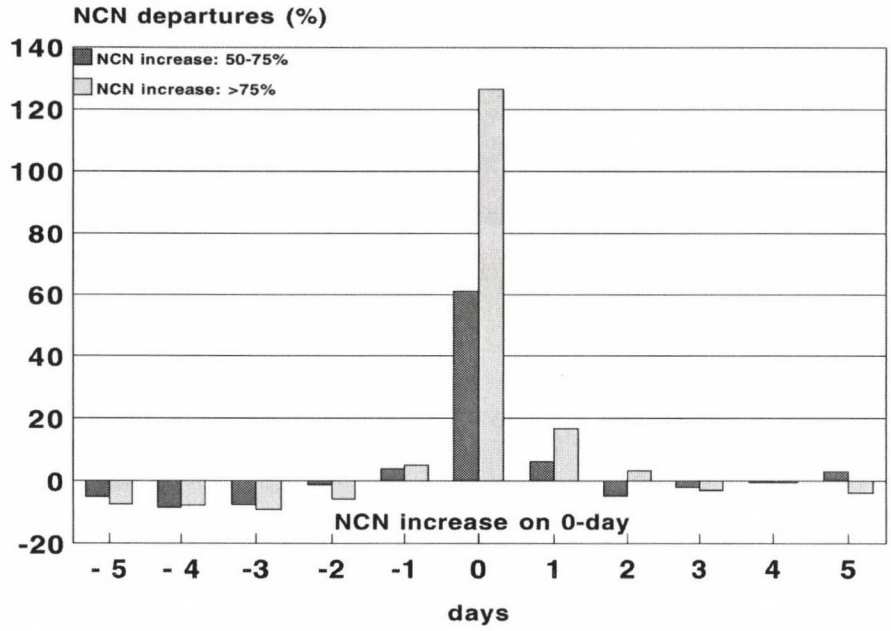


Fig. 1. Mean departures of the number of condensation nuclei (NCN) from the actual monthly average for Świder in the evening; both in the case of moderate (50–75%) and strong (>75%) NCN events on the 0-days. Investigated interval: 1965–1995

Appropriate evening averages of the air conductivity (determined for the interval from 1700 to 2000 UT) have been analysed by the superposed epoch method. At first, the occurrences of moderate events (444 cases) were used for setting key days. Based on experience about the dependence of the air conductivity on NCN (mentioned in the introductory part), it was expected that certain changes in this atmospheric electric parameter might be detected on the applied time scale. Figure 2 shows the mean values of air conductivity departures (in percents) from the actual monthly mean on the selected key days, as well as five days before and after them. Before the key days, the air conductivity hardly differs from the actual average value. Nevertheless, a sharp decrease appears simultaneously with the increased NCN on the 0-day which hints at an immediate response of the air conductivity to the change in the number of condensation nuclei. Air conductivity is quite quickly returning to its normal level, however, it remains somewhat below the actual average on day +1. The decrease of air conductivity on the 0-day is significant in relation to the preceding and following values as indicated by the error bars. (The lengths of the error bars are two standard deviations of the mean.) The results derived by analysing a quite large number of data confirm that air conductivity changes are opposite to those in NCN, on the time scale used here.

The second group consisted of 631 cases with an NCN increase of >75%. Using these occurrences for setting key days, the effect of rather strong condensation nuclei events on air conductivity could be checked. The results of this further superposed epoch analysis are given in Fig. 3. It is shown that a strong decrease

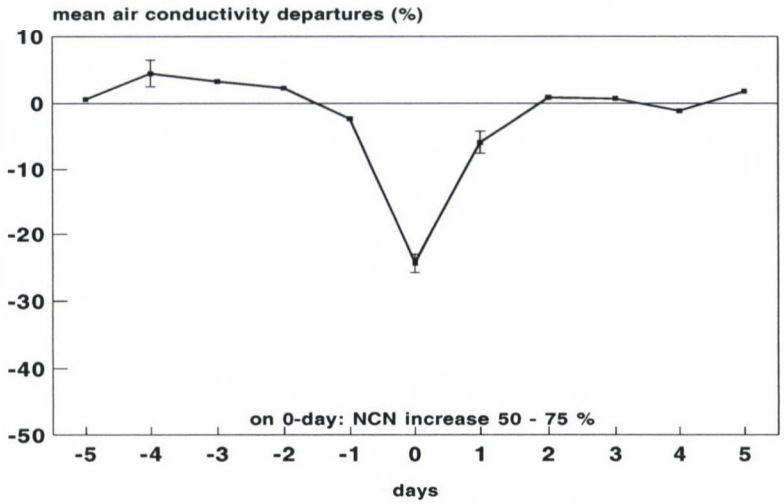


Fig. 2. Mean departures of positive air conductivity (measured in the evening at Świder) from the actual monthly average around days with moderately increased number of condensation nuclei (NCN: 50–75% above the corresponding monthly average). Investigated interval: 1965–1995

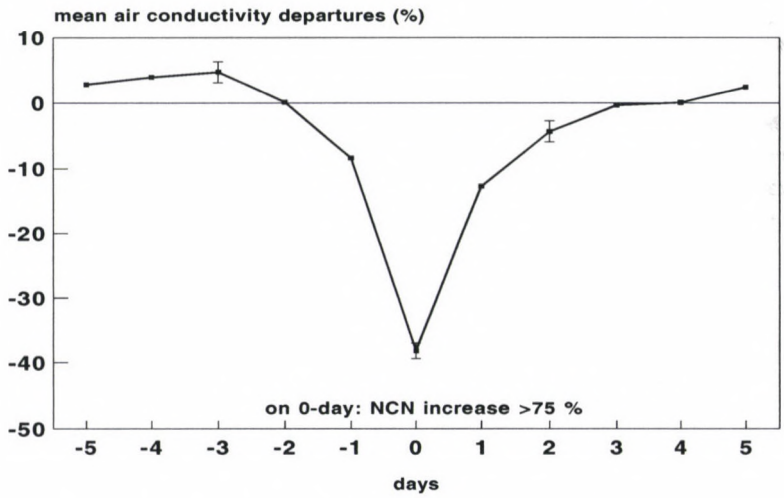


Fig. 3. Mean departures of positive air conductivity (measured in the evening at Świder) from the actual monthly average around days with strongly increased number of condensation nuclei (NCN: >75% above the corresponding monthly average). Investigated interval: 1965–1995

of air conductivity again appears on the key days (i.e. simultaneously with the increased NCN events) and this decrease is more significant than that in Fig. 2. (Error bars are given in a similar way as in Fig. 2.) Moreover, air conductivity returns to its normal level only on day +2, thus the effect of strong NCN events on air conductivity is more durable. A small negative departure also appears on day

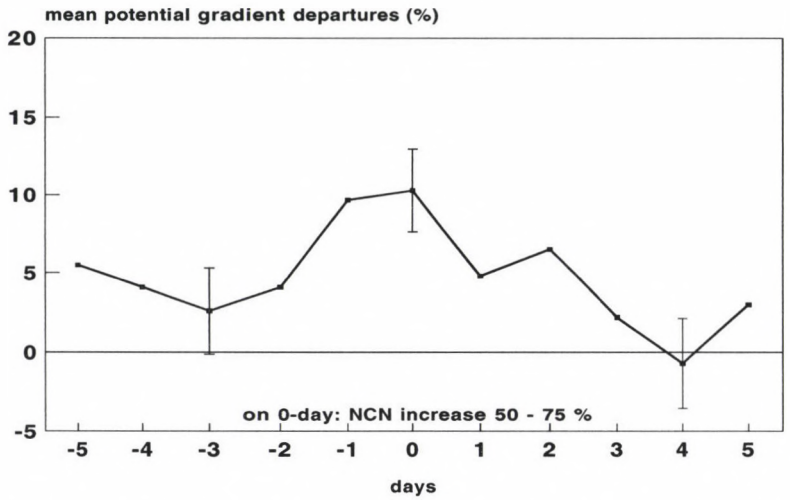


Fig. 4. The same as in Fig. 2, but for the atmospheric electric potential gradient

−1 which can partly be attributed to the fact that NCN is generally increased to a certain extent even before the selected key days.

3.2 Potential gradient around days with increased NCN

Using the combined selection criteria described in Section 2, the behaviour of the atmospheric electric potential gradient (PG) has been analysed by the superposition of 284 epochs in the case of moderately increased NCN events. Figure 4 shows the mean departures of PG evening averages from the actual monthly means around the selected key days (NCN increase: 50–75%). An enhanced PG appears one day before the selected days and there is a PG maximum on the key days followed by a return to the normal PG level on the successive days. The PG increase preceding the key day might be attributed to the fact that even NCN (being the highest on the selected 0-days) began to increase one day earlier. For indicating the significance of the results, error bars are given at appropriate sections of the curve displayed in Fig. 4.

The number of stronger NCN events (with increase of >75%) was 386. The PG variations around these selected events are presented in Fig. 5. An enhancement of the PG starts two days before the key day, however, the maximum increase again appears on the 0-day, and it is more significant than that found in the case of moderate NCN events (Fig. 4). The return to the normal PG level is quite similar to that shown in Fig. 4. (Error bars are drawn in the same way as in the case of the previous figures.)

According to experience, changes in the number of condensation nuclei influence the local environmental conditions and they should be quite effective factors regarding the atmospheric electric phenomena at a given site. The changes in the number of condensation nuclei could also lead to variations in the potential gradient mea-

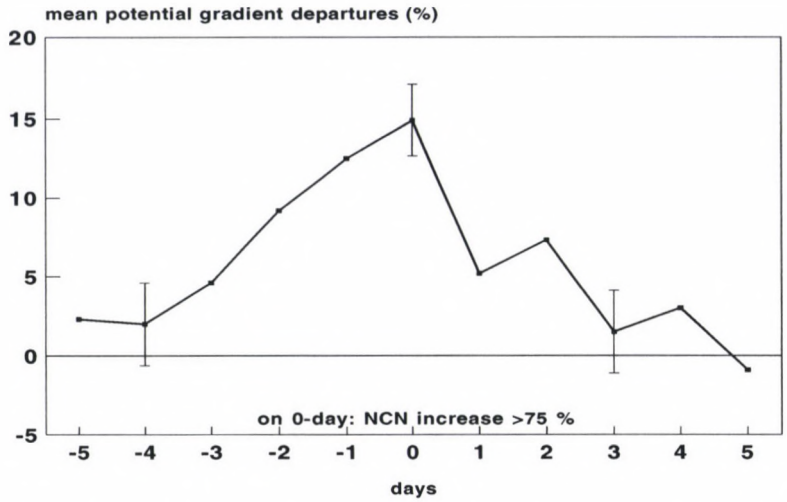


Fig. 5. The same as in Fig. 3, but for the atmospheric electric potential gradient

sured at the ground. The results presented in Figs 4 and 5 allow us to conclude that these local factors really play an important role in the enhancement of the atmospheric electric potential gradient. A comparison of the results on air conductivity (Figs 2 and 3) with those on PG (Figs 4 and 5) confirms that the two atmospheric electric parameters respond in an opposite way to changes in condensation nuclei.

4. Summary and conclusions

As can be learned from several basic works, the simultaneous measurement of different atmospheric electric parameters is reasonable for checking the interrelation of the individual parameters. The field, the air-earth current density and the total air conductivity are the three basic elements of atmospheric electricity. These elements might be interrelated by Ohm's Law, however, only in that case if the measurements are carried out at the same place on a plane area, under steady state conditions. The fulfilment of all these requirements is rather difficult at the individual locations of the measurements. At Świder station, environmental conditions favourable for the measurement of certain atmospheric electric parameters are generally common. There is no major industry in Świder and in the town Otwock situated nearest to it. Thus, the air at the measuring site is certainly not a severely polluted one. The measuring and recording instruments are located in an area overgrown by pine and other (deciduous) trees with a few clearings. In spite of these advantageous circumstances, the condensation nuclei content of the air sometimes distinctly differs from the normal one, as shown by the observations. The potential gradient and the air conductivity of positive polarity are the elements out of the three basic ones which are continuously recorded, while the number of condensation nuclei has only been measured three times daily. In the present paper, data determined in the evening interval have been chosen for studying the interrelations between condensation nuclei and the recorded atmospheric electric elements.

In the case of our initial analyses, two different groups of key days have been used. Selecting days for the first group, a so-called moderate increase (between 50% and 75% above the average) in the number of condensation nuclei was prescribed. On the selected days, the air conductivity of positive polarity was by about 25% lower than the average, as shown in Fig. 2. It is known from earlier measurements (Israël 1957, Table 8), that the value of air conductivity of negative polarity is generally similar to that of positive polarity. Consequently, it can be assumed that the changes in the two polar conductivities (associated with the increase of the condensation nuclei content of the air) should also be similar. Figure 3 reveals that a stronger increase of the number of condensation nuclei (>75%) results in a more decreased air conductivity. This is especially confirmed by the results on the key days: air conductivity is significantly below the average level (by about 40%). The sense of changes in air conductivity due to the enhanced number of condensation nuclei is in accordance with the expectations; moreover the relation between the two parameters is rather distinct. Nevertheless, it should be mentioned that further factors might also play an important role in conductivity changes. Thus, interrelations between small ions (being mainly responsible for air conductivity) and other charged, as well as uncharged particles should sometimes lead to quite complicated conditions.

As the influence of condensation nuclei is important on the small ion density and, from this, on electric conductivity, it has been assumed that the potential gradient should also departure from the normal value if air conductivity is changing. The results presented in Figs 4 and 5 do confirm this assumption. Namely, on days with increased number of condensation nuclei (used as key days), the decrease of air conductivity (seen in Figs 2 and 3) is indeed accompanied by a significant enhancement of the potential gradient measured at the ground (as shown in Figs 4 and 5). In the case of strongly increased number of condensation nuclei (>75%), the mean enhancement of the potential gradient mounts up to about 15% in relation to its average value (on 0-days in Fig. 5). There is an almost steady increase of the potential gradient on days before the NCN events and a gradual decrease after them. (The former is partly due to the somewhat increased number of condensation nuclei even before the selected NCN events.)

The interrelation between the air conductivity and the potential gradient might be described by Ohm's Law under the circumstances mentioned in the beginning of the present Section. The results derived in this study convincingly show that the decrease of air conductivity resulting from the increased number of condensation nuclei is generally associated with an increase of the potential gradient. Thus, on the investigated time scale, the sense of both changes is in accordance with Ohm's Law. Nevertheless, the complexity of the interrelations prevents us determining an accurate quantitative proportionality between the individual changes. Some of our earlier and recent studies (e.g. Märcz 1976, 1990, 1997) dealt with the relation between extraterrestrial events and atmospheric electric parameters. The detection of extraterrestrial effects in atmospheric electric parameters is a quite hard task since they are masked by local influences which are usually stronger than the extraterrestrial ones. In spite of this fact, certain relations could be revealed due to

the appropriate data selection used. Results of the present study seem to confirm that changes in atmospheric electric parameters associated with local factors are well detectable, however, sometimes an accurate quantification of the interrelations existing between the individual components might also be difficult even in these cases.

Acknowledgements

The author is grateful to the colleagues of the Institute of Geophysics, Polish Academy of Sciences, for providing the regularly published observatory reports of their institute. The investigations were supported by the Hungarian State Grant T 023111.

References

Chalmers J A 1957: Atmospheric Electricity. Pergamon Press, London, Paris, New York

Israël H 1957: Atmosphärische Elektrizität, Teil I, Grundlagen, Leitfähigkeit, Ionen. Akademische Verlagsgesellschaft, Geest and Portig K-G, Leipzig

Märcz F 1976: *J. Geoph. Res.*, 81, 4566–4570.

Märcz F 1990: *Ann. Geophysicae*, 8, 525–530.

Märcz F 1997: *J. Atm. Solar-Terr. Phys.*, 59, 975–982.

Prace Observatorium Geofizycznego im. S. Kalinowskiego w Świdrze, Warsaw, 1965

Publications of the Institute of Geophysics, Polish Academy of Sciences, Warsaw, 1966–1995

Reiter R 1992: Phenomena in atmospheric and environmental electricity. Elsevier, Amsterdam, London, New York, Tokyo

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 143–148 (2000)

APPLICATION OF HANKEL TRANSFORM IN THE INTERPRETATION OF MAGNETIC ANOMALY OVER SEMI-INFINITE VERTICAL ROD

N P SINGH[1], T LAL[1], M BANERJEE[1]

[Manuscript received February 3, 1999]

Application of Hankel transform in the interpretation of magnetic anomaly, is illustrated for a vertically magnetized semi-infinite vertical rod. Theoretical anomaly curves, generated using linear digital filter algorithm for the three synthetic model, have been used for the estimation of depth and magnetic moment using the Hankel transform approach. The interpreted results obtained by the Hankel transform technique match significantly well with those selected for the generation of synthetic anomaly curves. Comparison of Hankel transform results with the corresponding Fourier transform results, indicates that the application of Hankel transform is more suitable for the estimation of depth and magnetic moment of a vertically magnetized semi-infinite vertical rod.

Keywords: digital filter algorithm; Hankel transform; magnetic anomaly; semi-infinite vertical rod

1. Introduction

Analyses of gravity and magnetic data are generally performed with the help of Fourier transform technique. Application of Fourier transform in the analysis of magnetic anomalies has been made by a number of researchers (Dean 1958, Solovyev 1962, and others). The use of Hankel transform for the inversion of gravity data was suggested for the first time by Meinardus (1970). Later, Patella (1980) applied Hankel transform for the interpretation of gravity anomalies over some regular shape bodies like sphere, cylinder and vertical rod, and pointed out that for the case of sphere and vertical rod, application of Hankel transform is more suitable than the Fourier transform.

In the present paper, an attempt has been made to investigate the applicability of Hankel transform in the interpretation of magnetic anomaly over a semi-infinite vertical rod, magnetized vertically along its axis. The character of magnetic anomaly is generally determined by the parameters of the causative body: depth, magnetic moments and its size. Normally, the physical parameters are determined by the transform methods provided the spectrums of such anomalies contain information regarding these variables. To verify the application of Hankel transform for magnetic case, synthetic anomalies generated using the linear digital filter algorithm (Anderson 1979), have been interpreted for estimating the model parameters. The comparison of Hankel transform with Fourier transform show that Hankel transform is more amenable to easier evaluation of the parameters.

[1]Department of Geophysics, Banaras Hindu University, Varanasi–221 005, U.P., India, e-mail: npsingh@banaras.ernet.in

1217-8977/2000/$ 5.00

2. Theory and methodology

Application of the Hankel transform in the interpretation of magnetic anomaly is based on the use of Hankel's inversion theorem. Hankel's inversion theorem states that (Erdelyi et al. 1954, Koefoed 1968, and Patella 1980),

$$\text{if } P(x) = \int_0^\infty H(\lambda) \cdot J_n(\lambda x) \cdot \lambda d\lambda \tag{1}$$

$$\text{then } H(\lambda) = \int_0^\infty P(x) \cdot J_n(\lambda x) \cdot x dx \,. \tag{2}$$

The functions $P(x)$ and $H(\lambda)$ are called Hankel transform pairs. In Eqs (1) and (2), λ is the frequency in cycles per unit length.

The vertical component of the magnetic field at a point due to a semi-infinite vertical rod, magnetized vertically along its axis, is written as following (Telford et al. 1976).

$$Z(x) = \frac{IS \cdot h}{(x^2 + h^2)^{3/2}} \,, \tag{3}$$

where I is the intensity of magnetization, S is the cross sectional area of the rod, h is the depth to the top of the rod, and x is the distance of the point of measurement.

The Hankel transform pair of expression (3) can written as

$$Z(x) = \int_0^\infty \left[IS \cdot e^{-\lambda h}\right] \cdot J_0(\lambda x) \cdot \lambda d\lambda \tag{4}$$

$$H_0^R(\lambda) = IS \cdot e^{-\lambda h} = \int_0^\infty Z(x) \cdot J_0(\lambda x) \cdot x dx \,, \tag{5}$$

where $H_0^R(\lambda)$ is the Hankel transform of order 0 for the rod.

On taking logarithm of Eq. (5), we get

$$\ln H_0^R(\lambda) = \ln IS - \lambda h \,. \tag{6}$$

On taking derivative w.r.t. λ of the both sides of Eq. (5), one would get

$$H_1^R(\lambda) = ISh \cdot e^{-\lambda h} = \int_0^\infty Z(x) \cdot J_1(\lambda x) \cdot x^2 dx \,. \tag{7}$$

Here $H_1^R(\lambda)$ is the Hankel transform of first order for the rod. Further, on taking logarithm of Eq. (7), we get

$$\ln H_1^R(\lambda) = \ln(ISh) - \lambda h \,. \tag{8}$$

The Eqs (6) and (8) represent a linear relationship. A plot of $\ln H_0^R(\lambda)$ versus λ, or $\ln H_1^R(\lambda)$ versus λ, on a linear scale would provide a straight line with the slope equal to the negative of depth, and the intercept related to the intensity of magnetization and cross sectional area of the rod. These straight lines are parallel, and the shift between them is equal to the natural logarithm of the depth of the rod. In this study, the plot of $\ln H_0^R(\lambda)$ versus λ has been used for the study of the application of Hankel transform in the interpretation of magnetic data. It can be further seen that the gradient of the lines for any two values of the λ, (i.e. for λ_1 and λ_2) can be given as

$$\frac{\ln H_0^R(\lambda_2) - \ln H_0^R(\lambda_1)}{(\lambda_2 - \lambda_1)} = -h \tag{9}$$

and the intercept value (for $\lambda = 0$) can be written as

$$H_0^R(0) = IS\,. \tag{10}$$

3. Numerical results and discussion

To illustrate the application of Hankel transform, theoretical curves for $H_0^R(\lambda)$ versus λ, computed using the linear digital filter algorithm (Anderson 1979), are interpreted for estimating the model parameters: depth and magnetic moment of the vertically magnetized semi-infinite vertical rod. The method has been tested on three synthetic anomalies computed with the depths of burial as $h = 1000$ m for the model A, $h = 2000$ m for the model B, and $h = 3000$ m for the model C. The magnetic moment for each of the model has been chosen as $IS = 10^6$ ampere-turn-meter. The plots of the vertical component of magnetic field, $Z(x)$, with the horizontal distance, x, are shown in Fig. 1 for the synthetic models given in the inset of the figure. The transform values $H_0^R(\lambda)$ plotted against λ, for each of the models A, B and C, are presented in Fig. 2.

From the Fig. 2, the negative of the gradient of the lines gives the depth of burial of the magnetic rods as 1000.006 m and 3000.004 m for the models A, B and C respectively. From the intercepts of the lines, which correspond to the single intercept value $\ln(IS) = 13.8159$, the magnetic moment of the rod is obtained as 1000089.4 ampere-turn-m. These results show that the interpreted values of depth and moment are quite close to the values taken for the computation of synthetic anomalies.

Equation (3), is an even function, and its Fourier transform can be given (Odegard and Berg 1965) as

$$F^R(\lambda) = \frac{2IS\lambda}{\sqrt{2\pi}} K_1(\lambda h)\,, \tag{11}$$

where $K_1(\lambda h)$ is the Bessel function of the second kind and order 1.

For the comparative study of the application of Hankel and Fourier transform in the interpretation of magnetic anomaly over semi-infinite vertical rod, the plots

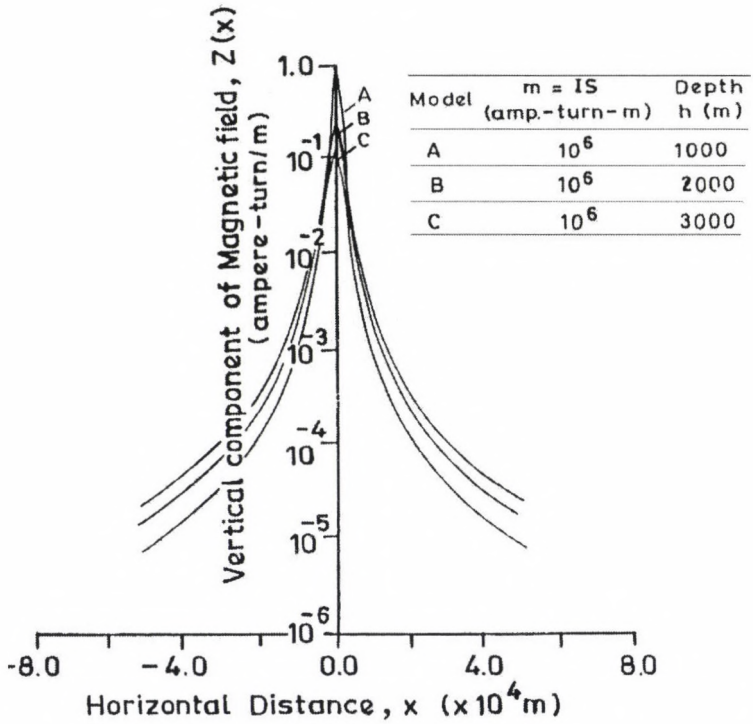


Fig. 1. Variation of vertical component of magnetic field, $Z(x)$ with horizontal distance, x for semi-infinite vertical rod models

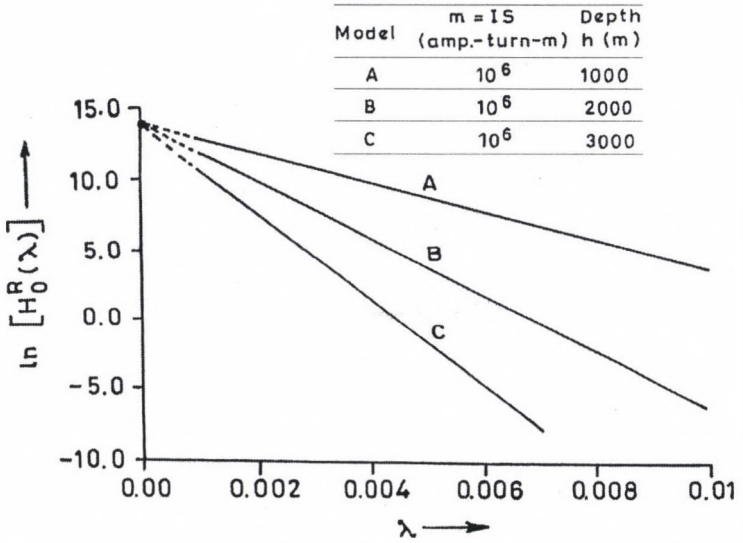


Fig. 2. Plot of the Hankel transform of the vertical component of magnetic field of semi-infinite vertical rod, $H_0^R(\lambda)$, versus λ

Model	m = IS (amp.-turn-m)	Depth h (m)
A	10^6	1000
B	10^6	2000
C	10^6	3000

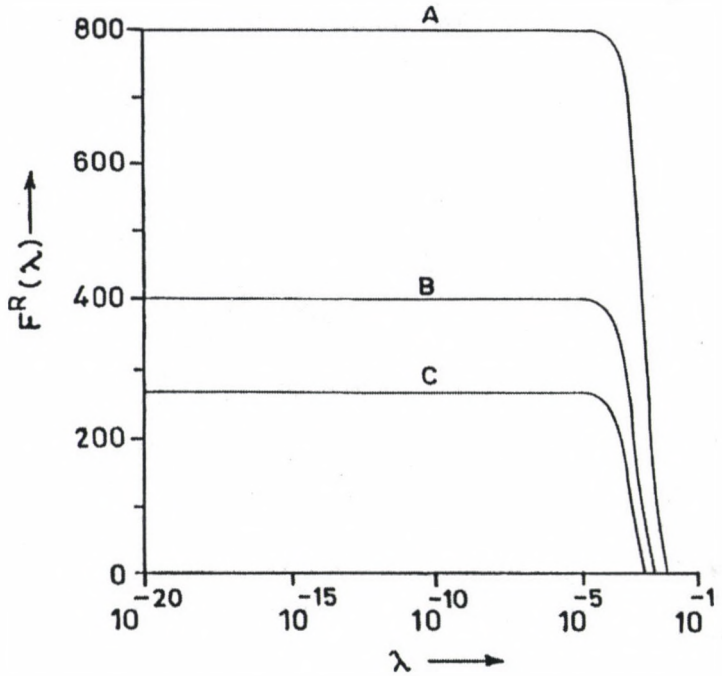


Fig. 3. Plot of the Fourier transform of the vertical component of magnetic field of semi-infinite vertical rod, $F^R(\lambda)$, versus λ

of the $F^R(\lambda)$ against λ are shown in Fig. 3 for the synthetic models under consideration, shown in inset of Fig. 3. The curves decrease asymptotically for large values of λ, from which the evaluation of the parameters, though possible, would be rather complicated (Odegard and Berg 1965). Therefore, in the present scenario, the application of Hankel transform is easy and preferable than the Fourier transform, for the interpretation of magnetic anomaly over a vertically magnetized semi-infinite rod.

Conclusion

In the present paper the application of Hankel transform has been demonstrated for the interpretation of magnetic anomalies over vertically magnetized semi-infinite vertical rod. Theoretical anomaly curves for three synthetic models are interpreted using the Hankel transform approach for the model parameters: depth and mag-

netic moment. Interpreted results show significant match with those taken for the synthetic models. Fourier transform shows its complicated character and non-comprehensive method for the easier evaluation of the parameters of causative body. The results indicate that the Hankel transform is more suitable than the Fourier transform for interpreting the magnetic anomalies over semi-infinite vertical rod.

Acknowledgements

One of the author (NPS) is thankful to the Council of Scientific and Industrial Research, New Delhi, India for providing the financial assistance in the form of Research Associateship.

References

Anderson W L 1979: *Geophysics*, 44, 1287–1305.

Dean W C 1958: *Geophysics*, 23, 97–127.

Erdelyi A, Magnus W, Oberhettinger F, Tricomi F G 1954: Tables of integral transforms 2. McGraw-Hill Book Company, New York

Koefoed O 1968: The application of Kernel function in interpreting geoelectrical resistivity measurements. Geoexploration Monographs, Series 1, No. 2, Gebrüder Borntraeger, Berlin-Stuttgart

Meinardus H A 1970: *Geophysics*, 35, 358.

Odegard M E, Berg J W 1965: *Geophysics*, 30, 424–438.

Patella D 1980: *Geophysical Prospecting*, 28, 744–749.

Solovyev O A 1962: *Academia Nauk SSSR, Sibirskoy Otedelniye, Geologiya i Geofizika*, 2, 122–125.

Telford W M, Geldart L P, Sheriff R E, Keys D A 1976: Applied geophysics. Cambridge University Press, London

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 149–175 (2000)

IMAGING PROPERTIES OF APPARENT RESISTIVITIES BASED ON ROTATIONAL INVARIANTS OF THE MAGNETOTELLURIC IMPEDANCE TENSOR

L SZARKA[1], M MENVIELLE[2], V V SPICHAK[3]

[Manuscript received August 1, 1999]

Apparent resistivity images of thin sheet like models having a complicated geometry are studied in a wide period-range by means of three-dimensional (3D) numerical modelling. Fifteen different definitions of the apparent resistivity are calculated using five different functions $[|f(\mathbf{Z})|, Ref(\mathbf{Z}), Imf(\mathbf{Z}), f(Re\mathbf{Z})$, and $f(Im\mathbf{Z})]$ of three rotational invariants f of the magnetotelluric impedance tensor $[f(\mathbf{Z}) = Z_1^2 = (Z_{xy} - Z_{yx})^2/4$; $f(\mathbf{Z}) = \det(\mathbf{Z}) = Z_{xx} \cdot Z_{yy} - Z_{xy} \cdot Z_{yx}$; $f(\mathbf{Z}) = \mathrm{ssq}(\mathbf{Z}) = Z_{xx}^2 + Z_{xy}^2 + Z_{yx}^2 + Z_{yy}^2]$. A simple visual analysis of apparent resistivity maps and of sounding curves averaged over and around a conductive (and a resistive) heterogeneity embedded in a homogeneous half space is presented. The imaging properties appear to depend much more on the apparent resistivity definition than on the rotational invariant itself. Except for the very short period range corresponding to the oscillating section of the sounding curves, a robust and regular behaviour of the imaging parameters is observed. In this so called "normal" period range the 3D imaging properties seem to be the best if the apparent resistivity is derived when using the function $Re\mathbf{Z}$, i.e. when computing the rotational invariant with the real parts of the four impedance tensor elements. For apparent resistivities derived from $Re\mathbf{Z}$, a rapid convergence over lateral resistivity contrasts and oscillations with small amplitude over homogeneous areas are actually observed in the apparent resistivity maps. In the period domain, they are characterised by a maximum rate of convergence to the underlying resistivity at longer periods and by a reasonably small standard deviation, even in presence of a near-surface disturbing body.

Keywords: electrical resistivity; electromagnetic methods; magnetotellurics

1. Introduction

The increasing performance of field instruments and computers gives rise to an increasing interest in the electromagnetic 3D problem, in both numerical modelling and field data acquisition and interpretation. Although there exist some tensor decomposition methods (e.g. Bahr 1988, Groom and Bailey 1988, Bahr 1991, Groom and Bailey 1991) there is in particular a need for quick-look interpretation techniques providing very fast, approximate but reliable views of the subsurface structures (Christensen 1996). The main requirement for such a quick-look interpretation is trivial: at whichever observation point, the value of the selected imaging parameter should be as close as possible to the 1D value corresponding to

[1]Geodetic and Geophysical Research Institute of the Hung. Acad. Sci., H–9401 Sopron, POB 5, Hungary, e-mail: szarka@ggki.hu
[2]CETP, 4 Avenue de Neptune, F-94107 Saint Maur des Fosses Cedex, France
[3]Geophysical Research Center, 113105 Moscow, Varshavskoe sh. 8, Russia

1217-8977/2000/$ 5.00

the actual resistivity profile just below the point. This implies a rapid convergence over lateral resistivity inhomogeneities and small oscillations over homogeneous areas, thus ensuring the anomaly maps reflect the shape of the subsurface target, and allowing a fairly accurate determination of the resistivity profile. The latter convergence- and oscillation-criteria can be regarded as a simple 3D re-formulation of the criteria set up by Spies and Eggers (1986). According to them an optimal 1D apparent resistivity sounding curve in the very short period range (that is in the so-called "oscillating" period domain) should have a minimum oscillation preceding the transition, and at longer periods (that is in the so-called "normal" period range) a maximum rate of convergence to the underlying resistivity, as a function of the period.

The use of rotational invariants of the magnetotelluric impedance tensor provides very handy tools for quick-look imaging of multidimensional structures, because these quantities are in fact independent of the directions of both the inducing electromagnetic field and recording axes. The 2D behaviour of some rotational invariants has already been investigated (see e.g. Ranganayaki 1984, Ingham 1988, Park and Livelybrooks 1989), thus leading to conclusions widely accepted by the induction community. These works resulted in focusing on the absolute value of some rotational invariants, as illustrated by the recent review by Weaver (1994) where the 3D apparent resistivity definitions are restricted to the "effective" apparent resistivity based on the absolute value of the determinant, and the "average" apparent resistivity based on the absolute value of the difference of the main impedance elements.

More recently, some experiences with tensor decomposition handling the real and the imaginary tensors separately found in field data (e.g. by Ingham and Brown 1998) that the real tensor based quantities are more efficient and reliable than the imaginary tensor based ones. These new experimental results clearly gave further evidence that the various interpretation parameters used in induction studies have not the same performances. But they do not allow to decide whether the better performances of the real tensor based interpretation parameters is a general feature or not. A numerical experience, based upon a systematic study of the imaging performances of various magnetotelluric interpretation parameters may provide useful information to answer this question.

The use of rotational invariants in 3D situations has been addressed by Szarka and Menvielle (1997). They showed that separating the elements of the magnetotelluric impedance tensor into their real and imaginary parts and considering rotational invariants computed with these quantities actually yielded significant experimental results. At the studied period (T = 10 s) it was observed that the geometry of a thin-sheet like 3D target having a complicated geometry is much better reflected in quantities computed exclusively with the real part of the tensor elements, that is with rotational invariants corresponding to the real-valued tensor $Re\mathbf{Z}$.

The present work aims at investigating the imaging properties of different rotational invariants throughout a wide magnetotelluric period range instead of at one single period, with particular attention to both their lateral- and depth-resolution power. Fifteen apparent resistivity definitions based on rotational invariants are

considered. Their choice is discussed in Section 2. The numerical models used for their computation are also briefly described in this Section. Their imaging properties are discussed in Section 3 through a visual comparison of maps and sounding curves, which allow to characterise both their horizontal and vertical imaging performances. Finally, the conclusions are summarised in Section 4.

2. Selection and computation of the imaging quantities

2.1 The use of apparent resistivity

For the sake of tradition and simplicity let us only consider apparent resistivities. These quantities, having a dimension of Ωm, are quite easy to interpret, and actually they have widely been used since the fifties (Tikhonov 1950, Cagniard 1953). The apparent resistivity is proportional to the product of the period by a quantity having the dimension of a squared impedance and there are accordingly as many different possible definitions of the apparent resistivity as real-valued quantities having the dimension of a squared impedance (see the Appendix).

In 1D situations, the inductive response of the conductive medium can be characterised by a scalar complex-valued impedance. Cagniard (1953) derived the apparent resistivity from the modulus of the impedance. Other definitions have then been introduced afterwards, and it is now well known that the apparent resistivity behaviour depends already in 1D medium on the used function of the impedance (see e.g. Spies and Eggers 1986, Schmucker 1987).

In 3D situations, the inductive response of the medium is characterised by the 2×2 complex-valued impedance tensor, and any second degree real-valued function of the elements of this tensor can be used to compute apparent resistivities. Since we are interested in imaging quantities which do not depend on the directions along which the tensor is expressed, let us consider apparent resistivities derived from functions of the elements of the tensor which are rotational invariants.

2.2 The selection of rotational invariants

Consider apparent resistivities derived from the three following different rotational invariants $f(\mathbf{Z})$ having the dimension of a squared impedance:

$$Z_1^2 = (Z_{xy} - Z_{yx})^2/4\,, \tag{1a}$$

$$\det(\mathbf{Z}) = Z_{xx} \cdot Z_{yy} - Z_{xy} \cdot Z_{yx}\,, \tag{1b}$$

$$\mathrm{ssq}(\mathbf{Z}) = Z_{xx}^2 + Z_{xy}^2 + Z_{yx}^2 + Z_{yy}^2\,. \tag{1c}$$

It is of fairly common use to characterise in the real plane a function of a complex impedance tensor either by means of its modulus, or by means of its real and imaginary parts. The former characterisation is the more commonly used in the induction community (see e.g., Weaver 1994). However, our results (Szarka and Menvielle 1997) and those of Lilley (1993) suggest to consider separately the real and the imaginary tensors. In order to compare the latter 'new' functions with

the more or less conventional ones, we decided to compute for each complex-valued rotational invariant f the following five real-valued invariants:

$$|f(\mathbf{Z})|,\ Ref(\mathbf{Z}),\ Imf(\mathbf{Z}),\ f(Re\mathbf{Z}),\ \text{and}\ f(Im\mathbf{Z})\,.$$

The resulting set of 15 invariants is given in Table I. Only seven of them can be independent (Szarka and Menvielle 1997). One possible set of seven independent rotational invariants is underlined in Table I. Once selected a set of seven independent rotational invariants, the other eight ones and any further rotational invariant can be expressed from subsets of these independent rotational invariants.

2.3 Computation of apparent resistivities and impedance phases

These 15 real-valued rotational invariants were then transformed into apparent resistivities, using the transformation formula given in Table II. Since apparent resistivities computed from $Ref(\mathbf{Z})$ (column 2) express merely a deviation from the resistivity corresponding to a homogeneous half space, for this case a special apparent resistivity formula was introduced as it stands in Table II.

Apparent resistivities were expressed as a function of relative wavelengths λ_1/h_1, where δ_1 is the skin depth in the hosting homogeneous medium of resistivity ϱ_1, and h_1 is the depth to the top of the conductivity heterogeneity:

$$\lambda_1/h_1 = 2\pi\delta_1/h_1 = (10^7\varrho_1 T)^{1/2}/h_1\,. \tag{2}$$

Several possible definitions of the invariants related to the magnetotelluric phase lag, $\Phi f(\mathbf{Z})$, are summarised in Table III. As expected from its very definition, the phase lag, $\Phi f(\mathbf{Z})$ is always a function of the real-valued $Re\mathbf{Z}$ and $Im\mathbf{Z}$ tensors, but the corresponding expressions are fairly more complicated. For a more detailed discussion about the complexity of apparent resistivity and phase definitions see the Appendix.

2.4 The numerical models

The models we used basically consist in a homogeneous half-space with heterogeneities at a certain depth. The structure to be imaged, hereafter called the target, is a thin sheet like body. Both the case of resistive and conducting targets were considered. In order to investigate the robustness of the imaging power of the rotational invariants to static shift effects, computations have also been made for models with the target and a near-surface disturbing body. Different models, corresponding to different target geometries have been considered, leading to similar results. Therefore we present here the results obtained with a U-shaped target only. Models having such a complicated geometry are rarely studied.

Three different model variants are considered in this paper. Model 1 (a U-shaped target of a resistivity of 1 Ωm at a depth of 1 km, in a homogeneous medium of 50 Ωm) is shown in Fig. 1. Model 2 is similar to model 1 but the resistivity of the target is 2500 Ωm. Finally, model 3 is derived from model 1 by adding a 100 m

Table I. The 15 real valued rotational invariants used in this study

$f(\mathbf{Z})$	$\lvert f(\mathbf{Z})\rvert$	$Ref(\mathbf{Z})$	$Imf(\mathbf{Z})$	$f(Re\mathbf{Z})$	$f(Im\mathbf{Z})$
Z_1^2	$\lvert Z_1\rvert^2 = \lvert Z_1^2\rvert$	ReZ_1^2	ImZ_1^2	$\underline{Re^2 Z_1}$	$\underline{Im^2 Z_1}$
$\det(\mathbf{Z})$	$\lvert\det(\mathbf{Z})\rvert$	$Re\det(\mathbf{Z})$	$\underline{Im\det(\mathbf{Z})}$	$\underline{\det(Re\mathbf{Z})}$	$\underline{\det(Im\mathbf{Z})}$
$\mathrm{ssq}(\mathbf{Z})$	$\lvert\mathrm{ssq}(\mathbf{Z})\rvert$	$Re\mathrm{ssq}(\mathbf{Z})$	$Im\mathrm{ssq}(\mathbf{Z})$	$\underline{\mathrm{ssq}(Re\mathbf{Z})}$	$\underline{\mathrm{ssq}(Im\mathbf{Z})}$

Table II. Apparent resistivity definitions based on 15 real-valued quadratic rotational invariants of the magnetotelluric impedance tensor

$f(\mathbf{Z})$	$\lvert f(\mathbf{Z})\rvert$	$Ref(\mathbf{Z})$	$Imf(\mathbf{Z})$	$f(Re\mathbf{Z})$	$f(Im\mathbf{Z})$
Z_1^2	$(1/\omega\mu)\lvert Z_1\rvert^2$	$\varrho_1 + (1/\omega\mu)ReZ_1^2$	$(1/\omega\mu)\lvert ImZ_1^2\rvert$	$(2/\omega\mu)Re^2 Z_1$	$(2/\omega\mu)Im^2 Z_1$
$\det(\mathbf{Z})$	$(1/\omega\mu)\lvert\det(\mathbf{Z})\rvert$	$\varrho_1 + (1/\omega\mu)Re\det(\mathbf{Z})$	$(1/\omega\mu)\lvert Im\det(\mathbf{Z})\rvert$	$(2/\omega\mu)\lvert\det(Re\mathbf{Z})\rvert$	$(2/\omega\mu)\lvert\det(Im\mathbf{Z})\rvert$
$\mathrm{ssq}(\mathbf{Z})$	$(1/2\omega\mu)\lvert\mathrm{ssq}(\mathbf{Z})\rvert$	$\varrho_1 + (1/2\omega\mu)Re\mathrm{ssq}(\mathbf{Z})$	$(1/2\omega\mu)\lvert Im\mathrm{ssq}(\mathbf{Z})\rvert$	$(1/\omega\mu)\mathrm{ssq}(Re\mathbf{Z})$	$(1/\omega\mu)\mathrm{ssq}(Im\mathbf{Z})$

Table III. Magnetotelluric phase definitions from Z_1, $\det(\mathbf{Z})$ and $\mathrm{ssq}(\mathbf{Z})$

$f(\mathbf{Z})$	$f(Im\mathbf{Z})/f(Re\mathbf{Z})$	$Imf(\mathbf{Z})/Ref(\mathbf{Z})$
Z_1	$\mathrm{atan}(\lvert ImZ_1/ReZ_1\rvert)$	
$\det(\mathbf{Z})$	$\mathrm{atan}(\lvert\det(Im\mathbf{Z})/\det(Re\mathbf{Z})\rvert^{1/2})$	$\mathrm{atan}(\lvert Im\det(\mathbf{Z})/Re\det(\mathbf{Z})\rvert^{1/2})$
$\mathrm{ssq}(\mathbf{Z})$	$\mathrm{atan}(\lvert\mathrm{ssq}(Im\mathbf{Z})/\mathrm{ssq}(Re\mathbf{Z})\rvert^{1/2})$	$\mathrm{atan}(\lvert Im\mathrm{ssq}(\mathbf{Z})/Re\mathrm{ssq}(\mathbf{Z})\rvert^{1/2})$

thick, square-shaped (side length: 1000 m), near-surface (depth: 100 m), conductive (resistivity: 1 Ωm) inhomogeneity.

The electromagnetic field components were computed by using Spichak's (1983) full finite-difference 3D numerical code (for more details see Spichak 1995 and Spichak and Popova 1998) and some tests were carried out by using the Geotools-version of Randie Mackie's 3D code (Mackie et al. 1993). The four complex elements of the magnetotelluric impedance tensor were determined at 21×21 sites (over a 10 km times 10 km area, with a grid distance of 500 m). The responses are shown here for six different periods (T = 0.03s, 0.1s, 0.3s, 1.0s, 3.0s and 10s), corresponding to $\lambda_r = \lambda_1/h_1$ between nearly 4 and 70.

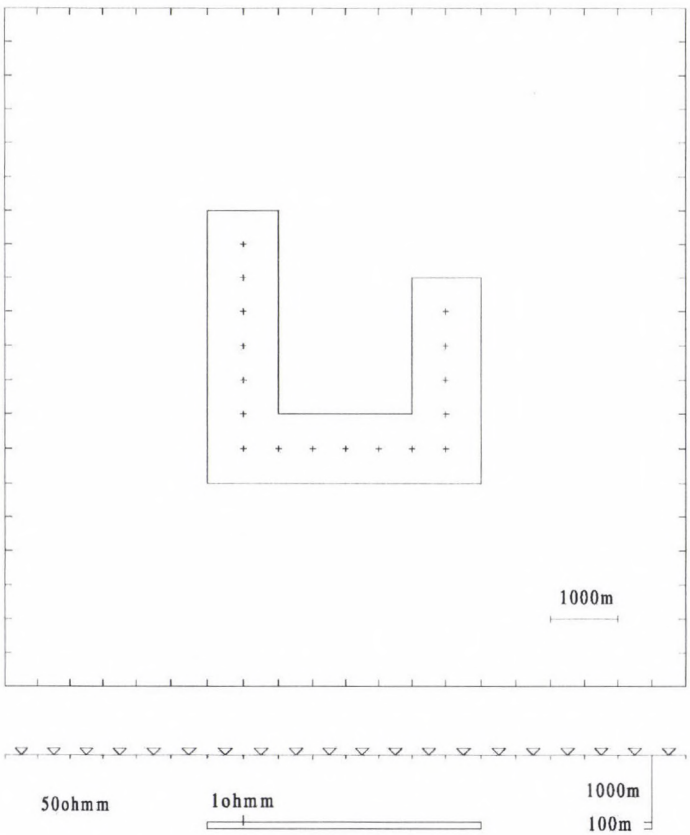


Fig. 1. Plan view and cross section of model 1 (an U-shaped heterogeneity with a resistivity of 1 Ωm). The measurement points are distributed on a 10 km × 10 km wide area. For the computation of sounding curves the 17 sites over the model (inside the contour line of the U-shaped heterogeneity, denoted by crosses) are handled as one group. In case of model 2 the resistivity of the U-shaped target is 2500 Ωm. In case of model 3, at a depth of 100 m an 1 km × 1 km large square of a resistivity of 1 Ωm, is added to model 1. Its location is in the elongation of the shorter leg of the U-shaped target

2.5 Way of presentation of the results

The imaging performances of the rotational invariants are described using both apparent resistivity maps and sounding curves, which allow together a qualitative discussion of both lateral and vertical imaging performances. Apparent resistivity maps have been established for fifteen rotational invariants and for six periods, that is altogether 90 maps for each model. The number of possible sounding curves is $21 \times 21 \times 15$ per model. In order to get a reasonable number of figures, the maps and curves are organised in an as compact way as possible.

The maps are organised in plates, each of them displaying all the maps corresponding to a given function $f[Z_1^2;\ \det(\mathbf{Z});\ \mathrm{ssq}(\mathbf{Z})]$ and a given model. In each plate, they are ordered according to the period (increasing from top to bottom) and the apparent resistivity definition [from left to right: $|f(\mathbf{Z})|$, $Ref(\mathbf{Z})$, $Imf(\mathbf{Z})$, $f(Re\mathbf{Z})$, and $f(Im\mathbf{Z})$].

For each model and for each of the 15 apparent resistivity definitions, four sounding curves are presented:

— the average of apparent resistivities observed at 17 sites, situated over the heterogeneity (indicated by crosses in Fig. 1), with the corresponding standard deviations;

— the average of apparent resistivities observed at 380 sites located outside the heterogeneity over a 10 km $\times$ 10 km area, with the corresponding standard deviations;

— the local 1D response over the heterogeneity, namely that over a three layered half space, with a second layer having the same thickness and resistivity as the U-shaped target;

— the local 1D response around the heterogeneity, namely that over a homogeneous 50 Ωm half space.

The sounding curves are also organised in plates, each of them displaying all the curves corresponding to a given function $f[Z_1^2;\ \det(\mathbf{Z});\ \mathrm{ssq}(\mathbf{Z})]$ and to a given model. Each plate consists in five panels corresponding respectively to the five apparent resistivity definitions we consider [from left to right: $|f(\mathbf{Z})|$, $Ref\mathbf{Z}$, $Imf(\mathbf{Z})$, $f(Re\mathbf{Z})$, and $f(Im\mathbf{Z})$]. In each panel, the variations of the above defined sounding curves are displayed with respect to the relative wavelengths $\lambda_r = \lambda_1/h_1$.

3. Results

3.1 Model 1 (conductive target)

Figures 2a, 3a and 4a show the apparent resistivity maps computed for a conductive (1 Ωm) target embedded in a homogeneous half space (50 Ωm), while the sounding curves are shown in Figs 2b, 3b and 4b. The main results are as follows.

I. In the different columns of the map plates the images are systematically different, but in the corresponding columns of Figs 2a, 3a and 4a the images are nearly the

same. It means that the observed imaging properties significantly depend on the definition of the apparent resistivity but not on the function f. With other words: the imaging properties of the different apparent resistivity definitions displayed in a given column of Table II are similar.

II. The $\lambda_r = \lambda_1/h_1$ range can be divided into two parts: a shorter λ_r range, characterised by irregular imaging behaviours and a longer λ_r range, where the observed images vary quite smoothly and regularly. These two domains correspond respectively to the "oscillating" and to the "normal" period sections in 1D situations (Spies and Eggers 1986). The limit between the shorter and longer λ_r ranges is quite soft (it is somewhere between 6 and 10, depending on the resistivity definition; for details see Szarka 1997). Somewhere at this limit, surprisingly good images are observed. (For details, see the keyhole image technique by Szarka and Menvielle 1999.) A keyhole image example can be seen e.g. for $|f(\mathbf{Z})|$ at $\lambda_r = 7.07$. In this paper we concentrate on the "normal" period range.

III. In the "normal" period domain, each of the five different real-valued functions [from left to right: $|f(\mathbf{Z})|$, $Ref(\mathbf{Z})$, $Imf(\mathbf{Z})$, $f(Re\mathbf{Z})$, and $f(Im\mathbf{Z})$] has a specific and a more systematic signature which can be seen for all the three complex valued functions [Z_1^2; $\det(\mathbf{Z})$; $\mathrm{ssq}(\mathbf{Z})$]. Robust anomaly maps actually appear very soon with increasing λ_r values. They are slowly attenuated and tend toward their long period asymptotes. It is remarkable that the most intense anomalies (e.g. which can be seen for ϱ_{ReZ_1} at $\lambda_r = 12.3$ in Fig. 2b), do not image the structure (look at ϱ_{ReZ_1} at $\lambda_r = 12.3$ or ϱ_{ImZ_1} at $\lambda_r = 22.4$ in Fig. 2a).

IV. Looking at the imaging properties of different apparent resistivity definitions in this "normal" period range, the maps corresponding to apparent resistivities derived from the real tensor $Re\mathbf{Z}$, (column number 4 of each figure) are characterised by a more rapid convergence over lateral resistivity contrasts and a smaller amplitude of the oscillations over homogeneous areas. This is confirmed by the sounding curves presented in Figs 2b, 3b and 4b: the standard deviations are the smallest in the case of apparent resistivities derived from $Re\mathbf{Z}$. At the same time, they are the largest in the case of apparent resistivities derived from $Im\mathbf{Z}$. Furthermore, the resistivity values throughout a wide period range indicate unambiguously only in case of $Re\mathbf{Z}$ if the observation point is located over or around the U-shaped target. The apparent resistivities deduced from $Re\mathbf{Z}$ tends to have a similar long period asymptotic value, smaller than 50 Ωm, at all the observation points located around the target, while the situation is more complicated for apparent resistivities derived from the four other functions.

As a summary it clearly appears from our results that apparent resistivities derived from $Re\mathbf{Z}$ provide the best images of a conductive target, while those derived from $Im\mathbf{Z}$ provide the worst ones.

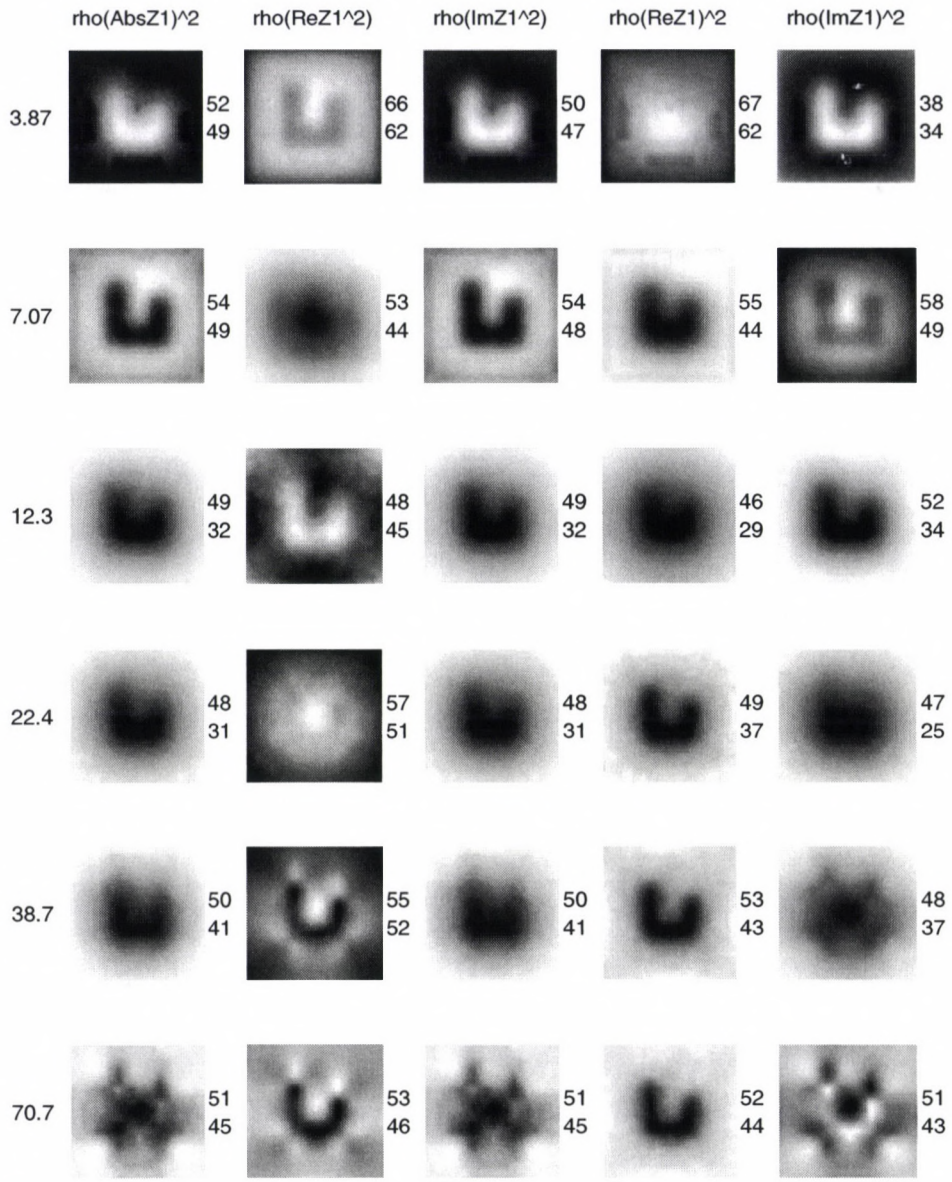


Fig. 2a. Apparent resistivity maps over the U-shaped heterogeneity presented in Fig. 1. They are computed at six different periods T = 0.03 s, 0.1 s, 0.3 s, 1.0 s, 3.0 s and 10 s increasing from top to bottom, and the on the left side of the figure the corresponding λ_r is indicated. The five apparent resistivity definitions based upon $f(\mathbf{Z}) = Z_1^2$, from left to right, are as follows: $|Z_1|^2$, ReZ_1^2, ImZ_1^2, Re^2Z_1 and Im^2Z_1. Each map is produced using the autoshading function 'colormap(gray)' of MATLAB individually. The maximum and minimum values in the right side of each map correspond to the most black and most white regions

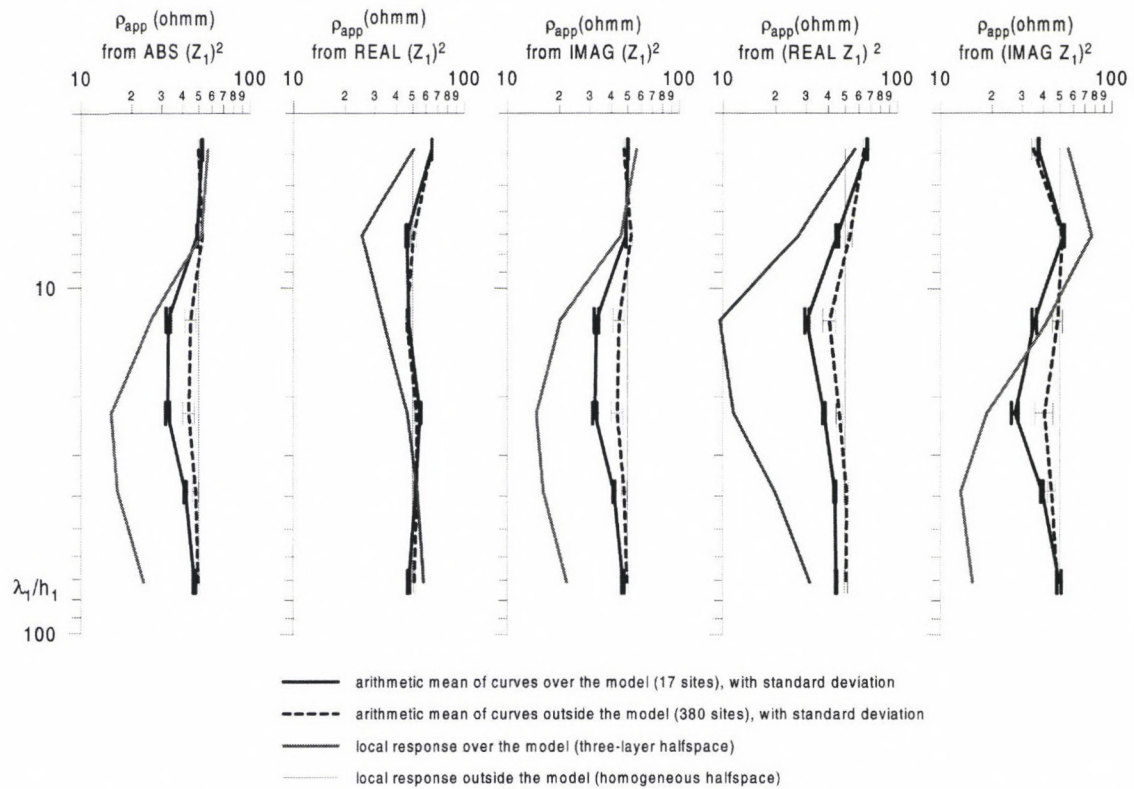


Fig. 2b. Five different apparent resistivity definitions, computed by using $f(\mathbf{Z}) = Z_1^2$, shown in five panels: a) $|Z_1|$, b) ReZ_1^2, c) ImZ_1^2, d) Re^2Z_1, and e) Im^2Z_1: – averaged apparent resistivity curve of 17 sites, situated inside the plan view of the heterogeneity, with the corresponding standard deviations; – averaged apparent resistivity curve of 380 sites, located outside the heterogeneity, with the corresponding standard deviations; – the local 1D response over the heterogeneity, namely that of a three-layer half space, with a second layer having the same thickness and resistivity as the U-shaped heterogeneous structure; – the local 1D response around the heterogeneity, namely that of a homogeneous half space, having a resistivity of 50 Ωm

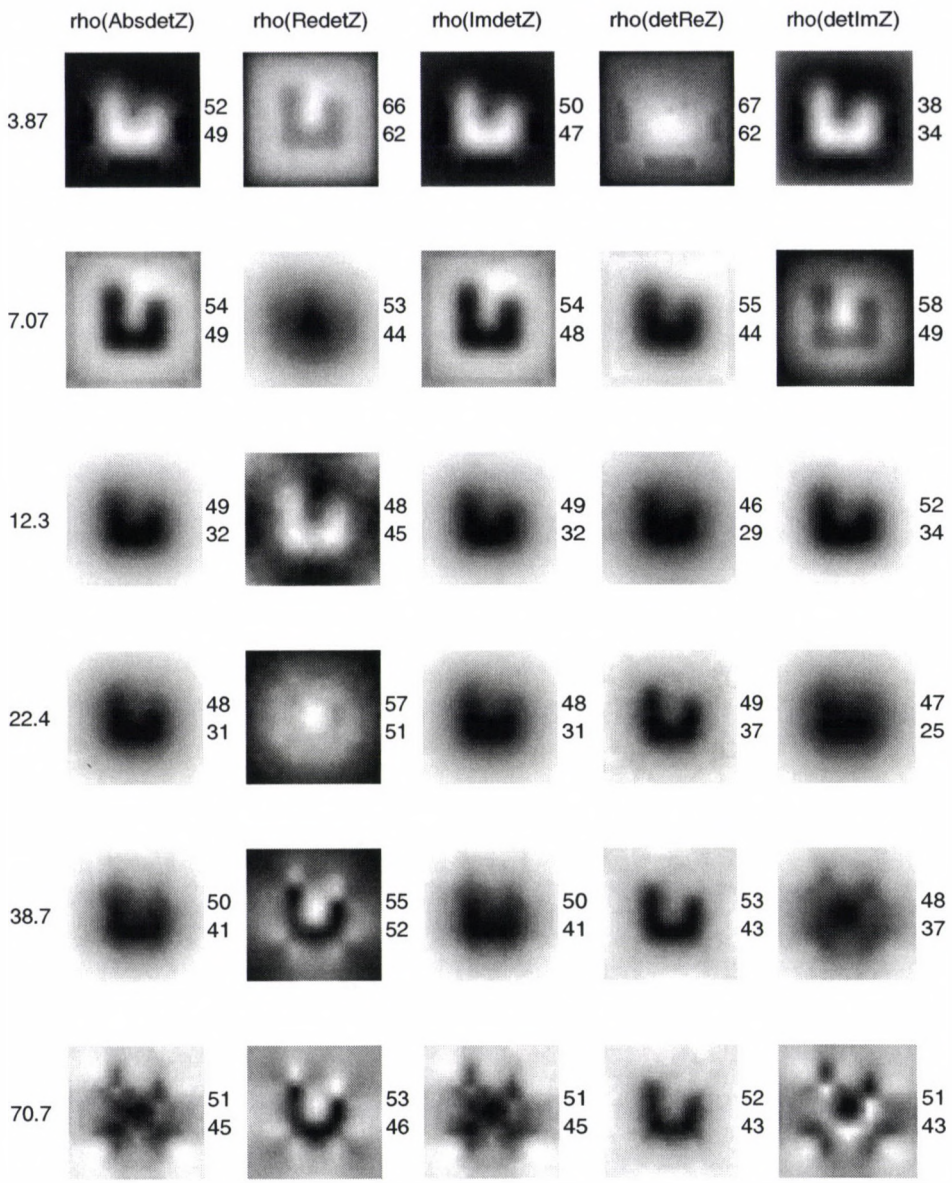


Fig. 3a. Apparent resistivity maps over the U-shaped heterogeneity presented in Fig. 1. They are computed at six different periods T = 0.03 s, 0.1 s, 0.3 s, 1.0 s, 3.0 s and 10 s, increasing from top to bottom, and the on the left side of the figure the corresponding λ_r is indicated. The five apparent resistivity definitions based upon $f(\mathbf{Z}) = \det(\mathbf{Z})$, from left to right, are as follows: $|\det(\mathbf{Z})|$, $Re\det(\mathbf{Z})$, $Im\det(\mathbf{Z})$, $\det(Re\mathbf{Z})$ and $\det(Im\mathbf{Z})$. Each map is produced using the autoshading function 'colormap(gray)' of MATLAB individually. The maximum and minimum values in the right side of each map correspond to the most black and most white regions

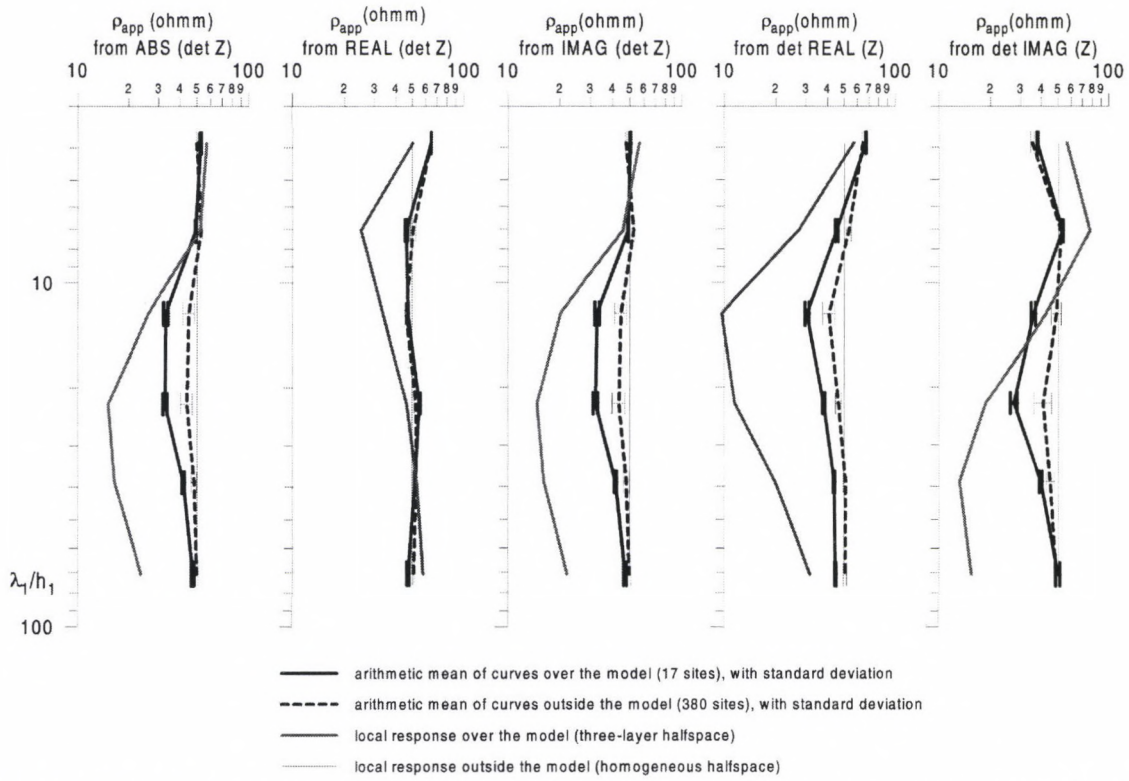


Fig. 3b. Five different apparent resistivity definitions, computed by using $f(\mathbf{Z}) = \det(\mathbf{Z})$ shown in five panels: a) $|\det(\mathbf{Z})|$, b) $Re\det(\mathbf{Z})$, c) $Im\det(\mathbf{Z})$, d) $\det(Re\mathbf{Z})$ and e) $\det(Im\mathbf{Z})$: – averaged apparent resistivity curve of 17 sites, situated inside the plan view of the heterogeneity, with the corresponding standard deviations; – averaged apparent resistivity curve of 380 sites, located outside the heterogeneity, with the corresponding standard deviations; – the local 1D response over the heterogeneity, namely that of a three-layer half space, with a second layer having the same thickness and resistivity as the U-shaped heterogeneous structure; – the local 1D response around the heterogeneity, namely that of a homogeneous half space, having a resistivity of 50 Ωm

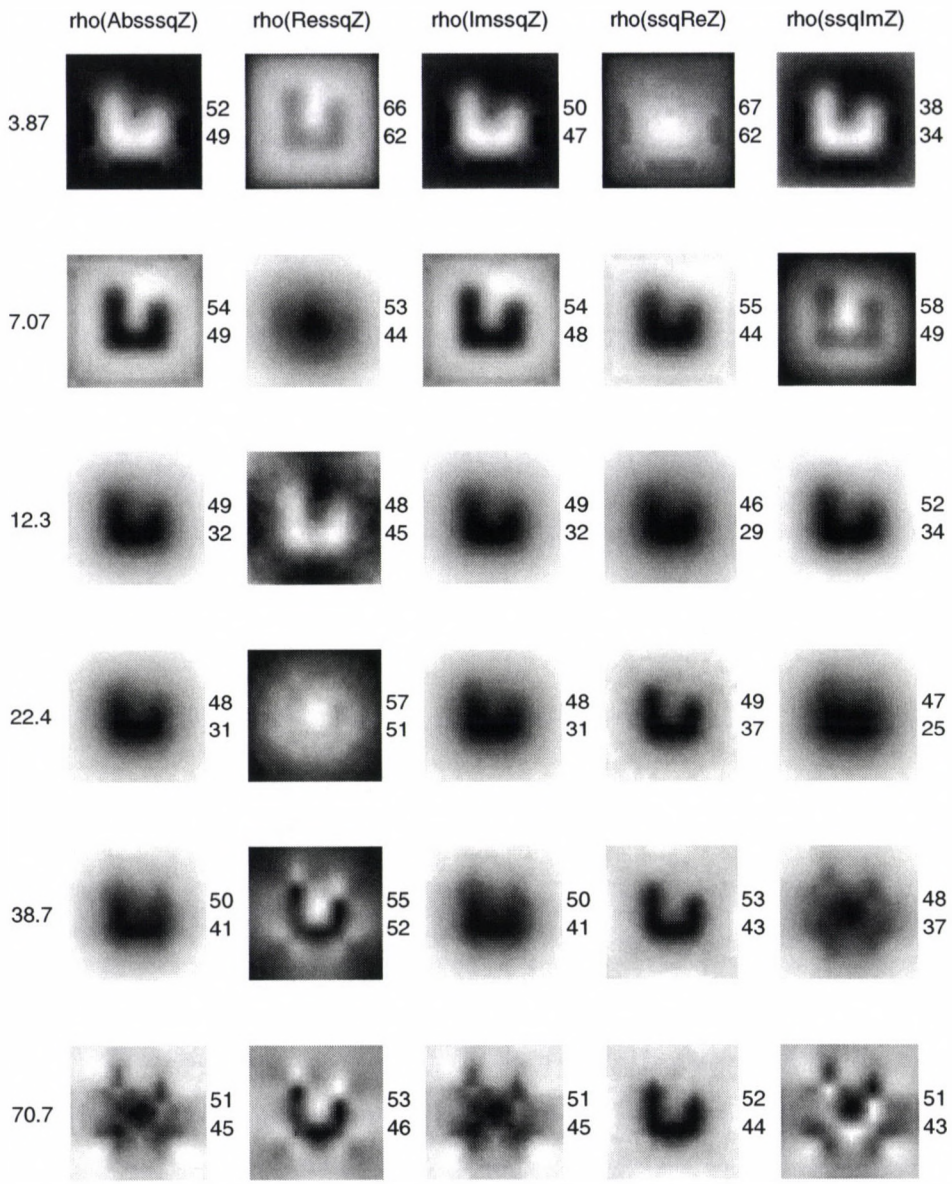


Fig. 4a. Apparent resistivity maps over the U-shaped heterogeneity presented in Fig. 1. They are computed at six different periods T = 0.03 s, 0.1 s, 0.3 s, 1.0 s, 3.0 s and 10 s, increasing from top to bottom, and the on the left side of the figure the corresponding λ_r is indicated. The five apparent resistivity definitions based upon $f(\mathbf{Z}) = \mathrm{ssq}(\mathbf{Z})$, from left to right, are as follows: $|\mathrm{ssq}(\mathbf{Z})|$, $Re\mathrm{ssq}(\mathbf{Z})$, $Im\mathrm{ssq}(\mathbf{Z})$, $\mathrm{ssq}(Re\mathbf{Z})$ and $\mathrm{ssq}(Im\mathbf{Z})$. Each map is produced using the autoshading function 'colormap(gray)' of MATLAB individually. The maximum and minimum values in the right side of each map correspond to the most black and most white regions

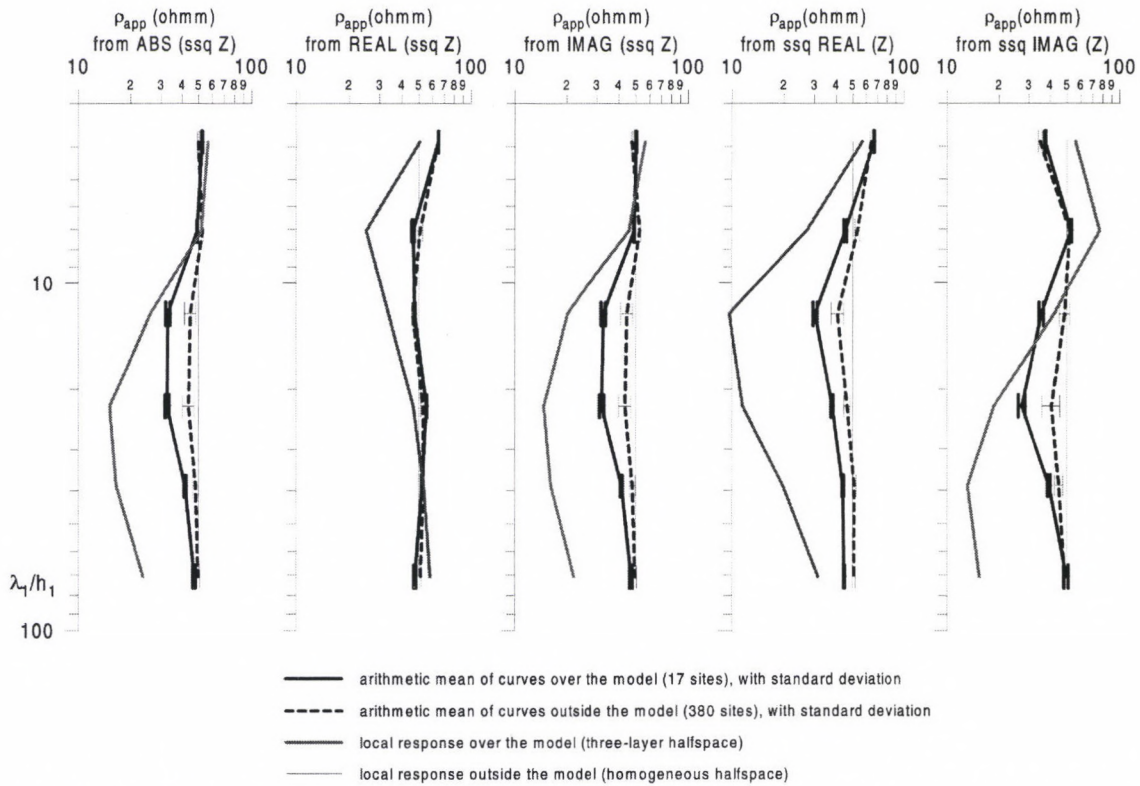


Fig. 4b. Five different apparent resistivity definitions, computed by using $f(\mathbf{Z}) = \mathrm{ssq}(\mathbf{Z})$, shown in five panels: a) $|\mathrm{ssq}(\mathbf{Z})|$, b) $Re\mathrm{ssq}(\mathbf{Z})$, c) $Im\mathrm{ssq}(\mathbf{Z})$, d) $\mathrm{ssq}(Re\mathbf{Z})$ and e) $\mathrm{ssq}(Im\mathbf{Z})$: – averaged apparent resistivity curve of 17 sites, situated inside the plan view of the heterogeneity, with the corresponding standard deviations; – averaged apparent resistivity curve of 380 sites, located outside the heterogeneity, with the corresponding standard deviations; – the local 1D response over the heterogeneity, namely that of a three-layer half space, with a second layer having the same thickness and resistivity as the U-shaped heterogeneous structure; – the local 1D response around the heterogeneity, namely that of a homogeneous half space, having a resistivity of 50 Ωm

3.2 Model 2 (resistive target)

Since the different functions f of the impedance tensor gave even closer results than in the case of conductive targets, only the results corresponding to the five different resistivity definitions based upon the det(**Z**) are presented. Figure 5a shows the apparent resistivity maps, while in Fig. 5b the sounding curves are shown.

The resistivity variations are of course much less intense in this case than in the case of model 1. In this case again, the homogeneity requirement over and around the target is fulfilled the best by the apparent resistivity definition based on det($Re\mathbf{Z}$). In particular, striking difference can be observed between the $\varrho_{f(Re\mathbf{Z})}$ images and the $\varrho_{f(Im\mathbf{Z})}$ ones, which are the worst ones.

3.3 Imaging properties in presence of a near-surface inhomogeneity

For the same reason as in the two preceding cases, again only det(**Z**) based apparent resistivity maps and sounding curves are presented. Figure 6a shows apparent resistivity maps, while in Fig. 6b sounding curves are shown.

Since the depth ratio between the near-surface inhomogeneity model and the U-shaped target is 0.1, the λ_r values for the "deep" target is 10 times greater than that corresponding to the near-surface one. Consequently, the effect of the near-surface inhomogeneity from beginning of the shortest period can be regarded nearly as a "static" effect.

The presence of this near-surface inhomogeneity has a dramatic effect on few images, and in particular on the $\varrho_{\det(Im\mathbf{Z})}$ and the $\varrho_{Re\,\det(\mathbf{Z})}$ ones (2nd and 5th columns of Figs 3a and 6a). On the contrary, the $\varrho_{\det(Re\mathbf{Z})}$ images clearly depict both the superficial heterogeneity and the target. In particular, the existence of a superficial heterogeneity does not induce any significant distortion of the image of the U-shaped target which remains clearly visible under the black spot corresponding to the near-surface heterogeneity (4th column of Figs 3a and 6a). The change in the grey level results from the used automatic determination of the grey scale. It is especially worth comparing Figs 3b and 6b, where the standard deviation values for $\varrho_{f(Re\mathbf{Z})}$ remain unchanged in presence of near-surface inhomogeneity, while those for $\varrho_{f(Im\mathbf{Z})}$ are dramatically increased. In this case again, the $\varrho_{f(Re\mathbf{Z})}$ images are the best ones, and the $\varrho_{f(Im\mathbf{Z})}$ images are the worst ones.

3.4 What is about the phase?

Figures 7a and 7b present respectively the maps and the sounding curves for the phase lags for model 1. Although the values may be different, all the phase map images are similar to the $\varrho_{Ref(\mathbf{Z})}$ apparent resistivity maps, shown in the second column of Figs 2a, 3a and 4a. This comes from the fact that the $\varrho_{Ref(\mathbf{Z})}$ maps correspond to the variations of $f(Re\mathbf{Z}) - f(Im\mathbf{Z})$ differences, while all phase maps correspond to those of the ratio between the same quantities. The oscillation range of the phase is shifted toward smaller λ_r ($\lambda_r < 4$) values, compared to those belonging to apparent resistivities. (Nice keyhole images can be seen at $\lambda_r = 3.87$.) For larger λ_r values, all phase maps indicate the presence of high conductivity

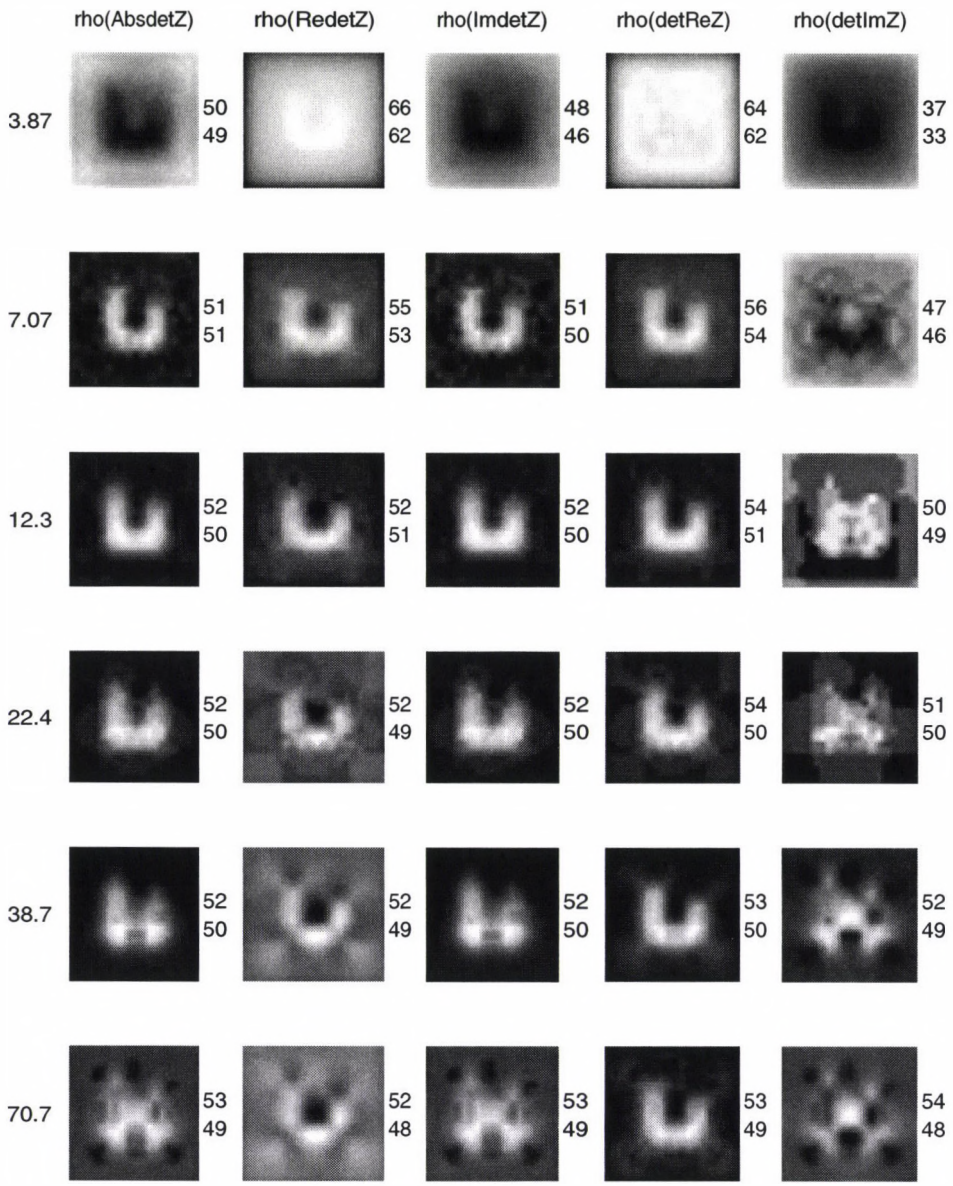


Fig. 5a. Apparent resistivity maps over a high-resistivity (2500 Ωm) U-shaped heterogeneity. All other parameters are those presented in Fig. 1. They are computed at six different periods T = 0.03 s, 0.1 s, 0.3 s, 1.0 s, 3.0 s, and 10 s, increasing from top to bottom, and the on the left side of the figure the corresponding λ_r is indicated. The five apparent resistivity definitions based upon $f(\mathbf{Z}) = \det(\mathbf{Z})$, from left to right, are as follows: $|\det(\mathbf{Z})|$, $Re\det(\mathbf{Z})$, $Im\det(\mathbf{Z})$, $\det(Re\mathbf{Z})$ and $\det(Im\mathbf{Z})$. Each map is produced using the autoshading function 'colormap(gray)' of MATLAB individually. The maximum and minimum values in the right side of each map correspond to the most black and most white regions

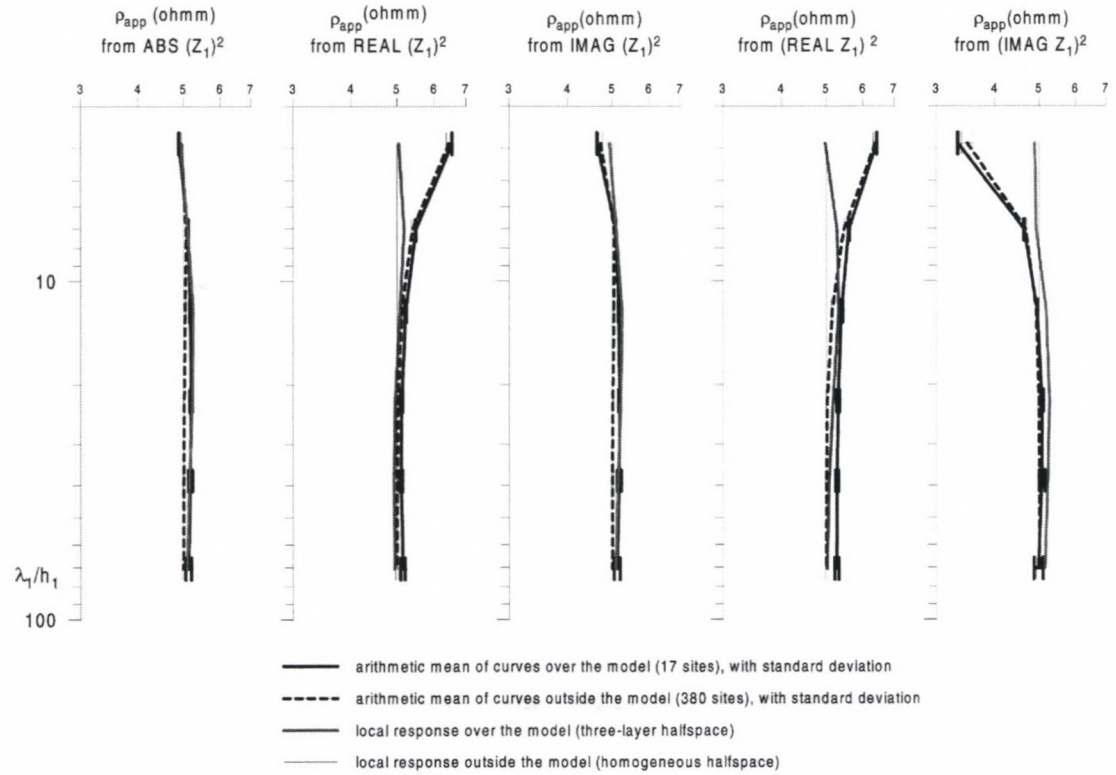


Fig. 5b. Five different apparent resistivity definitions over a high-resistivity (2500 Ωm) U-shaped heterogeneity, computed by using $f(\mathbf{Z}) = \det(\mathbf{Z})$, shown in five panels: a) $|\det(\mathbf{Z})|$, b) $Re\det(\mathbf{Z})$, c) $Im\det(\mathbf{Z})$, d) $\det(Re\mathbf{Z})$ and e) $\det(Im\mathbf{Z})$: – averaged apparent resistivity curve of 17 sites, situated inside the plan view of the heterogeneity, with the corresponding standard deviations; – averaged apparent resistivity curve of 380 sites, located outside the heterogeneity, with the corresponding standard deviations; – the local 1D response over the heterogeneity, namely that of a three-layer half space, with a second layer having the same thickness and resistivity as the U-shaped heterogeneous structure; – the local 1D response around the heterogeneity, namely that of a homogeneous half space, having a resistivity of 50 Ωm

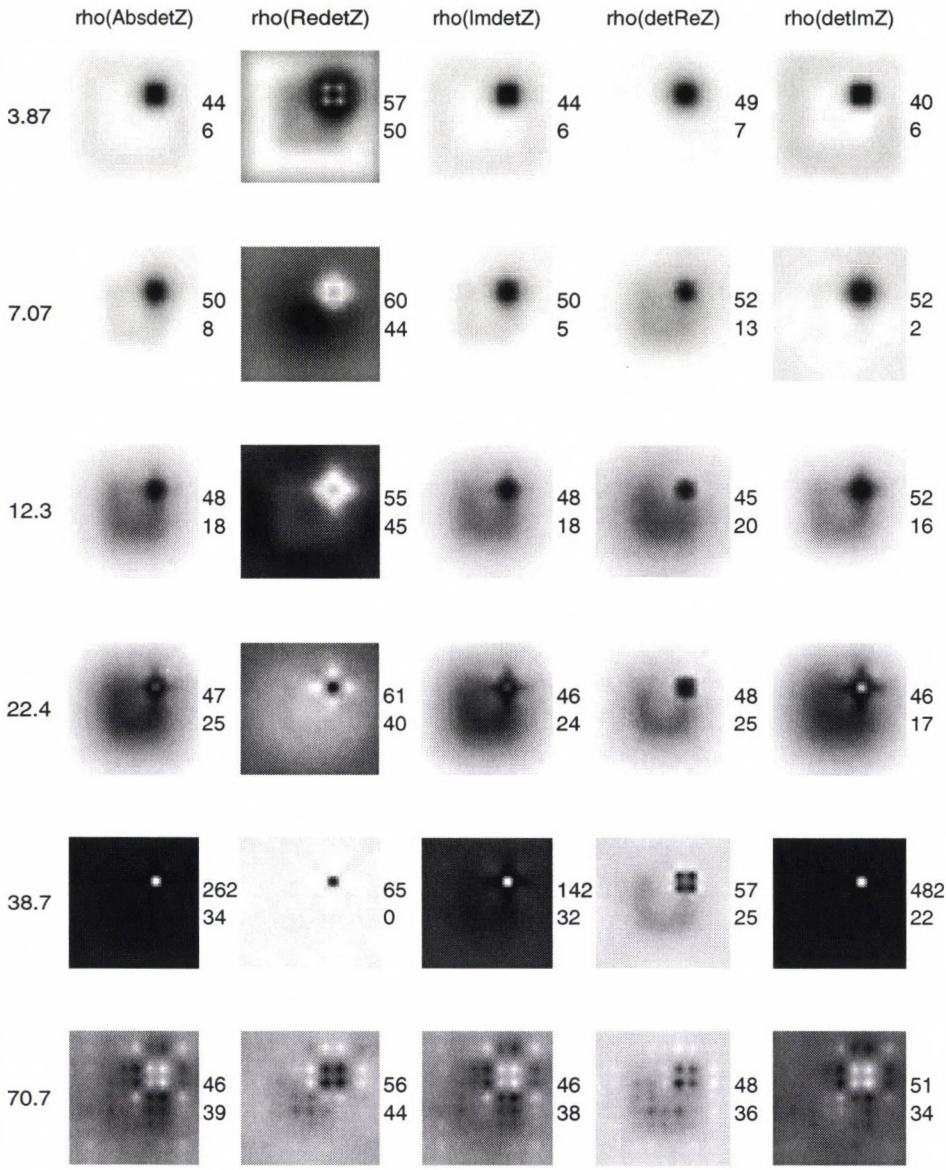


Fig. 6a. Apparent resistivity maps over the U-shaped heterogeneity, in presence of a 1km × 1km large and thin (100 m) conducting body, at a depth of 100 m. The near-surface body is situated over the missing leg of the U-shaped heterogeneity shown in Fig. 1. The maps are computed at six different periods T = 0.03 s, 0.1 s, 0.3 s, 1.0 s, 3.0 s, and 10 s, increasing from top to bottom, and the on the left side of the figure the corresponding λ_r is indicated. The five apparent resistivity definitions based upon $f(\mathbf{Z}) = \det(\mathbf{Z})$, from left to right, are as follows: $|\det(\mathbf{Z})|$, $Re\det(\mathbf{Z})$, $Im\det(\mathbf{Z})$, $\det(Re\mathbf{Z})$ and $\det(Im\mathbf{Z})$. Each map is produced using the autoshading function 'colormap(gray)' of MATLAB individually. The maximum and minimum values in the right side of each map correspond to the most black and most white regions

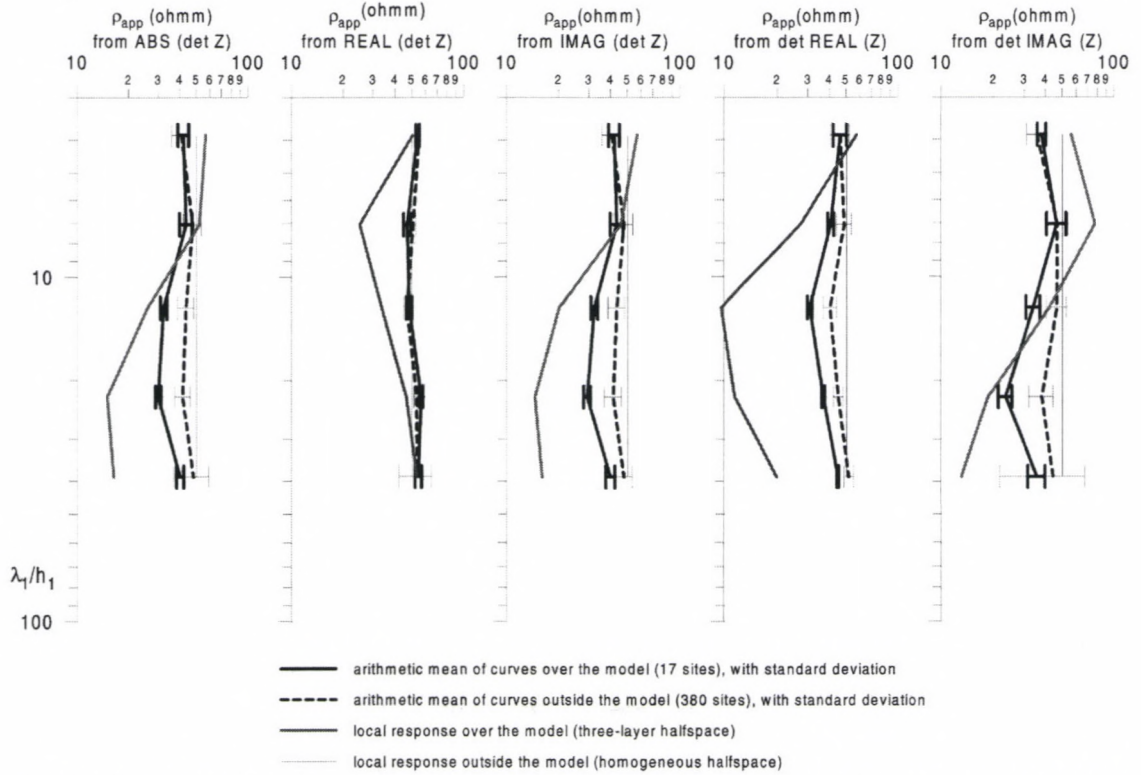


Fig. 6b. Five different apparent resistivity definitions over the U-shaped heterogeneity, in presence of the near-surface body, computed by using $f(\mathbf{Z}) = \det(\mathbf{Z})$, shown in five panels: a) $|\det(\mathbf{Z})|$, b) $Re\det(\mathbf{Z})$, c) $Im\det(\mathbf{Z})$, d) $\det(Re\mathbf{Z})$ and e) $\det(Im\mathbf{Z})$: – averaged apparent resistivity curve of 17 sites, situated inside the plan view of the heterogeneity, with the corresponding standard deviations; – averaged apparent resistivity curve of 380 sites, located outside the heterogeneity, with the corresponding standard deviations; – the local 1D response over the heterogeneity, namely that of a three-layer half space, with a second layer having the same thickness and resistivity as the U-shaped heterogeneous structure; – the local 1D response around the heterogeneity, namely that of a homogeneous half space, having a resistivity of 50 Ωm

model, although the oscillations around the U-shaped target remain relatively quite important.

3.5 Relations between images based on rotational invariants

In all above numerical experiments the apparent resistivity images computed from the real tensor proved to be the best and the images computed from the imaginary tensor proved to be the worst ones. Looking at the phase maps, it is very evident that all oscillations around the target origin from those in the apparent resistivity maps derived from $Im\mathbf{Z}$.

At this moment it is useful to overview the relations between different images based on different quadratic functions of the magnetotelluric impedance tensor.

It is easy to demonstrate that for any quadratic function f, $Ref(\mathbf{Z})$

$$Ref(\mathbf{Z}) = f(Re\mathbf{Z}) - f(Im\mathbf{Z})\,. \tag{3}$$

For the three functions f the corresponding equations are as follows.

$$ReZ_1^2 = Re^2 Z_1 - Im^2 Z_1 \tag{4a}$$

$$Re\,\det(\mathbf{Z}) = \det(Re\mathbf{Z}) - \det(Im\mathbf{Z}) \tag{4b}$$

$$Re\,\mathrm{ssq}(\mathbf{Z}) = \mathrm{ssq}(Re\mathbf{Z}) - \mathrm{ssq}(Im\mathbf{Z})\,. \tag{4c}$$

According to Eqs (4a, b and c) the apparent resistivity images based on the real part of any quadratic function of the complex $\mathbf{Z}$ reflect merely the difference between the images belonging to the same quadratic function of the real tensor and that of the imaginary one.

At the same time, $Imf(\mathbf{Z})$ is always the product of real and imaginary elements:

$$\begin{aligned} ImZ_1^2 &= 2ReZ_1 ImZ_1 = \\ &= (ReZ_{xy}ImZ_{xy} - ReZ_{xy}ImZ_{yx} - ReZ_{yx}ImZ_{xy} + ReZ_{yx}ImZ_{yx})/2 \end{aligned} \tag{5a}$$

$$Im\det(\mathbf{Z}) = ReZ_{xx}ImZ_{yy} - ReZ_{xy}ImZ_{yx} + ReZ_{yy}ImZ_{xx} - ReZ_{yx}ImZ_{xy} \tag{5b}$$

$$Im\mathrm{ssq}(\mathbf{Z}) = 2(ReZ_{xx}ImZ_{xx} + ReZ_{xy}ImZ_{xy} + ReZ_{yx}ImZ_{yx} + ReZ_{yy}ImZ_{yy}\,. \tag{5c}$$

Looking at $Imf(\mathbf{Z})$ images, we can see that they represent some intermediate quality between the real tensor based and the imaginary tensor based images.

The situation is the same for the $|f(\mathbf{Z})|$ images: it is straightforward to express $|f(\mathbf{Z})|$ in terms of $Ref(\mathbf{Z})$ and $Imf(\mathbf{Z})$, where $Ref(\mathbf{Z})$ can be substituted by the difference between $f(Re\mathbf{Z})$ and $f(Im\mathbf{Z})$.

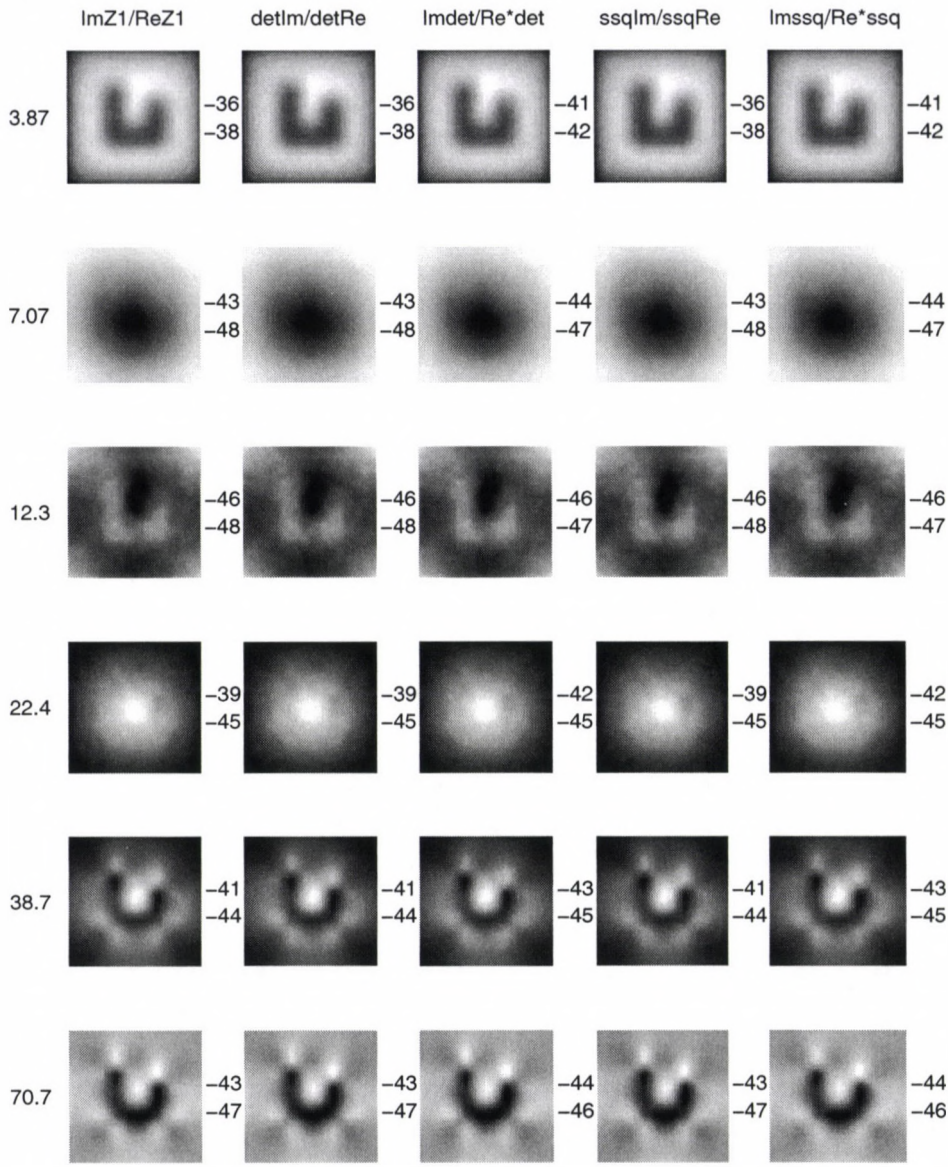


Fig. 7a. Phase maps over the model presented in Fig. 1. They are computed at nine different periods T = 0.03 s, 0.1 s, 0.3 s, 1.0 s, 3.0 s, and 10 s, increasing from top to bottom, and the on the left side of the figure the corresponding λ_r is indicated. The five different phase definitions are those shown in Table III. Each map is produced using the advanced function 'colormap(gray)' of MATLAB individually. The maximum and minimum values in the right side of each map correspond to the most black and most white regions

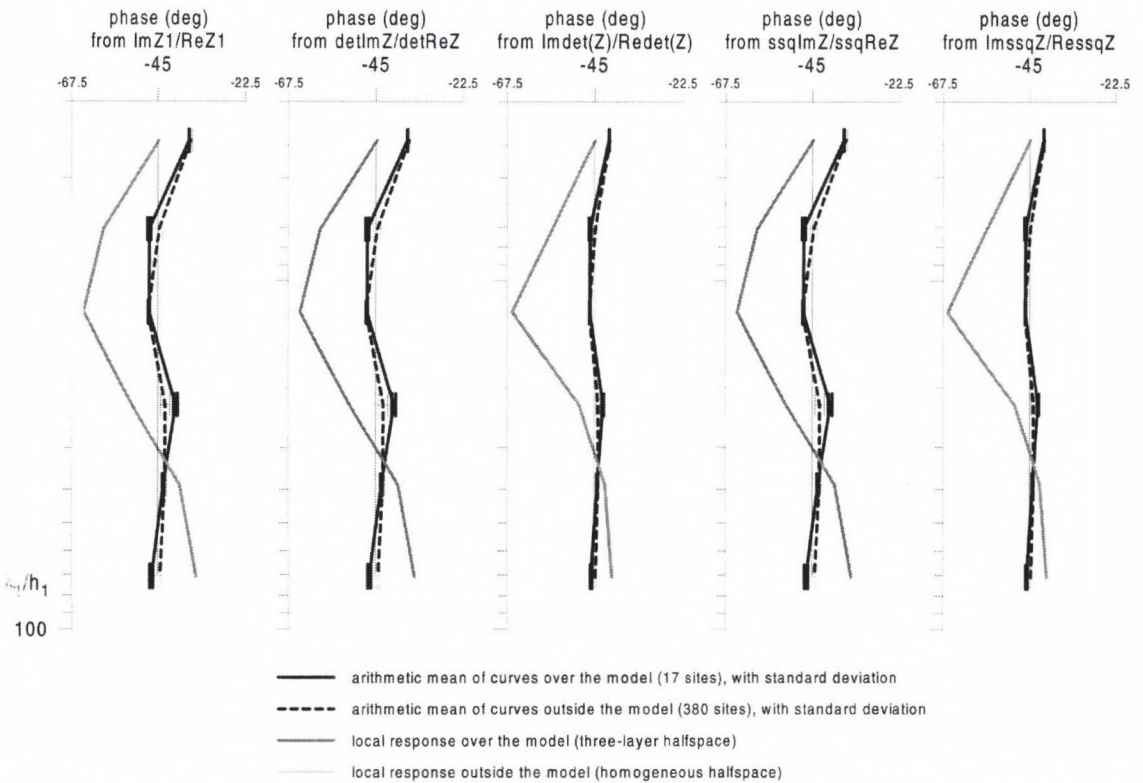


Fig. 7b. Phase sounding curves, based on definitions, shown in five panels: a) ImZ_1/ReZ_1, b) $\det Im\mathbf{Z}/\det Re\mathbf{Z}$, c) $Im\det(\mathbf{Z})/(Re\det(\mathbf{Z}) + Im\det(\mathbf{Z}))$, d) $\mathrm{ssq}Im\mathbf{Z}/\mathrm{ssq}Re\mathbf{Z}$ and e) $Im\mathrm{ssq}(\mathbf{Z})/(Re\mathrm{ssq}(\mathbf{Z}) + Im\det(\mathbf{Z}))$ over model 1: – averaged apparent resistivity curve of 17 sites, situated inside the plan view of the heterogeneity, with the corresponding standard deviations; – averaged apparent resistivity curve of 380 sites, located outside the heterogeneity, with the corresponding standard deviations; – the local 1D response over the heterogeneity, namely that of a three-layer half space, with a second layer having the same thickness and resistivity as the U-shaped heterogeneous structure; – the local 1D response around the heterogeneity, namely that of a homogeneous half space, having a resistivity of 50 Ωm

4. Conclusions

In this paper a wide period summary of 3D magnetotelluric results are given and a qualitative estimate of the imaging properties of different apparent resistivity definitions is presented. This estimate is mostly based upon a simple visual analysis of apparent resistivity maps and of sounding curves averaged over and around a conductive (and a resistive) heterogeneity embedded in a homogeneous half space. The computations are made for a set of six periods selected in order to cover a wide magnetotelluric period range.

Fifteen definitions of the apparent resistivity are considered. They are calculated using five different functions [$|f(\mathbf{Z})|$, $Ref(\mathbf{Z})$, $Imf(\mathbf{Z})$, $f(Re\mathbf{Z})$, and $f(Im\mathbf{Z})$] of three rotational invariants f of the magnetotelluric impedance tensor [$f(\mathbf{Z}) = Z_1^2 = (Z_{xy} - Z_{yx})^2/4$; $f(\mathbf{Z}) = \det(\mathbf{Z}) = Z_{xx} \cdot Z_{yy} - Z_{xy} \cdot Z_{yx}$; $f(\mathbf{Z}) = \mathrm{ssq}(\mathbf{Z}) = Z_{xx}^2 + Z_{xy}^2 + Z_{yx}^2 + Z_{yy}^2$].

The main results can be summarised as follows.

I. The observed imaging properties significantly depend on the definition of the apparent resistivity but not on the function f.

II. The period range can be divided into two parts. For the shorter (so called: "oscillating") period domain, the period dependence of images seems to be very complicated, leading to irregular variations of the image. On the contrary, for the longer periods (in the so-called "normal") range, all the real-valued rotational invariants we consider leads to robust and characteristic images of the target.

III. In the "normal" period the evolution of the images is very regular. It is remarkable that the most intense anomalies, appearing in the beginning of this range, do not image the target.

IV. In the "normal" period range, the real-tensor based apparent resistivity definitions have the best imaging properties for both conductive and resistive targets: the convergence towards the asymptotic value is rapid in both period and space domains; and the oscillations — both over the heterogeneity and around it — have the smallest amplitude. These results can be regarded as a three-dimensional extension of 1D conclusions by Spies and Eggers (1986) and by Schmucker (1987).

This latter result means that in 3D environments the robust and model-geometry connected behaviour of real-tensor based apparent resistivities makes it possible to carry out the most reliable interpretation from a small (but still large enough) number of magnetotelluric soundings. There have been already published several field results (e.g. Ingham and Brown 1998) in which the real tensor based interpretation parameters proved to be better than the imaginary tensor based ones. Although the reason for this is not yet known, this paper clearly support the use of real tensor based quantities. We hope, these experimental results based on numerical modelling will orient the attention to further theoretical and experimental investigations on these interesting imaging properties.

Acknowledgements

This publication is a product of a scientific co-operation, supported by the Hungarian OMFB and the French Ministry of Foreign Affairs, in the frames of the French-Hungarian inter-governmental co-operation in science and technology. Some support from the Hungarian Research Foundation (L.S. contract Nr. T014882) and the French Institut National des Sciences de l'Univers (M.M., contract Nr. 94/TOM/22) was received, too. This work was done in the frame of P.A.S.T. position of L Szarka and V V Spichak at Université Paris Sud. Contribution of V Spichak also was supported by the Russian Basic Research Foundation (grant 99–05–64552).

Appendix

The non-uniqueness of the magnetotelluric apparent resistivity $\varrho_{\rm app}$ comes inherently from the generalisation of the following homogeneous half-space relation:

$$\varrho_{\rm app} = (1/i\omega\mu)Z^2 = (1/\omega\mu)[2ReZImZ + i(Im^2Z - Re^2Z)]\,, \tag{A.1}$$

where $Z^2 = (ReZ + iImZ)^2 = (Re^2Z - Im^2Z) + i2ReZImZ$.

$\varrho_{\rm app}$ is therefore real if, and only if $|ReZ| = |ImZ|$, that is in the case of a homogeneous half space. In any other situation, $\varrho_{\rm app}$ becomes a complex number, and obtaining a real quantity requires either to take a real function of $\varrho_{\rm app}$ (most frequently the real or imaginary part, or its modulus) or to replace the complex impedance by a real function of it. Cagniard (1953) used the absolute value of Z. In order to keep all information, he introduced another real-valued parameter, the MT phase ('the phase of the impedance'):

$$\Phi_Z = \mathrm{atan}(ImZ/ReZ)\,. \tag{A.2}$$

A list of apparent resistivity definitions in 1D

Let us summarise now the single value based apparent resistivity definitions. There are five very evident definitions (after some correction of those, used by Spies and Eggers 1986):

1. The classical magnetotelluric definition for the apparent resistivity (Cagniard 1953) is based on $|Z^2| = |Z|^2$, since

$$|Z^2| = [(Re^2Z - Im^2Z)^2 + (2ReZImZ)^2]^{1/2} = (Im^2Z + Re^2Z)^{1/2} = |Z|^2\,. \tag{A.3}$$

   Therefore

$$\varrho_{|Z^2|} = \varrho_{|Z|^2} = (1/\omega\mu)(Im^2Z + Re^2Z)\,. \tag{A.4}$$

2. An apparent resistivity can formally be computed from the real part of Z^2, ($ReZ^2 = Re^2Z - Im^2Z$, see Eqs 3 and 4), but it should be observed that both the short-period and the long-period asymptotes of the ReZ^2 curve over a 1D halfspace are equally zero. In this way, ReZ^2 is just a measure of the non-homogeneity. Nevertheless, in the case of a heterogeneity embedded in a homogeneous half space, a realistic apparent resistivity definition based on ReZ^2 can be obtained by adding ϱ_1 to ReZ^2. As a result,

$$\varrho_{ReZ^2} = \varrho_1 + (1/\omega\mu)(Re^2Z - Im^2Z) \tag{A.5}$$

   is obtained.

3. It is possible to derive an apparent resistivity using the imaginary part of Z^2, that is the real part of the right hand side member of Eq. (A1). It comes:

$$\varrho_{|ImZ^2|} = (2/\omega\mu)(ReZImZ)\,. \tag{A.6}$$

4. Apparent resistivity definitions, providing realistic values can be defined using ReZ:

$$\varrho_{ReZ} = (2/\omega\mu)Re^2Z\,. \tag{A.7}$$

5. A similar definition is based on ImZ:

$$\varrho_{ImZ} = (2/\omega\mu)Im^2Z\,. \tag{A.8}$$

In 1D situations, the classical MT apparent resistivity definition is just the arithmetic mean of ϱ_{ReZ} and ϱ_{ImZ}:

$$\varrho_{|Z^2|} = \varrho_{|Z|^2} = (\varrho_{ReZ} + \varrho_{ImZ})/2 \tag{A.9a}$$

$\varrho_{|ImZ^2|}$ is their geometric mean:

$$\varrho^2_{|ImZ^2|} = \varrho_{ReZ}\varrho_{ImZ}\,, \tag{A.9b}$$

while $\varrho_{|ReZ^2|}$ is related with their difference:

$$\varrho_{|ReZ^2|} = \varrho_1 + (1/2)(\varrho_{ReZ} - \varrho_{ImZ})\,. \tag{A.9c}$$

The apparent resistivity formulas are directly applicable to the main elements (Z_{12} and Z_{21}) of the impedance tensor, since they describe the relationship of the corresponding electric and magnetic field components. It is not true for Z_{11} and Z_{22}. Therefore they are beyond the concept of apparent resistivity, which is given in Eq. (A1).

Apparent resistivity definitions in the 3D general case

In multidimensional medium, the MT response can no longer be described by a single complex valued number. It is a complex tensor linking the horizontal components of the electric and magnetic fields. The definitions of the apparent resistivity we just recalled can be generalised by replacing the square of the 1D complex impedance by any function having a dimension of $(V/A)^2$ of the elements of the impedance tensor.

Another alternative is to use some function of $\mathbf{Z}$. If these functions are rotational invariants, the apparent resistivity at a certain site will not depend on the directions along which the magnetotelluric tensor is expressed. Such functions may be dimensionless, or of dimension V/A, $(V/A)^2$, etc., so they should be transformed to have a dimension of $(V/A)^2$. Furthermore, care should be taken for the transforming coefficient, whether it is $1/2\omega\mu$, $1/\omega\mu$, $2/\omega\mu$ or something else.

A list of alternative phase definitions

In 3D environments the magnetotelluric phase definition (see Eq. (A2)) becomes also ambiguous. The phase of the impedance needs a complex invariant, having a dimension of (V/A). With Z_1

$$\Phi_{Z_1} = \mathrm{atan}(|ImZ_1/ReZ_1|)\,. \tag{A.10}$$

From any $f(\mathbf{Z})$ quadratic function like the determinant or the sum of squares, two alternative definitions can be formally derived:

$$\Phi_{\det(ImZ)/\det(ReZ)} = \mathrm{atan}(|\det(Im\mathbf{Z})/\det(Re\mathbf{Z})|^{1/2}) \tag{A.11a}$$

and

$$\Phi_{Im\det(Z)/Re\det(Z)} = \mathrm{atan}(|Im\det(\mathbf{Z})/Re\det(\mathbf{Z})|^{1/2}) \tag{A.11b}$$

In order to fulfil the asymptotic requirements of phase curves, Eq. (A11b) should be modified, e.g. as follows:

$$\Phi_{Im\det(Z)/Re\det(Z)} = \mathrm{atan}(|Im\det(\mathrm{Z})|/|Re\det(\mathbf{Z}) + Im\det(\mathbf{Z})|)^{1/2}\,. \tag{A.11c}$$

In an analogous way, from the sum of squares the two following definitions can be given:

$$\Phi_{\mathrm{ssq}(ImZ)/\mathrm{ssq}(ReZ)} = \mathrm{atan}(|\mathrm{ssq}(Im\mathbf{Z})/\mathrm{ssq}(Re\mathbf{Z})|^{1/2}) \tag{A.12a}$$

and

$$\Phi_{Im\mathrm{ssq}(Z)/Re\mathrm{ssq}(Z)} = \mathrm{atan}(|Im\mathrm{ssq}(\mathbf{Z})|/|Re\mathrm{ssq}(\mathbf{Z}) + Im\mathrm{ssq}(\mathbf{Z})|)^{1/2}\,. \tag{A.12b}$$

The same is true for Z_1^2

$$\Phi_{Im^2Z_1/Re^2Z_1} = \mathrm{atan}(|Im^2Z_1/Re^2Z_1|^{1/2}) \tag{A.13a}$$

and

$$\Phi_{ImZ_1^2/ReZ_1^2} = \mathrm{atan}(|ImZ_1^2|/|ReZ_1^2 + ImZ_1^2|)^{1/2} \tag{A.13b}$$

References

Bahr K 1988: *Journal of Geophysics*, 62, 119–127.

Bahr K 1991: *Physics of the Earth and Planetary Interiors*, 66, 24–38.

Cagniard L 1953: *Geophysics*, 18, 605–635.

Christensen N 1996: Electromagnetic subsurface imaging, Review paper presented at the 13th EM workshop of IAGA, Onuma, Japan

Groom R W, Bailey R C 1988: *J. Geoph. Res.*, B93, 1913–1925.

Groom R W, Bailey R C 1991: *Geophysics*, 56, 496–518.

Ingham M R 1988: *Geophysical Journal*, 92, 165–169.

Ingham M, Brown C 1998: *Geophys. J. Int.*, 135, 542–552.

Lilley F E M 1993: *Geophysics*, 58, 1498–1506.

Mackie R L, Madden T, Wannamaker P E 1993: *Geophysics*, 58, 215–226.

Park S K, Livelybrooks D W 1989: *Geophysics*, 54, 1483–1490.

Ranganayaki R P 1984: *Geophysics*, 49, 1730–1748.

Schmucker U 1987: *PAGEOPH*, 125, 341–367.

Spichak V V 1983: In: Algorithms and Programs for Solution of the Direct and Inverse Problems of Electromagnetic Induction in the Earth (in Russian), IZMIRAN, Moscow, 58–68.

Spichak V V 1995: In: Proc. of SDR Symposium on 3D Electromagnetics, Ridgefield, SEG, Tulsa, OK

Spichak V V, Popova I V 1998: *Izvestiya, Physics of the Solid Earth*, 34, 33–39.

Spies B R, Eggers D E 1986: *Geophysics*, 51, 1462–1471.

Szarka L 1997: *Geophysical Prospecting*, 45, 763–777.

Szarka L, Menvielle M 1997: *Geophys. J. Int.*, 129, 133–142.

Szarka L, Menvielle M 1999: *Geophysical Prospecting*, 47, 59–71.

Tikhonov A N 1950: *Dokl.Akad. Nauk. SSSR* (in Russian), 73, 295–297.

Weaver J T 1994: Mathematical methods for geo-electromagnetic induction. John Wiley and Sons Inc., New York

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 177–194 (2000)

BEHAVIOUR OF THE MAGNETOTELLURIC (MT) SOUNDING CURVES AT THE SOUTHERN BOUNDARY OF THE TRANSDANUBIAN CONDUCTIVITY ANOMALY (TCA) AND ITS REMOTE EFFECT (REGIONAL MT ANISOTROPY)

A Ádám[1], L Verő[2], V Wesztergom[1]

[Manuscript received March 17, 2000]

It is known from earlier MT soundings that at the southern boundary of the TCA, north of the Lake Balaton, the MT sounding curves strongly differ from the 1D ones (opposite tendencies in TE and TM modes, overshooting in TM mode, etc.). It has also been experienced that the directional characteristics of the impedance polar diagrams can be followed over a large area, i.e. they represent a remote effect or structural MT anisotropy.

Behaviour of the MT sounding curves near the boundary of the TCA has been studied in detail at the MT station Aszófő using the decomposition method. To clarify the geoelectric structure which can generate the observed MT curves a series of 2D halfspace models, including one with the edge of a conductor at the depth of the supposed source of the TCA, was computed.

Keywords: conductivity anomaly; decomposition; inversion; magnetotellurics; regional MT anisotropy

1. Introduction

Source of the observed regional MT anisotropy has been investigated for a long time (Ádám 1969, 1996) and it was found that it should be looked for within the basement. The southern boundary of the TCA has been chosen to study in detail one of the assumed causes — deep conductor — of the regional MT anisotropy. Its effect on the MT sounding curves is shown and analysed by various methods (decomposition, inversion and forward modelling).

2. Indication at the boundary of the TCA

It is known from the telluric experience (Ádám and Verő 1967) that the major axes of the telluric ellipses north of the Lake Balaton are perpendicular to the characteristic structural lines in the Pannonian Basin, i.e. they point towards N-NW–S-SE, as shown in Fig. 1. This peculiarity can be followed over a distance of about 30 km northward of the Lake Balaton. Consistently with the azimuth of the MT anisotropy — described in the above mentioned papers — the same direction

[1]Geodetic and Geophysical Research Institute of the Hung. Acad. Sci., H–9401 Sopron, POB 5, Hungary, e-mail: adam@ggki.hu

[2]Eötvös Geophysical Institute, H–1145 Budapest, Kolumbusz u. 17–23, Hungary

1217-8977/2000/$ 5.00

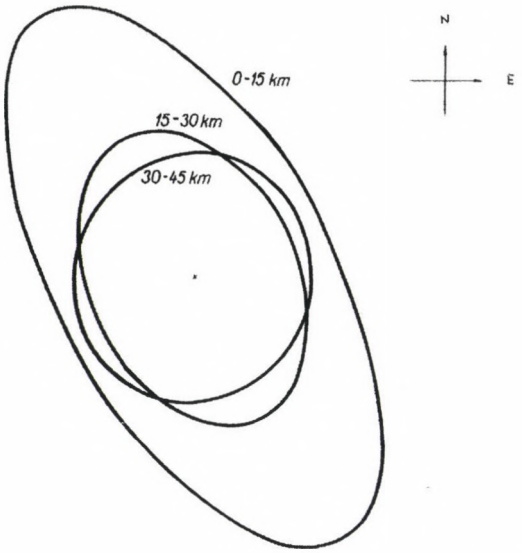


Fig. 1. Relative telluric ellipses — related to the Nagycenk observatory — north of the Lake Balaton with distances from it (Ádám and Verő 1967)

can be obtained from the major axes of MT polar diagrams in this area. On the northern shore of the Lake Balaton the observed anisotropy is much stronger, this can be attributed to the closeness of the TCA's source (Fig. 2).

According to Ádám (1971) the TCA's source consists of graphitic black shale at a depth of 4–6 km. Resistivity of this conductor can be lower than 1 Ωm, this supports the hypothesis of its graphitic composition. The impact of the TCA's source boundary can clearly be seen on our earlier MT sounding curves, e.g. at Zánka, Kékkút, Balatonöszöd, Nagycsepely, etc., at both sides of the Lake Balaton (Ádám 1976) (Fig. 3), and on ELGI's sounding curves MK1 34, KA3s 22–32, too (G Varga, personal communication). Opposite tendencies of the extreme sounding curves represent this impact. The same phenomenon can be observed on the very old, scalar MT curves at the Tihany Magnetic Observatory which were measured by Chen Leshou (1961) and Ádám (1968). Already Ádám (1968) drew the attention to the exceptional current channelling observed here and to its consequence, the strong MT anisotropy. Repetition of the MT measurement at Tihany to get tensorial relation would hardly be possible, due to the increased noise level, although it would have been needed for a deep mantle study, too (Semenov et al. 1997). A new, more reliable long period MT sounding was carried out in 1997 at a noise-free site near Tihany, at Aszófő, using the digital MT instrument DR-02 manufactured by the Geophysical Institute of the Polish Academy of Sciences. The objective was to study the impact of the TCA's source boundary on the MT sounding curves in detail in the period range between 10 s and 20000 s. Location of the Aszófő MT site is also shown in Fig. 2, on Majoros's tectonic map (1983). Direction of the main tectonic lines in the area under study can also be seen in this map (see e.g. the Balatonfő line), this direction is roughly parallel with the boundary of the TCA.

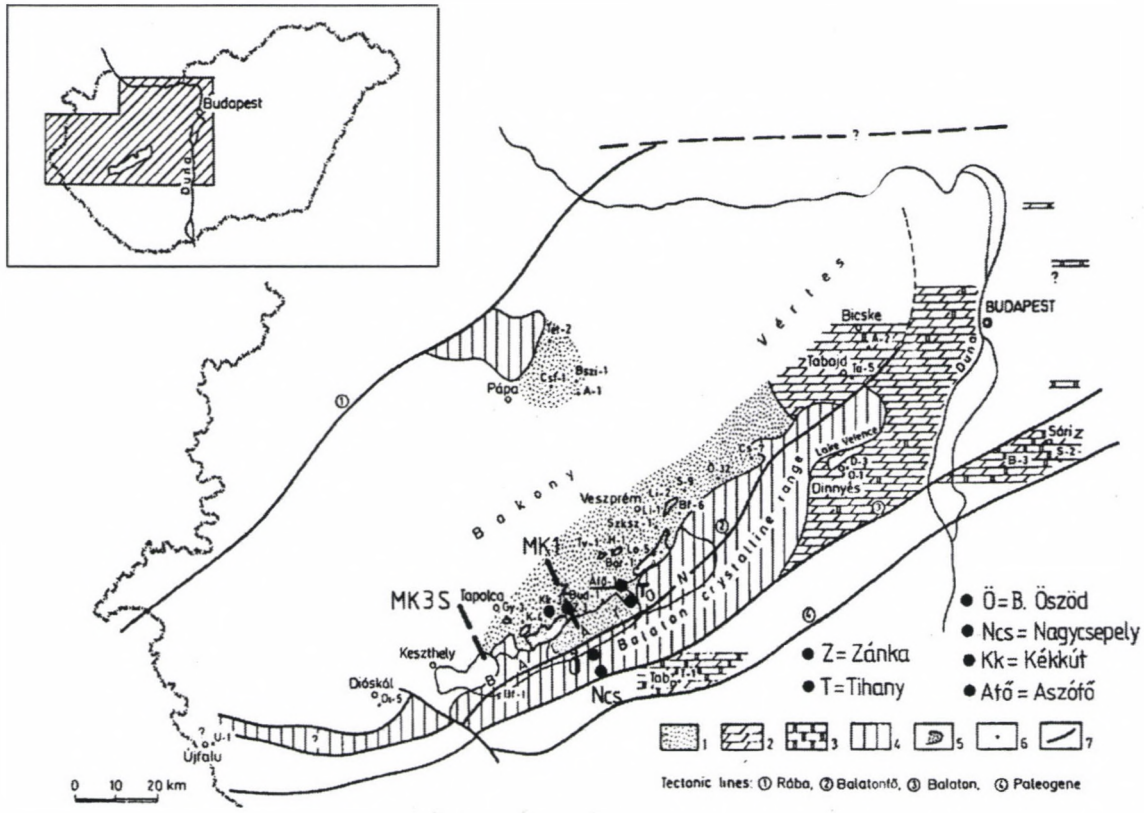


Fig. 2. Extension of the Upper Permian formations in the Transdanubian Central Mountains. 1. Red Sandstone of the Balaton Highland; 2. Dinnyés Dolomite – Tabajd Formation; 3. Nagyvisnyó Limestone; 4. Early Palaeozoic; 5. outcrop of the Permian; 6. remarkable boreholes exploring Permian formations; 7. Primary tectonic lines bordering the structural units (modified after Majoros 1983)

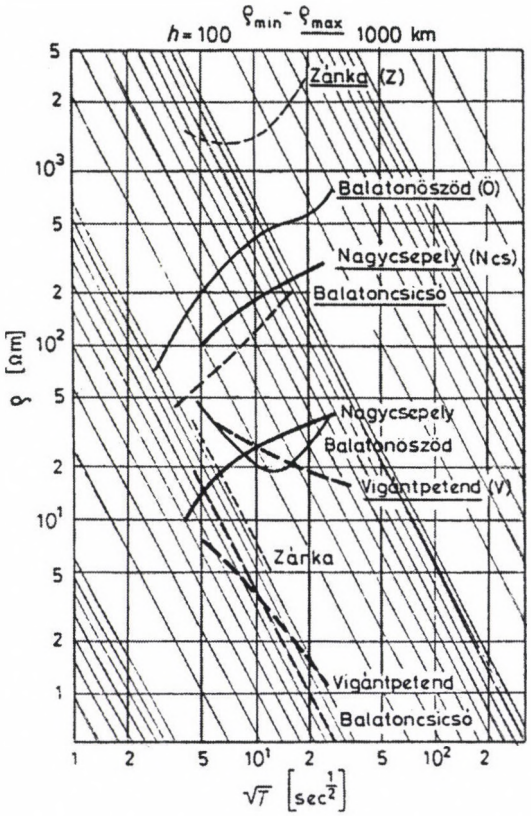


Fig. 3. MT sounding curves measured south of the boundary of the TCA (Ádám 1976)

The sounding curves measured at Aszófő are analysed using various mathematical methods to understand their information content concerning the geoelectric structure.

3. Analysis of the MT sounding at Aszófő

3.1 1D inversion and directional characteristic of the MT sounding curves

MT data were processed using J Verő's program (1972). Direction of the Z_{xy} polar diagrams' maxima is N-NW–S-SE in the whole frequency range, i.e. perpendicular to the main tectonic lines in the Pannonian Basin (and thus to the Balatonfő line as well.) This is consistent with the regional MT anisotropy (Ádám 1996). Extreme values of the impedance were determined by co-ordinate transformation and the Rhomax and Rhomin versus frequency curves were calculated (Figs 5 and 6). The difference between the two resistivity curves exceeds three orders of magnitude (the lowest Rhomin is 0.4 Ωm, while the highest Rhomax equals to 3000 Ωm). The Rhomax values can be determined with high accuracy even at $T = 10000$ s, but scattering (standard deviation) of Rhomin values is quite high.

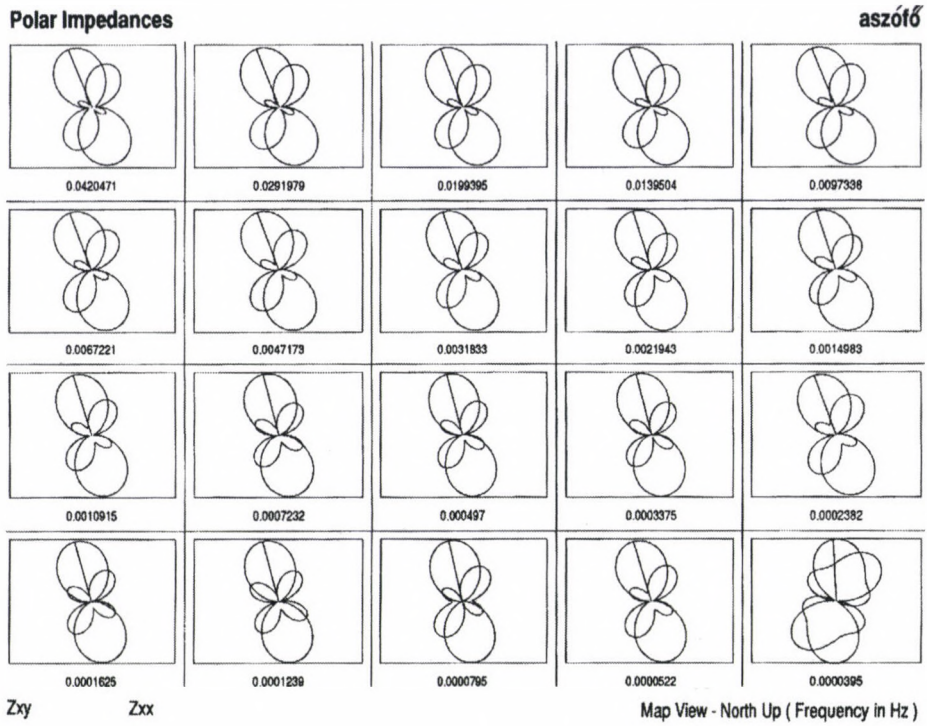


Fig. 4. Z_{xy} and Z_{xx} polar diagram of Aszófő MT sounding as a function of frequency

The two resistivity curves, even the Rhomin curve in spite of the high scattering, can be interpreted separately using 1D models. However, both the resistivity and depth values obtained strongly differ from each other.

1D inversion of the Rhomax curve provides a physically absolutely unrealistic sequence of layers (Fig. 5). Significant decrease in resistivity takes place at the depth of 2000 km, although it should appear at about 60 km, this is the average depth to the asthenosphere in the Pannonian Basin, or at 400 km, where in general the phase transition of mantle material takes place. It is worth noting that the 1D curve obtained by inversion fits well to the measured data, this might generate the illusion of a real 1D structure. This phenomenon will be explained by forward modelling.

Fitting a 1D model to the Rhomin curve is much more difficult (Fig. 6). Its shape suggests a four layer structure. Under the high resistivity uppermost layer ($\varrho > 100$ Ωm) a conductor (1 Ωm) appears at the depth of about 2.5 km. It is questionable whether this is either only a side effect or the real depth of the TCA's source. Indication of the asthenosphere can be seen at a depth of about 40 km. This is less than the average depth to the asthenosphere in the Pannonian Basin, probably due to the static shift caused by the assumed conductor.

It is generally accepted that in the case of low-degree anisotropy (e.g. caused by a 3D structure) the geometric mean of Rhomax and Rhomin values provides more

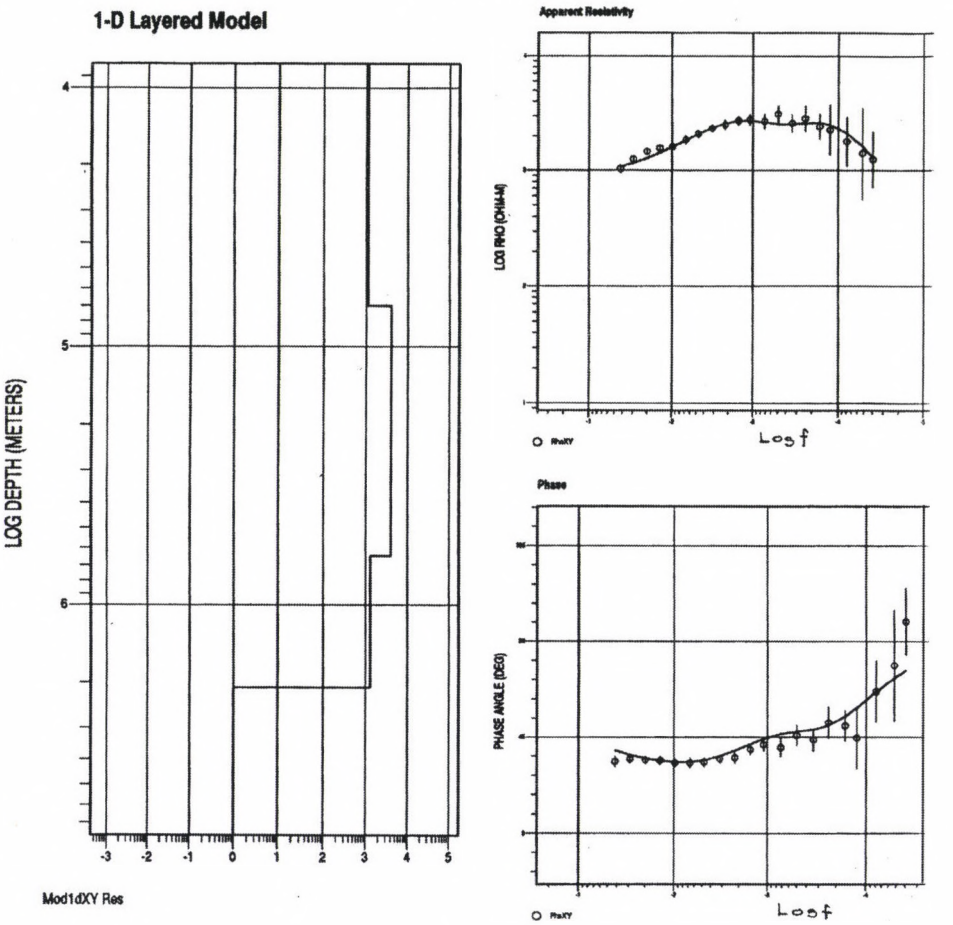


Fig. 5. 1D inversion of Rhomax curve of Aszófő

realistic results (Berdichevsky and Dmitriev 1976). Nevertheless, in our case the anisotropy is too strong, therefore 1D inversion of the geometric mean also results in an unrealistic layered structure (top of the crustal conductor was found at about 10 km, while that of the upper mantle at about 300 km).

As we have already mentioned the character of the anisotropy is quite similar to that of the regional anisotropy in the Pannonian Basin, but due to the closeness of the inhomogeneity it is much stronger here.

3.2 Effect of decomposition

An experiment has been made to clarify the 2D structure based on the decomposition of the measured MT curves because this method decreases the possible 3D effects. The basic idea of the Groom-Bailey decomposition is similar to that of others (e.g. Bahr 1988), the separation of the deformation tensor related to the 3D

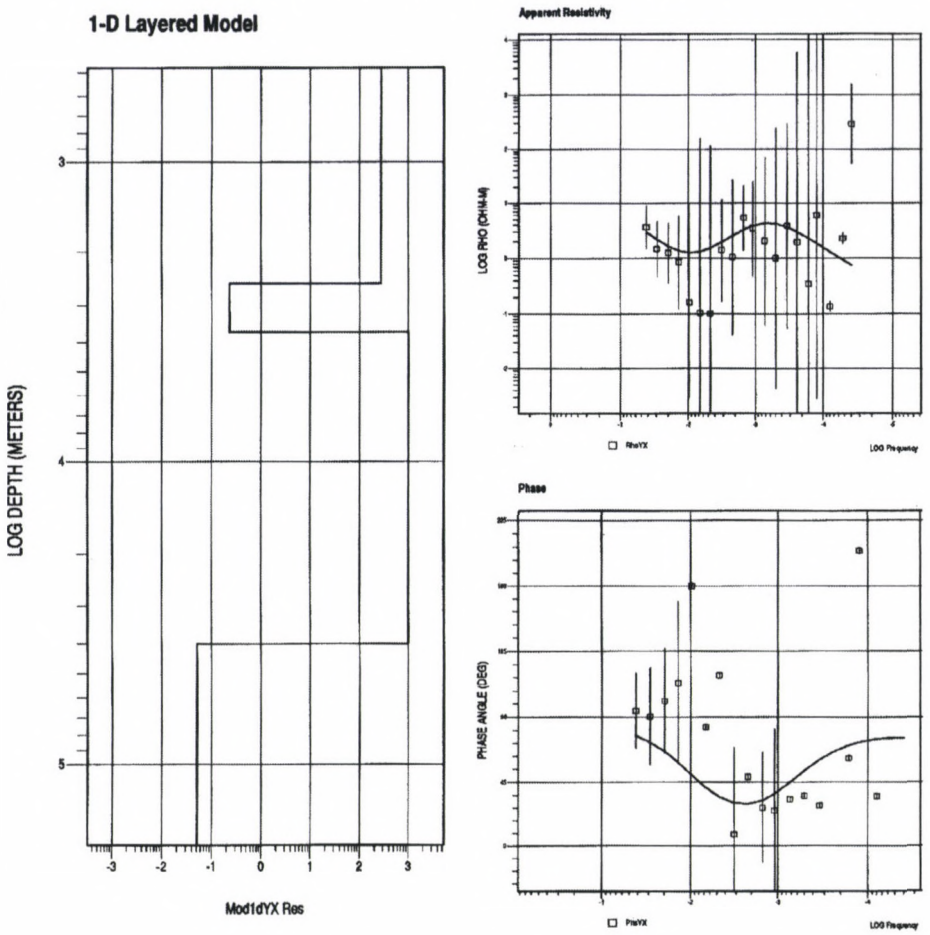


Fig. 6. 1D inversion of Rhomin curve of Aszófő

near-surface inhomogeneity which distorts the 2D impedance. Groom and Bailey proposed the factorization of **C** electric deformation tensor:

$$\mathbf{C} = g\mathbf{TSA}\,,$$

where g is the station's gain factor (related to the static shift), **A** is the anisotropy tensor, **S** is the so called shear tensor and **T** is the twist tensor.

The Groom-Bailey decomposition provides deformation parameters, the regional strike direction and the modified impedance values as well (the strike direction and deformation parameters can be fixed on request). Summarising the main features of decomposition carried out on the Aszófő impedance tensor:

- the strike direction (clockwise (CW) from north) is very close to the Z_{xymax} direction ($\sim -40°$) at short periods, but it turns towards N-S between 300 and 5000 s. The Groom-Bailey strike strongly deviates from the longitudinal tectonic direction in the Pannonian Basin (e.g. the Balatonfő line) (Fig. 7a);

– both the twist and shear values (Figs 7b and c) are extremely large and change significantly with the period. Twist shows a sinusoidal variation between 20° and 40°, while the shear decreases from 0° (at $T = 25$ s) to −40° (at $T = 10000$ s). For comparison: The mean twist and shear values in the Great Hungarian Plane are $\sim 11°$ and $\sim 7°$, respectively (Ádám 1998). These parameters obviously reflect the anomaly which appears in the difference between the Rhomax and Rhomin values.

It is interesting to note that the decomposition results in significantly smaller differences between the sounding curves of TE and TM mode (the difference is reduced by two order of magnitudes) (Fig. 8a). The deviation of the Rhoxy GB (GB = Groom-Bailey) values from the fitted 1D curve is smaller than in the case of the lower Rhoyx GB resistivity values (Figs 9 and 10). The effect of a conductor appears on the Rhoxy GB curve at short periods, similarly to the original Rhomin

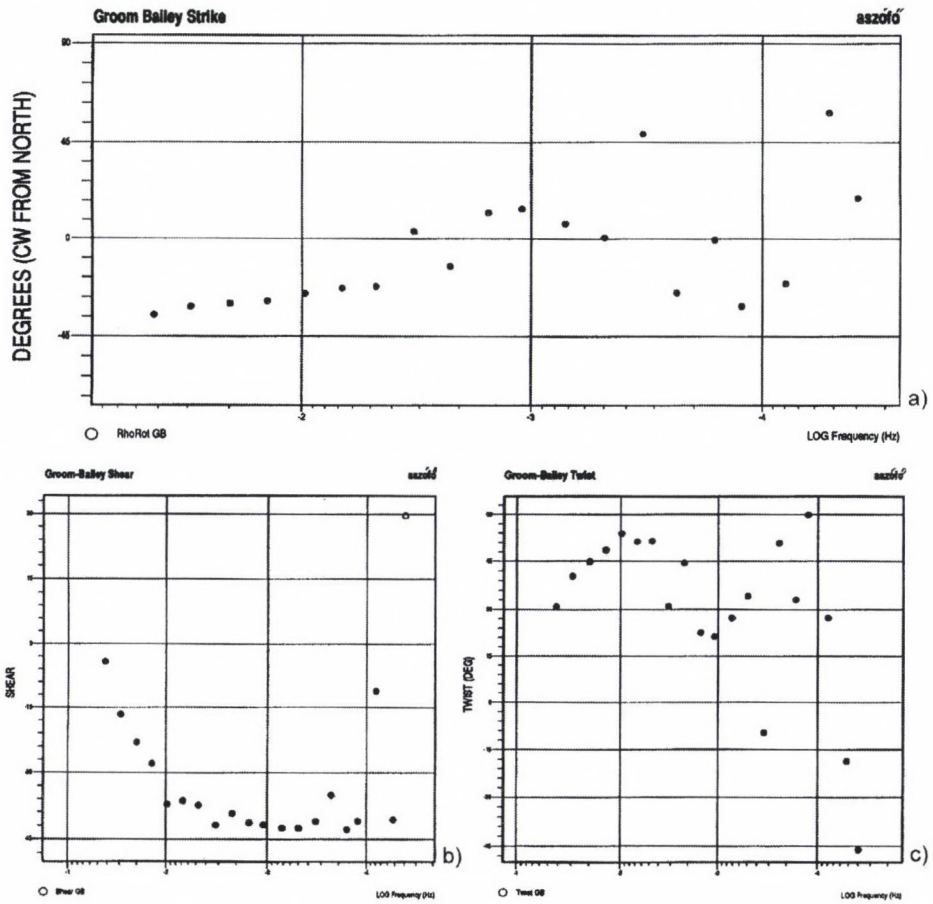


Fig. 7. Groom-Bailey (GB) decomposition: a) strike values versus frequency, b) shear versus frequency, c) twist versus frequency

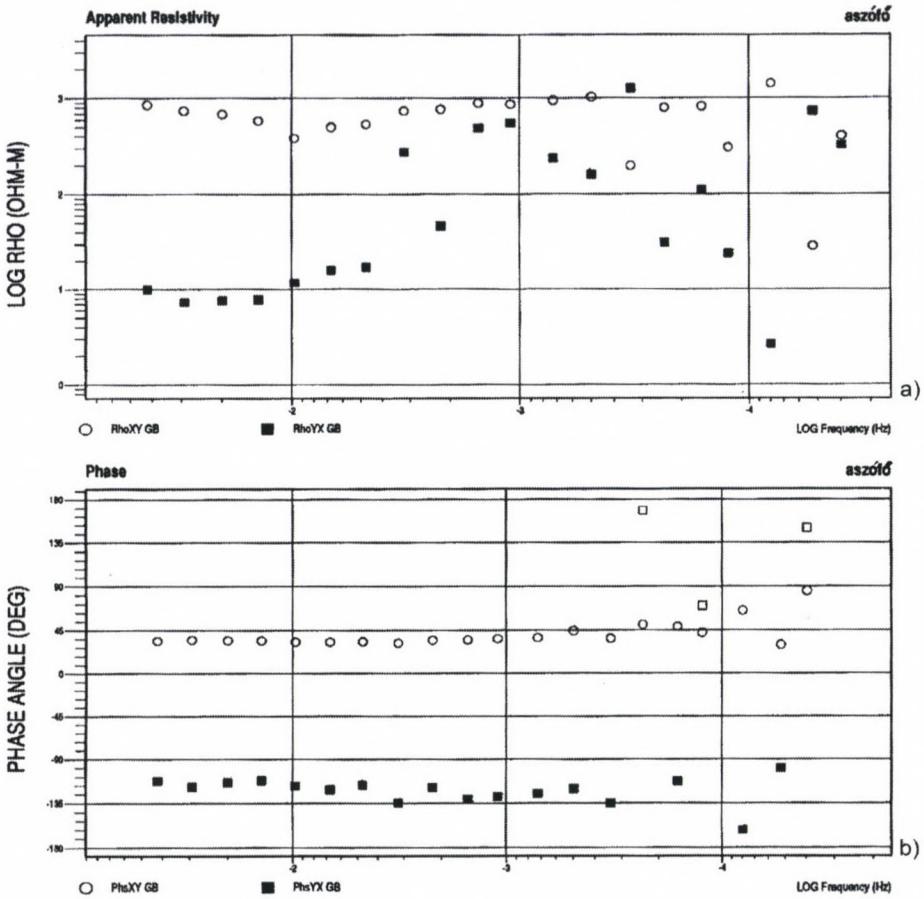


Fig. 8. Rho (a) and phase (b) values calculated after GB decomposition versus frequency

(yx) curve. Unfortunately, beyond $T = 10^3$ s the RhoXY GB values do not form a curve, therefore fitting of a theoretical curve to the data is of no use. None of the applied 1D inversions (Fischer, Occam, Marquardt) led to appropriate simultaneous fitting of resistivity and phase curves in the case of RhoYX GB values. Although the best fitting 1D curves indicate a conductor in the upper mantle between 68 and 97 km (Figs 9 and 10), the inversion is very uncertain due to the large deviations from the theoretical curves. The real geoelectric structure is far from the 1D model obtained by inversion and also far from the ideal case assumed in decomposition.

It can be concluded that large inhomogeneities, like the edge of the source of the Transdanubian conductivity anomaly, have a very strong impact on the MT sounding curves and they can be interpreted neither by 1D inversion techniques, nor by the decomposition method.

As we have already mentioned, the effect of such structures extends far away and causes regional MT structural anisotropy, i.e. difference of the extreme sounding

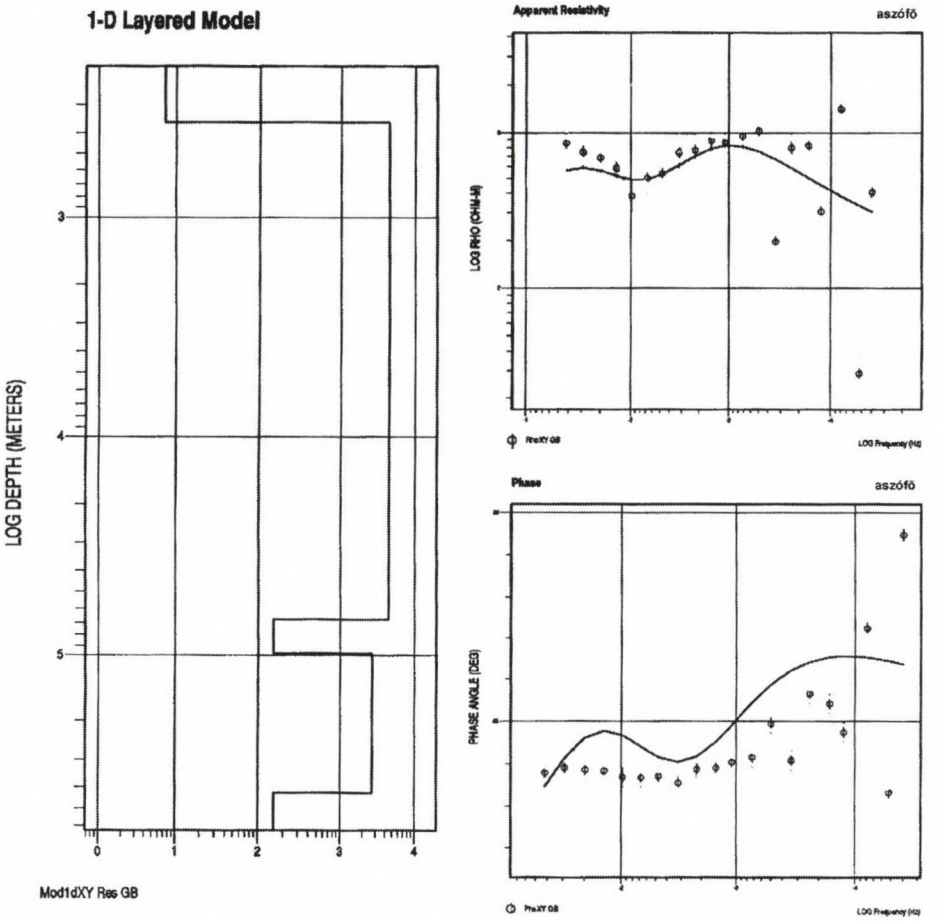


Fig. 9. 1D inversion of GB Rhoxy and phase

curves. This can be observed in MT curves and telluric ellipses measured in the area of the TCA, nevertheless it becomes weaker with increasing distance from the edge of the TCA's source (Ádám and Verő 1967). This weakening can clearly be observed south of the edge, while north of it the anisotropy is influenced by the inner structure of the supposed conductor, too.

Finally, let us see some modelling results to demonstrate the reason for the difference between the Rhomax and Rhomin curves.

4. Forward modelling

Using the 2D modelling programs of GEOTOOLS forward modelling was performed to explain the MT anisotropy observed at Aszófő. The models include a surface conductivity contrast, a so called quarter-space structure, without and with a conductive asthenosphere, and a model with a vertical conductive layer boundary (model of the TCA's source).

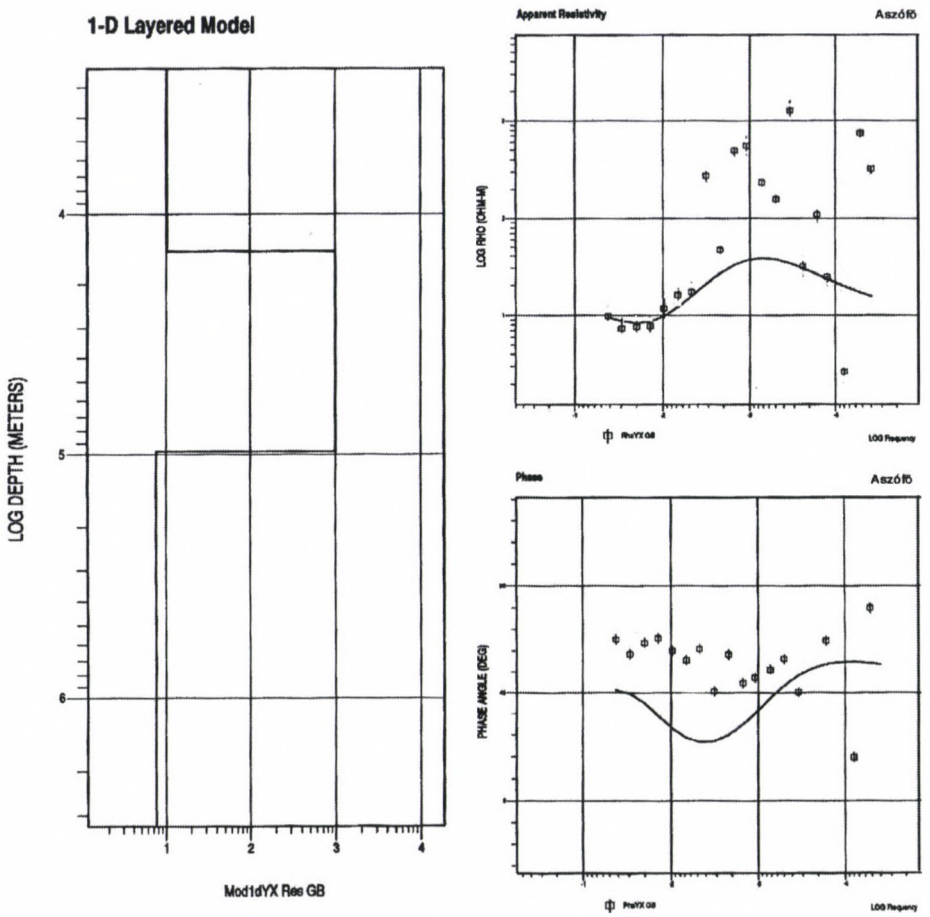


Fig. 10. 1D inversion of GB Rhoyx and phase

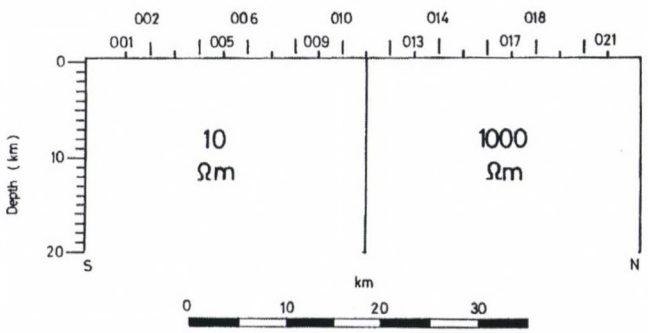


Fig. 11. A half space divided by a vertical plane into two sections (10 and 1000 Ωm) so called quarter-space structure

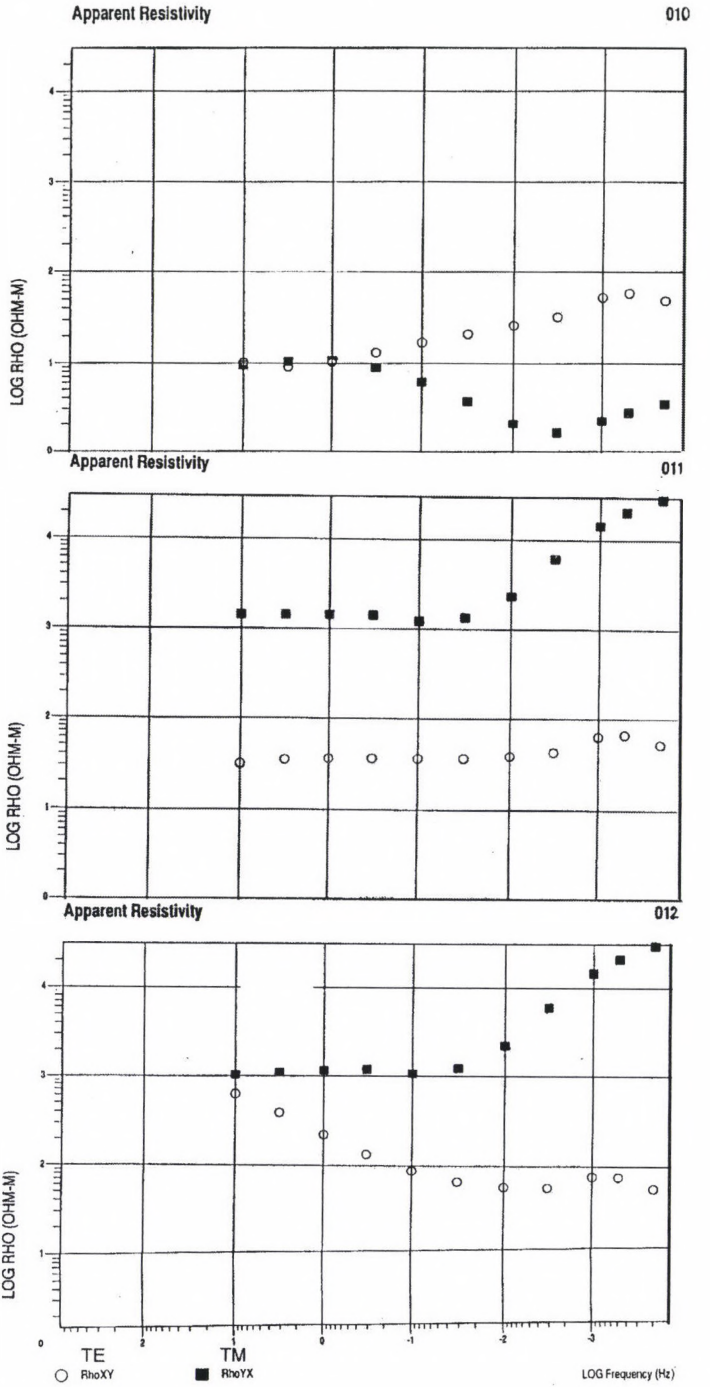


Fig. 12. Rho curves of TE and TM mode near the boundary of different resistivities in the "quarter-space" structure of Fig. 11

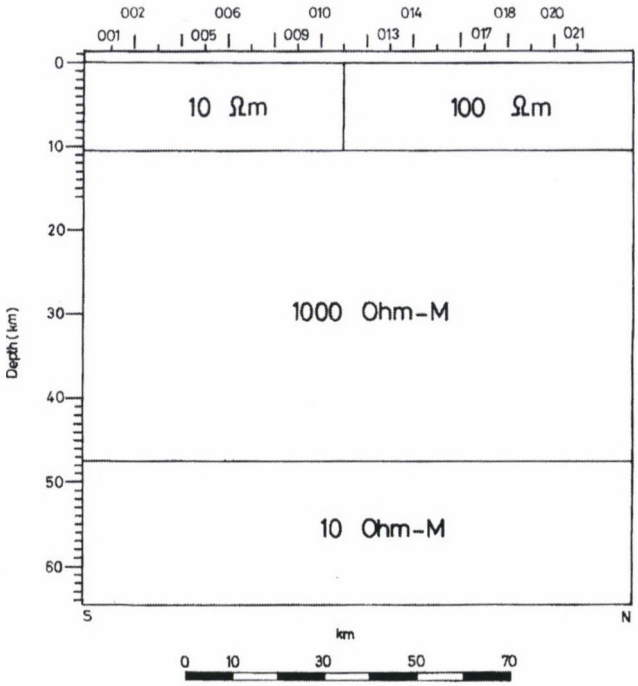


Fig. 13. A quarter-space model with conducting basement (asthenosphere)

— Model I: Halfspace, divided by a vertical plane into two parts (Fig. 11), having resistivities of 10 and 1000 Ωm, respectively. The resistivity sounding curves in the vicinity of the vertical plane are shown (Fig. 12), i.e. at points denoted by 010, 011 (vertical plane) and 012 in Fig. 11 (2.5 km apart from each other). As a result of the accumulated charges at the boundary, low frequency section of TM curves (calculated perpendicularly to the vertical boundary) lies higher over the high resistivity part than the real resistivity (10^3 Ωm), while the same section over the low resistivity part falls below the resistivity of the medium, down to $\sim$ 1 Ωm. The TE mode resistivities suffer from the side effect. It means that correct resistivity values are obtained at $T = 10$ s at both sides of the boundary, but they start to increase with increasing periods to about ten times higher values at the low resistivity side, while the opposite tendency can be observed at the high resistivity side. At $T = 10^3$ s the TE mode resistivity values are almost equal at points 010, 011 and 012, independently of their position relative to the boundary. This common value of 100 Ωm is an intermediate one between the resistivities of the two parts. Except for a part of the short period resistivity values 1D inversion of both TE and TM mode curves leads to a non-realistic layer sequence.

— Model II: The resistivity ratio in the quarter-space structure — located in the upper 10 km — is 10/100. Beneath this upper structure there is a high resistivity (1000 Ωm) layer which is underlain by a 10 Ωm layer (asthenosphere) at

a depth of 47.5 km (Fig. 13). The TM mode curves reflect the effect of charge accumulation at the boundary, similarly to those of the previous model. The sounding curve branches corresponding to the asthenosphere are shifted either upwards or downwards. The descending parts of the TE mode curves at points 009 and 012, however, provide the depth to the asthenosphere with small error (Figs 14a and b). The phase curves also differ from the 1D ones, but in the case of TE mode the deviation is much smaller than in TM mode resistivity curves because the phases are in derivative relation with the resistivity values. The general effect of a lateral resistivity contrast within a quarter-space structure was discussed by Fischer et al. (1992); their figure (Fig. 15) shows the discontinuity of the TM mode and smooth transition of TE mode resistivity curves across the boundary, together with the phase values.

— Model III: The boundary of the TCA's source was modelled. The conductor's

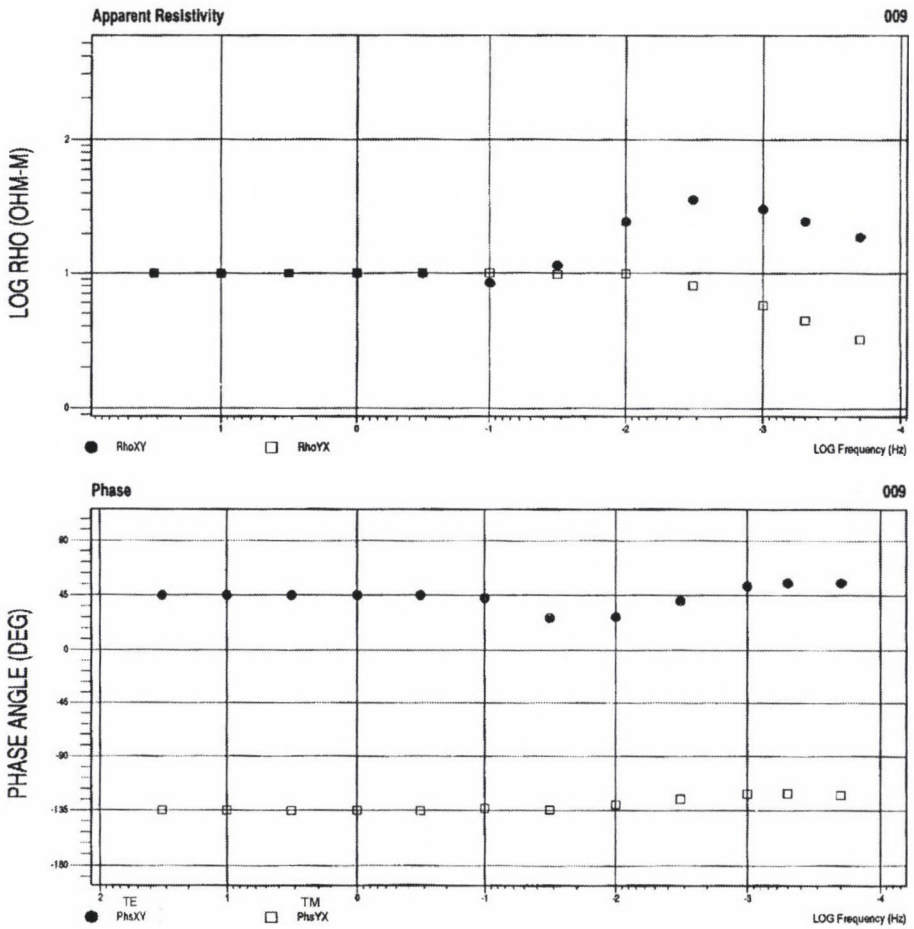


Fig. 14a. Rho and phase curves near the boundary of different resistivities in the quarter-space structure of Fig. 13 at the site 09

resistivity is 0.3 Ωm in a 500 Ωm halfspace. It lies at a depth of 2.5 km, this depth was obtained from the Rhomin curve at Aszófő (Fig. 16). The calculated TE mode and TM mode curves show opposite tendencies (Fig. 17), similarly to the measured ones. The right hand side branch of TM mode resistivity curves rises high above 1000 Ωm.

These modellings also confirmed that in the case of a 2D structure inversion of the resistivity curves measured in strike direction TE mode approximates better the parameters of a conductive layer if 1D inversion is applied.

5. Conclusions

As clearly shown by the sounding curves from Aszófő — and other ones measured earlier near the southern edge of the TCA's source (e.g. Tihany, Zánka, Balatonöszöd, etc.) — this inhomogeneity has a strong impact. This is felt as "regional

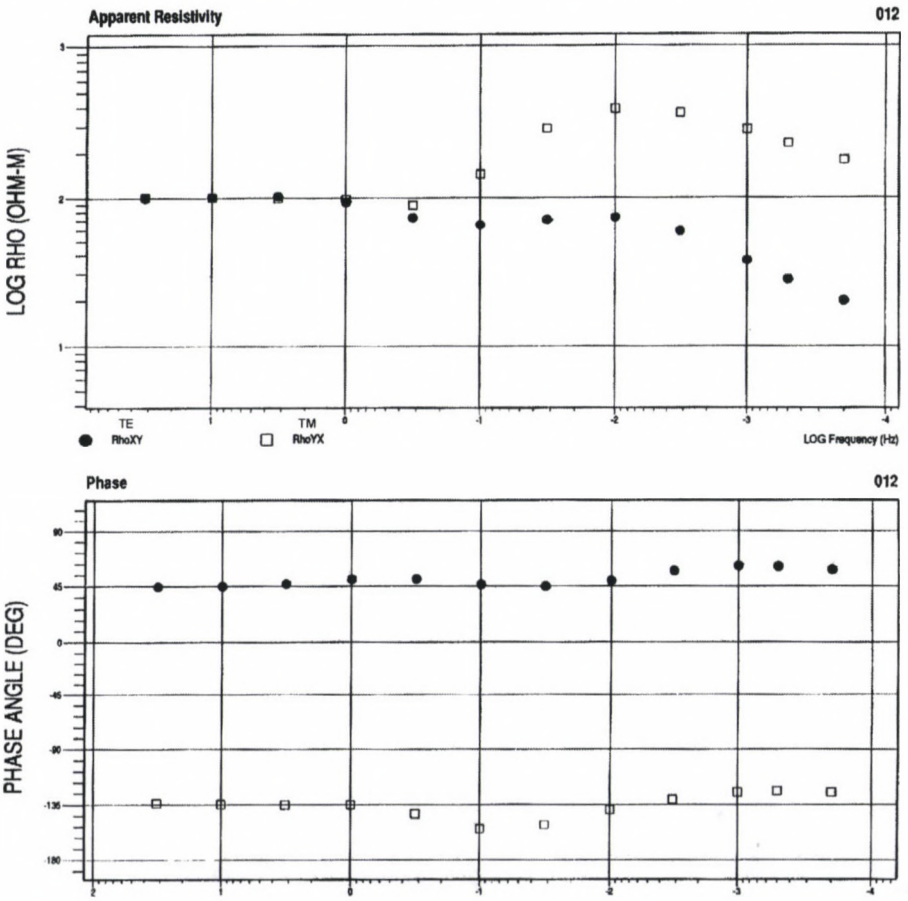


Fig. 14b. Rho and phase curves near the boundary of different resistivities in the quarter-space structure of Fig. 13 at the site 14

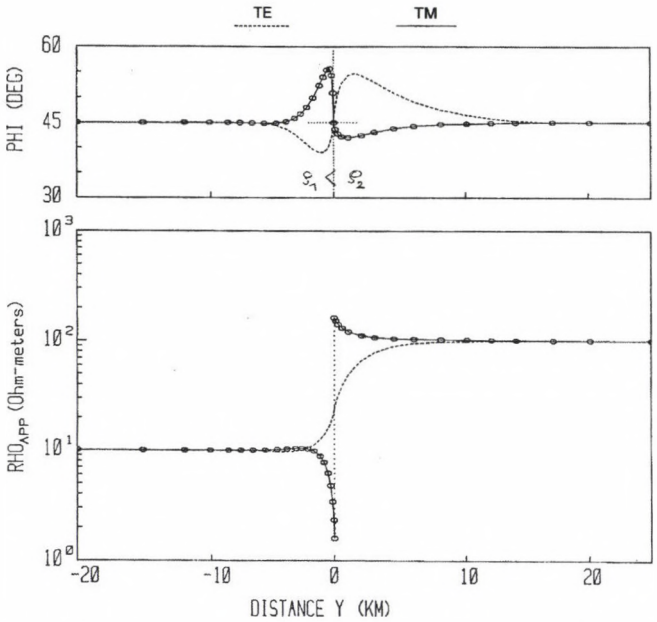


Fig. 15. Behaviour of Rho and phase values of TE and TM mode above a quarter-space structure (Fischer et al. 1992)

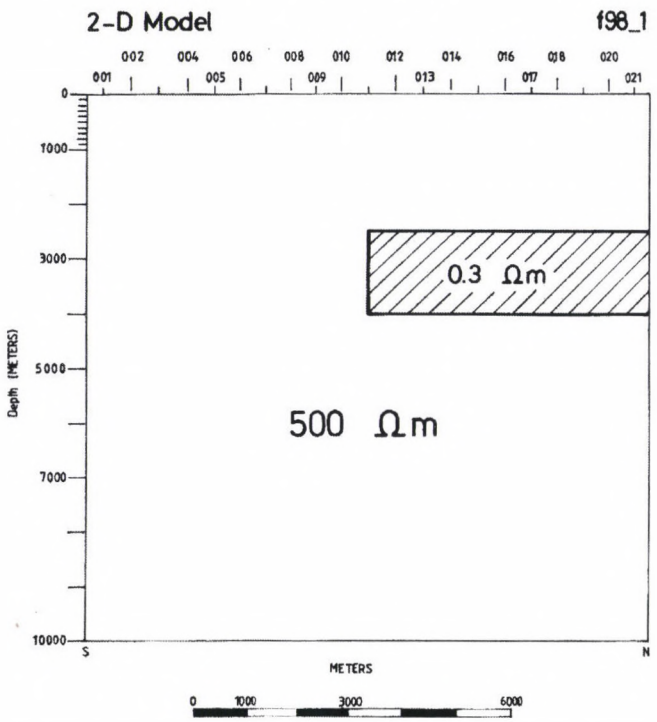


Fig. 16. Model for the boundary (edge) of the TCA

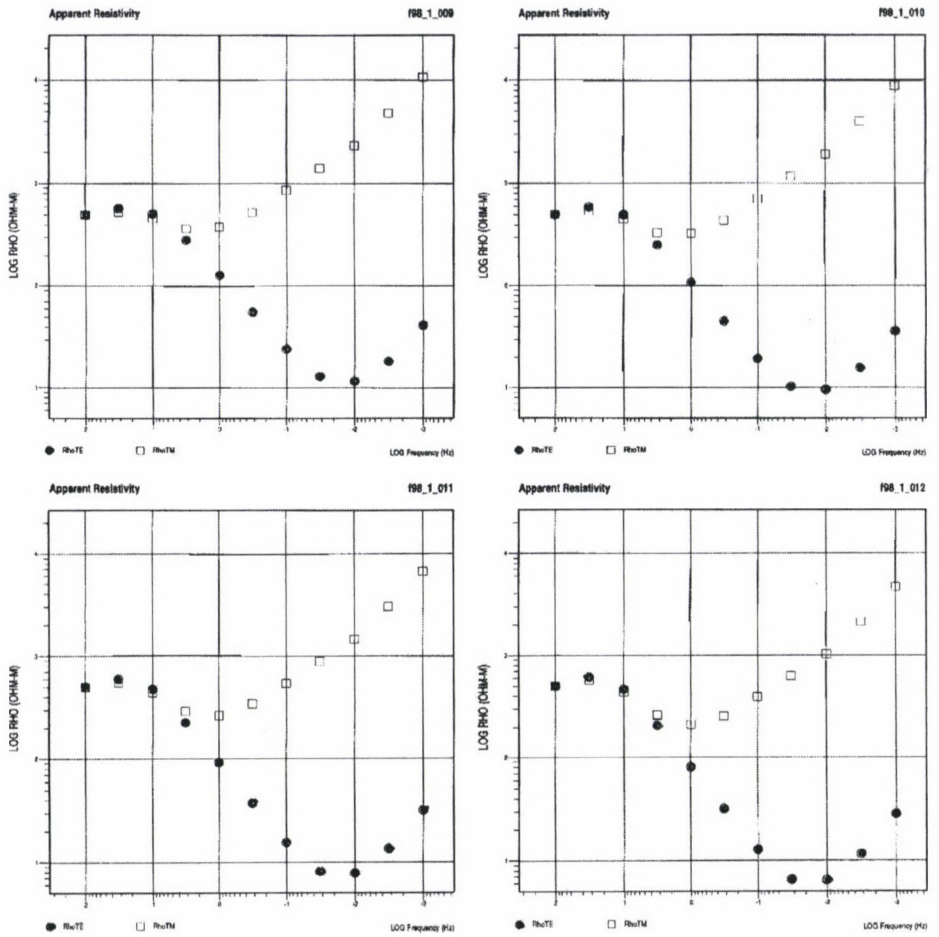


Fig. 17. Sounding curves of TE and TM mode in case of model of Fig. 16

MT anisotropy" over a large area both in telluric and magnetotelluric results. As a consequence direction of resistivity maxima is N-NW–S-SE. As demonstrated by 2D forward modelling, the effect results in the case of TM mode in curve branches ascending high above the medium resistivity. Inversion of these curves provide quite unrealistic 1D structures.

None of the treatments, including the Groom-Bailey and Bahr decomposition, is able to produce sounding curves suitable for 1D inversion, moreover, they increase the uncertainty of data, especially over $T = 1000$ s. Any kind 1D inversion of these curves can provide, even in the best case, only a very rough approximation of the real 1D structure. 3D measurements and 2D/3D inversion may certainly result in better solution.

Acknowledgement

The authors thank for the grant of the Hungarian Research Foundation (OTKA T014882 and T029443) which helps to carry out this study.

References

Ádám A 1968: *Geofizikai Közlemények*, 17, 39–50.
Ádám A 1969: *Acta Geod. Geoph. Mont. Hung.*, 4, 187–197.
Ádám A 1971: *Geonómia és Bányászat*, 4, 297–308.
Ádám A 1976: In: Geoelectric and Geothermal Studies (East-Central Europe, Soviet Asia), KAPG Geophysical Monograph, A Ádám (ed.), Akadémiai Kiadó, Budapest, 547-561.
Ádám A 1996: *Acta Geod. Geoph. Hung.*, 31, 191–216.
Ádám A 1998: *Acta Geod. Geoph. Hung.*, 33, 187–213.
Ádám A, Verő J 1967: *Geofizikai Közlemények*, 16, 25–52.
Bahr 1988: *J. Geophys.*, 62, 119–127.
Berdichevsky M N, Dmitriev V I 1976: *Acta Geod. Geoph. Mont. Hung.*, 11, 447–483.
Chen Leshou 1961: Study of the electromagnetic short period variations in Tihany with special emphasis on the design of the instrument recording the geomagnetic pulsations. Thesis of Candidate degree (in Hungarian)
Fischer G, Szarka L, Ádám A, Weaver J T 1992: *Geophys. J. Int.*, 180, 778–786.
Groom R W, Bailey R C 1989: *J. Geophys. Res.*, B, 93, 1913–1925.
Majoros Gy 1983: *Acta Geol. Hung.*, 26, 7–20.
Semenov V Yu, Ádám A, Hvozdara M, Wesztergom V 1997: *Acta Geod. Geoph. Hung.*, 32, 151–168.
Verő J 1972: *Acta Geod. Geoph. Mont. Hung.*, 7, 333–351.

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 195–204 (2000)

EARTHQUAKE TIME CLUSTER IN NORTH-EAST INDIA DURING FEBRUARY TO APRIL 1988

D SHANKER[1], BHAWANI SINGH[2], V P SINGH[3]

[Manuscript received September 23, 1998]

During the time interval of 3 February to 20 April 1988 four independent main shocks occurred in North-East (NE) India: 03/02/1988, M = 5.5; 06/20/1988, M = 5.8; 15/04/1988, M = 6.2 and 20/04/1988, M = 5.8. The seismicity rate within this 78 day interval is increased by a factor of 15 and 6 with respect to the mean, long term seismicity for M $\geq$ 5.5 and M $\geq$ 6.1, respectively. In terms of probability, it has been found out that the probability of observing by chance four events of M $\geq$ 5.5 or one events of M $\geq$ 6.1 in NE-India is equal to only 0.4243 and 0.3680, respectively. These results imply that the observed seismicity has a non-random time clustering. Similar earthquake time clusters were identified to have occurred in NE-India in 1930 and 1951. A triggering mechanism has been proposed to interpret the earthquake clustering: the first event of the earthquake sequence produces transient stress changes that cause an acceleration to the static stress loading, and then to seismic failure, to remote highly pre stressed regions.

Keywords: Coulomb's law; Poisson model; static stress; time clustering; triggered earthquake; triggering mechanism

Introduction

The NE-region of India is seismically active as two large earthquakes of magnitude $M \geq 8.4$ occurred in this zone since 1912. Besides, the area shows wide-spread distribution of seismic activity where a large number of earthquakes of $M \geq 6.0$ occurred during the past hundred years. Deviation of seismicity rates from the normal before and after large earthquakes has been reported by Khattri and Wyss (1978). The NE-India is geologically and tectonically extremely complex. Several authors have described the geology and tectonics of the region (Desikachar 1974, Evans 1964, Nandy 1980, Nedoma 1997). Seismicity of this region could be considered as the interaction of four major tectonic elements (Fig. 1): eastern syntaxis (Zone-I), Arakan Yoma and Naga thrust fold belt (Zone-II), the Shillong plateau (Zone-III) and the Main central thrust and Main boundary fault of the Himalayan Frontal Arc (Zone-IV). However, the overall seismic activity of the region may be assumed to be dependent on stress generated and resistance offered in each zone.

From 3 February 1988 to 20 April 1988, that is within a time interval of only 78 days, four strong independent earthquakes with magnitudes ranging between 5.5 to 6.2 occurred in the NE-India. The focal parameters of these earthquakes

[1]Department of Earthquake Engineering, University of Roorkee, Roorkee–247667, India, e-mail: dayasfeq@rurkiu.ernet.in
[2]Department of Civil Engineering, University of Roorkee, Roorkee–247667, India
[3]Department of Geophysics, Banaras Hindu University, Varanasi–221 005, U.P., India

1217-8977/2000/$ 5.00

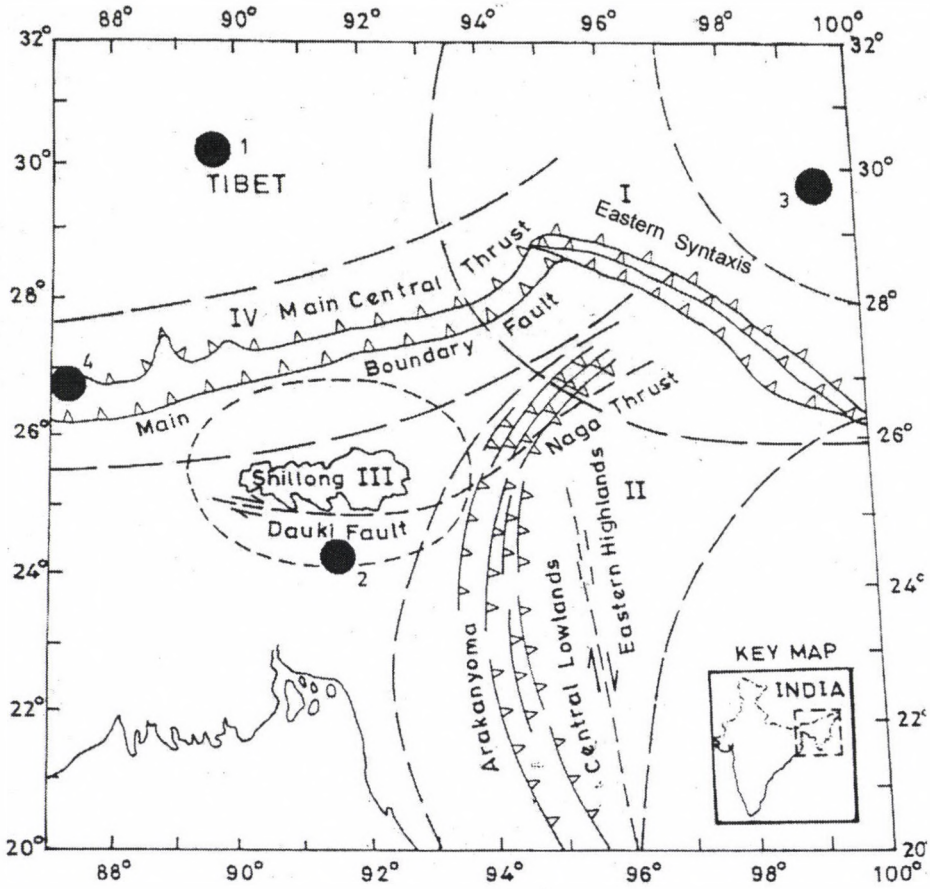


Fig. 1. Tectonic map of NE-India and its vicinity. The time cluster of Feb.-April 1988 is also shown. Solid circles 1, 2, 3 and 4 indicate the epicentres of the mainshocks reported in Table I, respectively

are listed in Table I, while their epicentres are shown in Fig. 1. The occurrence of the February to April 1988 earthquake sequence raises two important questions: one, whether the earthquake sequence is accidental in time and secondly, if the earthquake sequence is non-random in time, then what is the plausible geophysical mechanism to interpret the earthquake time clustering?

In this paper an effort has been made to respond to these two questions in terms of probability and, then, to describe a geophysical triggering process accounting for the earthquake time clustering.

Table I. Focal parameters of the earthquakes involved in the three time clusters. ϕ_N^o = geographical latitude and λ_E^o = geographical longitude. Last column indicates the distance between the epicenter of a triggering event (No. 1) and the epicenter of a triggered event

No.	Date	Origin time	ϕ_N^o	λ_E^o	Depth (km)	M	Distance (km)
			Time cluster of 1988				
1	February 3	17:49:58.6	30.23	89.96	10	5.5	0.0000
2	February 6	14:50:42.7	24.65	91.52	15	5.8	643.13
3	April 15	20:34:11.7	29.98	99.24	33	6.2	1119.22
4	April 20	06:40:22.6	26.94	86.61	28	5.8	521.18
			Time cluster of 1951				
1	November 18	09:35:37.0	30.5	91.5	60	7.9	0.0000
2	December 21	08:37:26.0	26.5	91.7	*	6.5	444.56
3	December 25	10:06:56.0	31.0	90.5	*	6.3	124.10
			Time cluster of 1930				
1	July 02	21:03:42.0	25.53	90.0	60	7.1	0.0000
2	September 21	23:04:14.0	25.83	98.4	*	6.5	932.99
2	September 22	14:19:11.0	25.03	94.0	*	6.3	447.46

Non-randomness seismicity behaviour of North-East India

The present existing literature shows that usually after removing aftershocks/foreshocks the simple Poisson (random) model is applicable for the description of the earthquake process in time (Papadopoulos and Voidomatis 1987, Dionysiou and Papadopoulos 1992, Singh et al. 1994). However, main shock occurrences might be also quasi-periodic or clustered (Kagan and Jackson 1991). In general, the Poisson model is useful for the seismicity time analysis except when the process is dominated by a single source (Cornell and Winterstein 1986). Then, models with memory are applicable. For example, in single fault segment and at simple plate boundaries the time distribution of strong or large, shallow main shocks appear to be adequately described by time-dependent model (Shimazaki and Nakata 1980, Wesnouski et al. 1984, Nishenko and Buland 1987).

From a number of statistical tests it has been concluded that in the different seismotectonic environments (Zone-I to Zone-IV), the main shock occurrence is stationary and random depending upon the time interval and magnitude class considered (Singh et al. 1994). This is possibly due to the high seismotectonic complexity and crustal heterogeneity. The random behaviour of NE-India region examined here shows a seismotectonically independent earthquake generation in each one of the seismogenic sources in a particular segment (Shanker and Singh 1996). Some main shock clustering, however, may occasionally occur. Dionysiou and Papadopoulos (1992) concluded that main shock clustering may result because of the occurrence of multiple events or triggering effects.

Quantitative measurement of seismicity

To examine quantitatively the February-April 1988 seismicity of the NE-India and to reply first question addressed in the introduction, a scanning of the earthquake catalogue has been made for the area 20° N–32° N and 88° E–100° E. The time interval from 1912 onwards has been considered. The total interval covered equals to 76-years (1912–1987). The examination showed that within this timespan 92 main shocks of $M \geq 5.5$ have occurred while the number of events $M \geq 6.1$ has been equal to 54. The above figures imply that the mean seismicity rates for $M \geq 5.5$ and $M \geq 6.1$ are equal to $\overline{r}_{5.5} = 3.317 \times 10^{-3}$ events/day and $\overline{r}_{6.1} = 1.947 \times 10^{-3}$ events/day, respectively. However, the corresponding seismicity rates observed during the 78 days of the earthquake (3 February–20 April 1988) are $r_{5.5} = 4/78 = 5.128 \times 10^{-2}$ events/day and $r_{6.1} = 1/78 = 1.282 \times 10^{-2}$ events/day. This quantitative measurements suggest that in the 3 February–20 April 1988 time-interval, very high seismicity rates occurred that are increased by factor of 15 and 6 with respect to the long-term mean rates for $m \geq 5.5$ and $M \geq 6.1$.

Probabilistic measurement

To express this result in terms of probabilities, the Poisson (random) model has been adopted as an adequate description of the main shock time distribution:

$$P(x) = \frac{\exp(-rt) \cdot (rt)^x}{x!}, \tag{1}$$

where $P(x)$ is the probability for a number of x events to be observed in a time interval t, given that the mean seismicity rate equal to r events/time unit. From Eq. (1), one may easily find out that the probability for observing four earthquake events of $M \geq 5.5$ in NE-India in a time interval of 78 days is equal to $P_{5.5}(4) = 0.195$, while the probability of observing one event of $M \geq 6.1$ within the same time interval is $P_{6.1}\ (1) = 0.3680$. These very low probabilities indicate that the February-April 1988 main shock sequence occurred far beyond the random chance which implies that the particular earthquake sequence has been a non-accidental time cluster.

On this basis, the available earthquake catalogue of NE-India has been searched in order to investigate past time cluster of earthquake main events. The result is summarised in Table I and illustrated in Fig. 2. More precisely, in 1930 an earthquake time cluster occurred with three main shocks of M = 6.3–7.1 within a time interval of only 83 days. Twenty-one year later another earthquake time cluster took place with three main shocks of M = 6.3–7.9 within a time interval of 39 days. The long-term seismicity rates for $M \geq 6.3$ and $M \geq 6.5$ are equal to $\overline{r}_{6.3} = 11/18 = 1.6743 \times 10^{-3}$ and $\overline{r}_{6.5} = 28/38 = 2.018 \times 10^{-3}$ events/day, respectively. The observed seismicity rates within the 83 days and 39 days time interval mentioned earlier are $r = 3/83 = 3.61 \times 10^{-2}$ and $r = 2/39 = 5.128 \times 10^{-2}$ events/day. As such the rates increased by factor of 21 and 24 with respect to long-term events were observed. The probabilities for observing by chance three main shocks of $M \geq 6.3$ in a 83 day time interval or two main shocks of $M \geq 6.5$

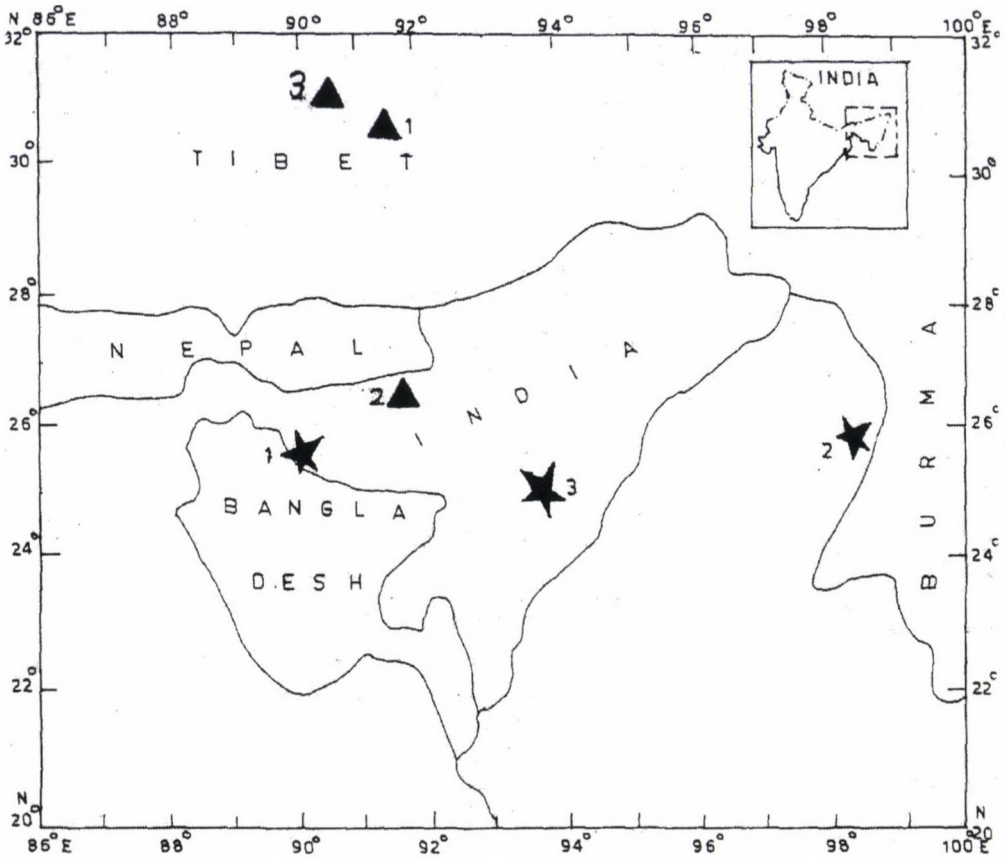


Fig. 2. The time clusters of Nov.-Dec. 1951 and Sept. 1930. Solid triangles 1, 2 and 3 indicate the epicentres of the events reported in Table I mainshocks, respectively; solid stars 1, 2 and 3 indicate the epicentres of the event in 1930 mainshocks respectively

in a 39 day time interval were found to be $P_{6.3}(3) = 0.2240$ and $P_{6.5}(2) = 0.2710$, respectively.

From the previous observations one may suggest that in the NE-India a triggering mechanism may account for a reasonable geophysical interpretation of the earthquake time clustering.

A plausible mechanism for triggering of earthquake by earthquake

In the last two decades several earthquake triggering processes have been proposed to explain the earthquake occurrences such as tidal forces, loading of artificial lakes and others. However, the possibility for an earthquake triggering other earthquakes; taking place at adjacent or spatially isolated areas has been rather ignored for a long time in the seismological literature. The phenomenon of a triggering process of this type of the earthquake synchronization has been observed in Japan

(Rikitake 1983, Ohtake 1986) and Greece (Papadopoulos 1988). Very recently, a sudden increase in seismicity within a time period of up to 83 days from the origin time of the M = 7.4 Landers (California) earthquake of 28 June 1992 and at epicentral distances up to 1250 km was observed and studied in detail (Hill et al. 1993, Anderson et al. 1994, Bodin and Gomberg 1994, Gomberg and Bodin 1994, Johnston et al. 1995). Lomnitz (1996) has tested a 90-year catalogue of shallow world earthquakes ($M > 7$) and found that triggering seismicity is commonly occurring at epicentral distances of 300 to 1000 km. In Taiwan the number of local earthquakes of $M \geq 4.5$ has been observed in the 15 days following a large $M \geq 6.5$ event occurring around the Philippine sea plate icnreases systematically at distances up to 300 km implying a mechanism of remote seismicity triggering (Wen et al. 1996).

In Greece, several cases of earthquake time coincidence were considered as triggering effect resulted by isostatic compensation (Galanopoulos 1995a, 1995b). Papadopoulos (1988, 1994) have made similar observations and tested the statistical significance of his observations and also evaluated them in terms of increased shear stress at the place of a strong earthquake because of the stress released by another earthquake in another, remote seismic region. Correlation between Greek and Global or between Greek and Calabrian seismicities have also described as being statistically valid (Bath 1984, De Natale et al. 1988, Mantovani et al. 1987).

The proposed geophysical model of triggering

A plausible geophysical mechanism explaining remote earthquake triggering is yet unknown. Recently, the remote triggering of seismicity as a result of the large Landers earthquake initiated an increased interest on this topic and some relevant hypotheses were proposed. Hill et al. (1993) suggested that the propagation of dynamic strains caused by the triggering event may account for the triggering effect. Similarly, according to Anderson et al. (1994) the cause of triggering could be a low-frequency pulse (period > 15 sec) of dynamic shear strain that initiates creep event at hypo central depth and that some of these creep events accelerate to seismic failure. Another possibility was presented by Bodin and Gomberg (1994), who suggested that fault connectivity can increase static strain changes and cause seismic failure at distances of at least 280 km. They have also attempted a more quantitative approach and concluded that if dynamic strains initiate remote triggering, the orientation of faulting most favourable for being triggered by a given strain transient change with depth. Moreover, they noted that estimated strains are too small to cause Mohr-Coulomb style failure unless the fault was exceptionally weak, and/or the dynamic strains trigger other process that lead to failure.

An alternative model has been proposed for the triggering of earthquake. The in situ stress data indicates that the in situ state on the ground level in rocks and at point close to ground are very near to the state of failure (Sheorey 1995). Figure 3 shows the plot of the data. This appears to be valid in young mountainous regions which are rising up due to tectonic stresses. Zoback et al. (1993) have inferred from deep in situ stress measurement that upper crust is in a state of failure. Thus slight increase in tectonic stresses may cause fault slips.

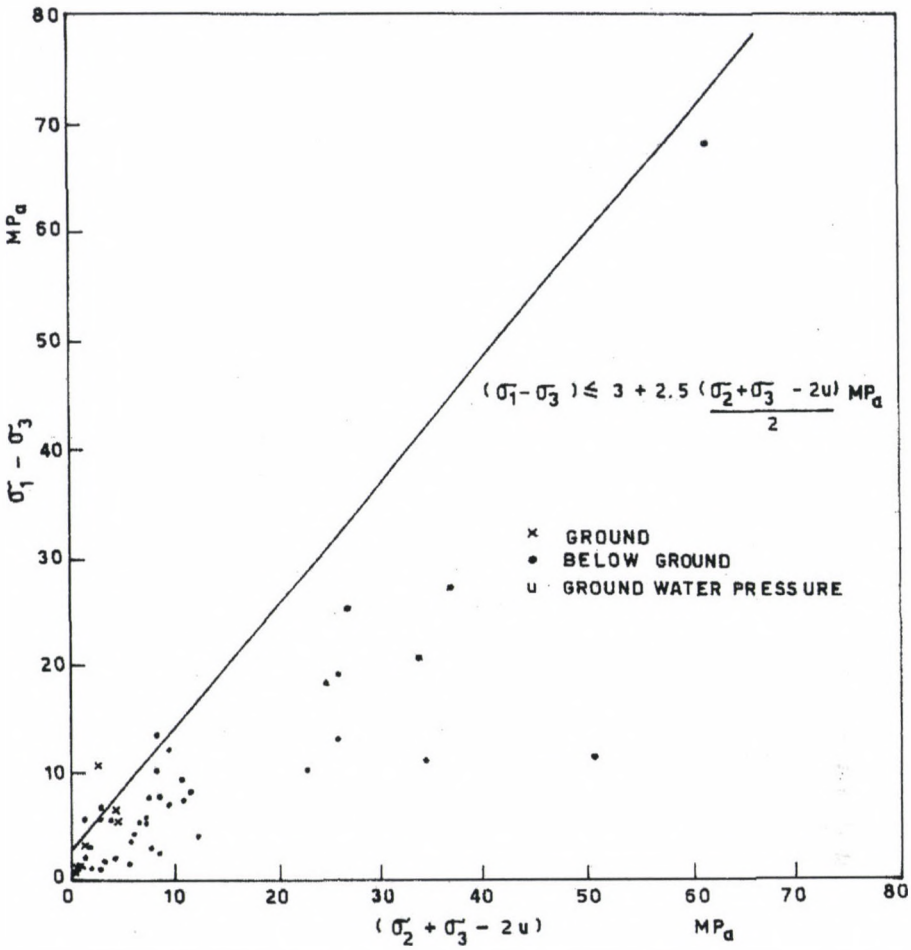


Fig. 3. Plot of in situ stresses showing state of failure at ground level. Here σ_1 represent major in situ principal stress in the rock mass, σ_2 intermediate in situ principal stress in rock mass, σ_3 minor in situ principal stress in the rock mass and u indicates pore water pressure in the rock mass

A sequence of earthquakes will be increasing the accumulated strain and stresses in addition to the static stresses along nearby faults. As such in a tectonically overstressed are such as Himalaya earthquake may cause triggered earthquakes along nearby faults which are already close to the failure limit. Obviously earthquake triggered earthquakes are likely to be of smaller magnitude.

Here a model has been proposed for the explanation of remote earthquake triggering observed in NE-India. This model is compatible with hypothesis of the low frequency pulse of dynamic shear strain that can cause acceleration to seismic failure in a fault statically prestrained to near failure level. Figure 4 shows possible sequence of earthquakes. Let us assume that a deep seated earthquake I occurs, on boundary of plates, it will send elastic waves which cause accumulation of strains and stresses in addition to the static stresses along shallow thrusts. As more earth-

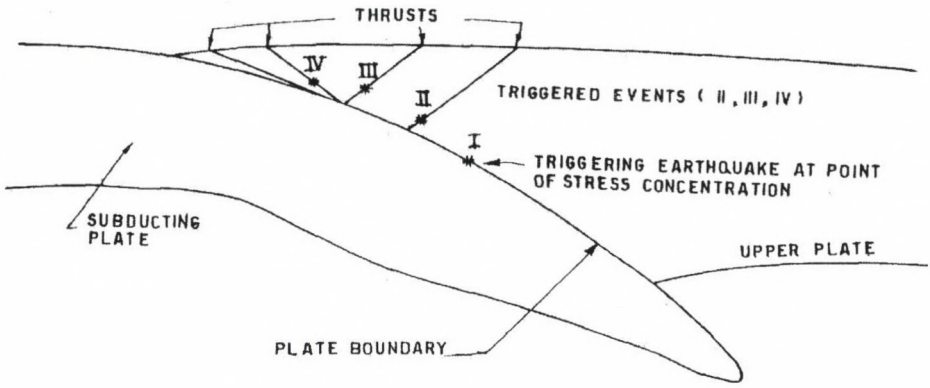


Fig. 4. Illustration of hypothesis on earthquake triggered swarm of earthquake due to accumulation of strains and stresses along plate boundary

quake shocks deform the block of rocks between thrusts, total shear stresses may (dynamic + static) exceed total strength along thrust were instability and earthquake II will occur. The chain reaction of earthquakes III and IV will continue until locked strain static energy in the earth crust is released to safe limit of total stresses locally.

It may be mentioned that the total shear strength along thrust is proportional to the total normal stresses minus static fluid pressure according to Coulomb' law. The fluid pressure may be on account of the ground water or gases coming out of the molten rocks below earth crust.

Discussion

The proposed hypothesis provides an adequate interpretation of some of the basic assumptions of remote triggering of seismicity. One of the main assumption involved in this hypothesis is the cause of transient stress/strain changes by the triggering event. The second assumption in the model proposed is that a triggered event occur only in a highly prestressed region. The deep foreshock of large magnitude sends elastic waves to major thrusts and faults where strains and stresses are accumulated. Then swarm of shallow foreshocks are triggered when stresses exceed shear strength along prestressed thrusts and faults. When a sequence of time cluster of events occurs, like that of February-April 1988 in the NE-India; it is reasonable to suggest that each event except the last event may act as a cause of further triggering process until shear stresses along faults are reduced below the shear strength.

Obviously, the triggering process terminates beyond a limiting distance as the influence of the stress pulse becomes small on statically prestressed faults which cannot slide. The non-linear geodynamic analysis of plate tectonics may predict epicentres of earthquake triggered earthquake swarm and their sequence approximately Nedoma (1997).

The proposed hypothesis may be applied in micro seismic zonation in young

mountainous areas. The major earthquakes ($M \geq 7.5$) are likely to occur at those points along thrust and plate boundaries where excessive accumulation of static strains/stresses has occurred. However, the stress accumulation will be restricted to local weak zones only along the thrust. Thus, there is a possibility that the adjoining areas may be of much lower seismic risk. Such areas of low seismic risk within the zone of high tectonic activity may be demarcated as those areas where major earthquakes have not occurred in the past two hundred years. It appears that major earthquake will continue to occur repeatedly only in the zone of strain accumulation and static stress concentration. Thus, the zones of lower earthquake risk may be demarcated reliably on the basis of seismic history. It is true that major earthquakes have been occurring in NE-Himalaya and Kashmir. As such many areas of lesser earthquake risk may exit in Himalaya in between Bihar and Jammu. Thus attention should be focussed on microseismic zonation.

There is growing feeling among geophysicists and engineers that seismicity of Himalayas and vicinity has perhaps decreased in the last 3-trhee decades. During this period a large number of dam reservoirs have been built by the civil engineers, feeding major faults and thrusts with reservoir water. The static water pressure reduces the shear strength along thrusts and thus causes reservoir induced earthquakes. Experience of mechanics suggests that the static water pressure will prevent excessive accumulation of static shear stresses and strain energy along water charged thrusts. So the large number dam reservoirs may reduce slightly the seismicity of certain areas in the Himalayas, as energy will be reduced by very large number of smaller earthquake.

The important future projects like minor hydroelectric projects and towns should preferably be located in the areas of lower seismic risk. It is time to increase confidence and interest among geophysicists and geoscientist to develop computer modelling of 3-dimensional non-linear geodynamic analysis of cause of triggered earthquakes. These computations will help us improving our understanding of the phenomenon of earthquake induced earthquake swarm.

Acknowledgements

First author is indebted to Dr. S Basu, Professor and Head, Department of Earthquake Engineering, University or Roorkee, Roorkee, India for his helpful suggestion, nice advice and providing excellent computer facilities. Financial support provided by Department of Science and Technology (DST), New Delhi is gratefully acknowledge by the first author.

References

Anderson J G, Brune J N, Louie J N, Zeng Y, Savage M, Yu G Chen Q, dePolo D 1994: *Bull. Seismol. Soc. Am.*, 84, 863–863.

Bath M 1984: *Tectonophysics*, 109, 345–351.

Bodin P, Gomberg J 1994: *Bull. Seismol. Soc. Am.*, 84, 835–843.

Cornell C A, Winterstein S R 1986: In: Seismic Hazard Methodology for the Central and Eastern United States. EPRI Research Report NP-4726, Electric Power Research Institute, Palo Alto, California

De Natale G, Musmeci F, Zollo A 1988: *Geophys. J.*, 95, 285–293.

Desikachar S V 1974: *J. Geol. Soc. India*, 5, 137–149.
Dionysiou D, Papadopoulos G A 1992: *Phys. Earth Planet. Int.*, 71, 154–165.
Evans P 1964: *J. Geol. Soc. India*, 5, 80–96.
Galanopoulos A 1995a: *Praktika Academias Athenon* (in Greek with English abst.), 30, 38–49.
Galanopoulos A 1995b: *Praktika Academias Athenon* (in Greek with English abst.), 30, 49–57.
Gomberg J, Bodin P 1994: *Bull. Seismol. Soc. Am.*, 84, 844–853.
Hill D P, Reasenberg P A, Michel A, Arabaz W J, Berozu G, Brune J N, Brumbaufg D, Castro R, Davbis S, dePolo D, Ellworth W L, Gomberg J, Harmsen S, Hause L, Jackson S M, Johnston M, Jones L, Keller L, Malone J C, Sanford A, Simpson R W, Smith R S, Stark M, Stickney M, Vidal A, Walter S, Wong V, Zollweg J 1993: *Science*, 260, 1617–1623.
Johnston M J S, Hill D P, Linde A T, Langbein J, Bilham R 1995: *Bull. Seismol. Soc. Am.*, 85, 787–795.
Kagan Y Y, Jackson D D 1991: *Geophys. J. Int.*, 104, 117–133.
Khattri K N, Wyss M 1978: *Geology*, 6, 685–688.
Lomnitz C 1996: *Bull. Seismol. Soc. Am.*, 86, 293–298.
Mantovani E, Albarello D, Mucciarelli M 1987: *Annales Geophysicae*, 5B, 143–148.
Nandy D R 1980: *Ind. J. Earth Sci.*, 7, 103–107.
National Geophysical Data Centre Boulder, Colorado, USA seismicity data for NE-India (1912–1990)
Nedoma J 1997: Geodynamic analysis of the Himalayas and the Andaman Island. The 29th General Assembly of the IASPEI-97, Symposium S-3, Geodynamics of the Alpine Mediterranean collision zone, Thessaloniki, Greece
Nishenko S P, Buland R 1987: *Bull. Seismol. Soc. Am.*, 77, 1382–1399.
Ohtake M 1986: *Earthq. Pred. Res.*, 4, 165–173.
Papadopoulos G A, Voidomatis P 1987: *Pure Appl. Geophys.*, 125, 613–628.
Papadopoulos G A 1988: *Tectonophysics*, 145, 343–347.
Papadopoulos G A 1994: In: Electromagnetic phenomena related to earthquake prediction, M Hayakawa, Y Fujinawa (eds), Terrapub, Tokyo, 97–102.
Rikitake T 1983: *Earthq. Pred. Res.*, 2, 75–81.
Shanker D, Singh V P 1996: *Acta Geod. Geoph. Hung.*, 31, 181–190.
Sheorey P R 1995: *Int. J. Rock Mechanics and Mining Sciences and Geomech Abstracts*, 10, 311–335.
Shimazaki K, Nakata T 1980: *Geophys. Res Lett.*, 7, 179–282.
Singh V P, Shanker D, Hamada K 1994: *Current Sciences*, 66, 922–926.
Wen Kuo-Liang, Igor Bresnev A, Shih-Nan Cheng 1996: *Bull. Seismol. Soc. Am.*, 86, 843–847.
Wesnouski S G, Scholz C H 1984: *Bull. Seismol. Soc. Am.*, 74, 687–708.
Zoback M D, Apel R, Baumgartner J, Brudy M, Emmermann R, Engeser B K, Kessels K, Rlschmuller H, Rummel F, Vermic L 1993: *Nature*, 305, 633.

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 205–208 (2000)

WIDESPREAD OCCURRENCE OF NOCTILUCENT CLOUDS

W SCHRÖDER[1]

[Manuscript received October 2, 1998]

The widespread, partially simultaneous occurrence of noctilucent clouds is discussed with respect to the frequency of occurrence of noctilucent clouds being a function of the meteorological conditions at the mesopause.

Keywords: mesopause; meteorology; noctilucent clouds; temperature profile

1. Introduction

The present topic belongs to the morphology of noctilucent clouds (NLC) and it still contains many unsolved problems, both empirical and theoretical ones.

Among many problems to be answered by a theory of NLC is the one about the frequency of occurrence in some extent limited by the empirical data. The maximum of the frequency of occurrence of NLC is during the months June to August, with a maximum generally for the northern hemisphere in July.

The appearance of NLC-s has been coupled several times to the various phenomena of the physics of the atmosphere. One of the aspects, however, namely that of the sometimes widespread and partially even simultaneous appearance of NLC should be illuminated in more details. The present paper discusses this aspect.

2. Selected observations

Observations are now available for nearly one century from the former Soviet Union (Bessonova 1963, Gromova 1963), from Scandinavia (Störmer 1935, Witt et al. 1965), Germany (Jesse 1886), England (Paton 1964, 1965) and North-America (Fogle 1964, 1965) Currie 1963). The selected data are presented in Table I.

An examination of Table I shows: (a) the frequent occurrence of NLC on the same day with visibility in many countries, (b) the continuous observation of NLC-s on several subsequent days. The latter fact is especially clear in the case of the events in 1959, 1963 and 1964. The first conclusion is therefore that NLC-s are observed in certain time intervals in a greater spatial extension. A more precise determination of the covered area is unfortunately impossible due to the lack of corresponding data (site of the observation, UT etc.).

It is to be remarked here as a restriction that the fact of the continuous appearance should not lead to a supposition about some "periodicity". A regular change

[1]Hechelstrasse 8, D-28777 Bremen Rönnebeck, Germany

1217-8977/2000/$ 5.00

Table I. Summary of the observations

Date[1]	Countries[2]	Date[1]	Countries[2]
30. 6. 1908	Cz, E, G, P, S	16. 7. 1963	E, Sc
1. 7. 1908	E, G, S	17. 7. 1963	E, Sc
25. 8. 1916	G, Sc	18. 7. 1963	Sc
30. 7. 1933	G	19. 7. 1963	Sc
31. 7. 1933	G, S	20. 7. 1963	E, G, Sc
30. 6. 1934	G, N, S	21. 7. 1963	Sc
18. 7. 1935	G, S	22. 7. 1963	A, G, Sc
1. 7. 1936	G, S	23. 7. 1963	E, Sc
10. 7. 1951	G, S	24. 7. 1963	Sc
11. 7. 1951	G, S	25. 7. 1963	A, E, Sc
10. 7. 1953	G, S	26. 7. 1963	Sc
11. 7. 1953	G, S	27. 7. 1963	A, G, Sc
6. 7. 1955	E, S	28. 7. 1963	G, Sc
7. 7. 1955	E, G, S	29. 7. 1963	G, I, Sc
9. 10. 1957	S	30. 7. 1963	G
10. 10. 1957	S	31. 7. 1963	C
12. 10. 1957	G	1. 8. 1963	C
26. 6. 1959	G, Sc, S	2. 8. 1963	A
27. 6. 1959	G, S	5. 8. 1963	A, C
28. 6. 1959	S	6. 8. 1963	A, C, I
30. 6. 1959	S	8. 8. 1963	A
1. 7. 1959	E, S	9. 8. 1963	I, Sc
2. 7. 1959	S	10. 8. 1963	I
3. 7. 1959	C, E, S	11. 8. 1963	A, C, Sc
4. 7. 1959	S	12. 8. 1963	A, C
5. 7. 1959	E, S	28. 6. 1964	C, G, Sc
6. 7. 1959	S	29. 6. 1964	C, Sc
7. 7. 1959	S	30. 6. 1964	A, C, E
8. 7. 1959	E, S	1. 8. 1964	A, C, Ni
10. 7. 1959	E, S	2. 8. 1964	A, C, N
11. 7. 1959	S	3. 8. 1964	A, C, Sc
12. 7. 1959	S	4. 8. 1964	Sc
13. 7. 1959	S	6. 8. 1964	A, Sc, I
14. 7. 1959	E, G, S	7. 8. 1964	A, Sc, I
15. 7. 1959	Sc, S	8. 8. 1964	A, C, Sc
7. 8. 1961	G, S	9. 8. 1964	A, C, Sc
11. 7. 1963	A, E, G, Sc	10. 8. 1964	Sc
15. 7. 1963	E, Sc	11. 8. 1964	A, C

[1]Dates indicate one night
[2]Countries: A Alaska, C Canada, Cz Czechoslovakia, E England, G Germany, I Iceland, N Norway, Sc Scandinavia, Ni Netherland, P Poland, S Soviet Union

in the frequency of occurrence cannot be concluded from the available data of the long observation period.

To the results of Table I a consideration of the geographical extent (longitude) can be added. Pavlova (1962) has already published an analysis of this topic with the result that the frequency of occurrence is in the western regions about twice of that in eastern regions is. Her results refer to the area of the former Soviet Union.

Table II. Geographical extent of NLC-s

Date[1]	λ^2	Date[1]	λ^2
9. 6. 1964	15	30. 6. 1964	65
15. 6. 1964	25	5. 7. 1964	70
20. 6. 1964	20	11. 7. 1964	30
24. 6. 1964	20	12. 7. 1964	75
27. 6. 1964	30	15. 7. 1964	40
29. 6. 1964	45	25. 7. 1964	45

[1]Dates refer to one night
[2]Geographical extent in longitude

Paton (1965) published data on the longitudinal extension from the observations of the year 1965 (Table II). His data prove that the extension of NLC-s can be sometimes very great.

3. Discussion

The results from Tables I and II (continuously high frequency of occurrence, simultaneous appearance on several days, geographical extension) should be now discussed. It is already mentioned that NLC are in connection with different other phenomena of the higher atmosphere. Villmann (1962) supposed a connection with geomagnetic activity. Kurilova (1962) discussed the influence of meteorological parameters on the frequency of occurrence of NLC. Grishin (1962) described a mechanism for the transport of water to the mesopause.

The influence of the different conditions on the NLC is very difficult to investigate due to several complications. In each case, however, the conditions at the mesopause and in the surrounding area should be taken into account. Similarly to dynamic meteorology, one has to apply the method "extrapolation to ideal cases" (Ertel 1938). There is an additional difficulty when considering the problem of NLC-s, which is unknown in a similar extent in the physics of the atmosphere in general, namely the lack of knowledge about the conditions at the mesopause. That is why many ideas to explain the appearance of NLC-s became unsure.

For the theory of NLC seems the fact of a meteorological control of the frequency especially important. This influence should be found in the temperature range at the mesopause. It was found that the appearance of NLC is coupled to low temperatures. Ascents made in Sweden (Witt et al. 1965) measured temperatures $T_{\min} = 133$ K at an altitude of 84.4 km when NLC was present. In absence of NLC, the temperature was $T_{\min} = 160$ K at an altitude of 91 km and $T_{\min}$ 140 K at 86 km. It follows that the temperature is when NLC-s are present, about 130 K, while when they are absent, the temperature lies in the range 170 to 190 K, and in winter it is about 240 K. The measurements prove at the same time important variations in the temperature of the mesopause, too.

If the appearance of NLC-s would be closely coupled to the temperature then it would be evident that temperatures being advantageous for the appearance of

NLC-s exist during a longer interval. This would mean — taking into account the data published in Table I — that the necessary temperatures a) for a certain area and b) in a greater area have to exist in order to explain the appearance of NLC-s in these time intervals. The mentioned data by Witt et al were obtained during one case of the presence of NLC-s (30. 7. 1963). Simultaneously, NLC-s were observed in Germany, too. The co-ordinates of the two observing sites are: (Sweden: $\varphi = 60.2°\text{N}$, $\lambda = 19°\text{E}$, Germany: $\varphi = 53.2°\text{N}$, $\lambda = 8.5°\text{E}$). In the other case (7. 8. 1963) there are no observations from Northern Europe. In the case of the third ascent (2. 8. 1963) no NLC-s were observed in Northern Europe. At this time, there is only a positive report from Alaska ($\lambda = 129.2°\text{E}$). This is, however, no serious objection, as the difference in conditions of the mesopause can be very different at the two sites (Sweden and Alaska).

The data in Table I show that in these time interval the conditions for the appearance of NLC-s were very advantageous. Taking into account the close connection between temperature and appearance of NLC-s, it follows that such low temperatures had to exist at the mesopause during these times. Moreover, the conditions had to be very stable during the same time, as expected as necessary during the appearance of NLC-s. The appearance of NLC-s is according to our present knowledge so closely coupled to the discussed conditions that one can conclude to the temperature profile at the mesopause from the appearance of NLC-s.

This view is represented also by Paton (1964) who referred to the close connection with temperature and by Currie (1963) who suggested that stable conditions have to co-exist with NLC-s.

As a summary it is to be concluded that for an explanation of the data in Table I, stability of the temperature profile as well as of the general conditions in the region of the mesopause are necessary pre-conditions.

Evidently NLC-s are tools for a better understanding of the dynamic processes at the mesopause; this supposition would be significantly confirmed by adequate rocket measurements.

References

Bessonova T D 1963: *Res. of Researches on the Progr. of IGY*, No. 6, 23.
Currie B W 1963: *Canad. J. Phys.*, 41, 1745.
Ertel H 1938: Methoden und Probleme der dynamischen Meteorologie. Berlin
Fogle B 1964: Cloud Report No. 2.
Fogle B 1965: *Nature*, 107, 696.
Grishin N I 1962: Conference on Noctilucent Clouds, Riga, 107.
Gromova L F 1963: *Res. of Researches on the Progr. of IGY*, No. 6, 64.
Jesse G 1886: *Meteor. Zschr.*, 2, 6.
Kurilova J 1962: Conference on Noctilucent Clouds, Riga, 131.
Paton J 1964: *Meteor. Mag.*, 93, 161.
Paton J 1965: *Meteor. Mag.*, 94, 180.
Pavlova T 1962: Conference on Noctilucent Clouds, Riga, 119.
Störmer C 1935: *Astrophys. Norvegica*, 1, No. 1.
Villmann Ch 1962: Conference on Noctilucent Clouds, Riga, 151.
Witt G, Martin-Löf N W, Smith W S 1965: *Space Res.*, 820.

Acta Geod. Geoph. Hung., Vol. 35(2), p. 209 (2000)

Meeting of the IAG, Section I (Positioning) Special Comission 4, Working Group 2 "Building Structures as Kinematic Systems" Sopron, March 30th, 1999

The Working Group "Building Structures as Kinematic Systems" was organised for the IAG period 1996–1999. The aim of the Working Group was to carry out research to investigate of building motions. The activity of the Working Group was determined in the following mean research topics:

1. Kinematic sensors and their application for measurements of motions,
2. Automation of kinematic measurements,
3. Application of kinematic measurements and methods in engineering geodesy,
4. Development of mathematical and physical models taking into account the properties of the object to be measured and the parameters of the environment,
5. Development of data processing methods.

The Working Group had 18 members from 13 countries. Three international seminars and a meeting of the Working Group were organized during the four year period. Papers presented at the meeting in Sopron were:

Kümpel H-J, Lehmann K, Fabian M: Background tilting of "fixed" structures at shallow sepths in a sedimentary basin
Dede K, Szabó M, Szücs L: Analysis of vertical movements in the area of oil-field Algyő
Katrycz W: A study on the design of an integrated pipe surveying system for the deformation analysis of landfill sites
Kapovic Z, Ratkajec M, Mastelic-Ivic S: Combination of GPS and precise levelling in measurements of vertical bridge deformations
Brimich L: Tilt measurement of the atomic power station in Jaslovské Bohunice
Dede K: Analysis of the Vertical Deformations of a Building
Katrycz W: Vermessung von Deponie-Überwachung
Csepregi Sz: Movement control of buildings under construction in Székesfehérvár
Schäfer M: Determination and modelling of dumping surfaces with real-time-GPS
Mentes Gy: Increasing the resolution of CCD line sensors by image matching

Five of these papers are published on the next pages.

Gy Mentes
chairman

Akadémiai Kiadó, Budapest

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 211–217 (2000)

A STUDY ON THE DESIGN OF AN INTEGRATED PIPE SURVEYING SYSTEM FOR THE DEFORMATION ANALYSIS OF LANDFILL SITES

W KATRYCZ[1]

In the course of a research program on the monitoring of structures which takes place at the Technical University, Braunschweig, we have set ourselves the goal to survey the pipe systems of landfill sites. The intended standard deviation of position is three centimeters on one hundred meters. Here, the process of choosing and planning the adaptation of a system apt to fulfil this task is summarized. We compared three systems: the deflectometer, the Swedish Maxibor system and an inertial measurement system comprising three Honeywell GG 1320 laser gyros and three Q-Flex QA 2000-30 accelerometers, which finally made the race.

Keywords: deflectometer; deformation; inertial system; landfill sites; monitoring; pipe system

1. The task: monitoring of landfill sites

At the present, the Institute of Geodesy and Photogrammetry, Technical University, Braunschweig, is participating in a research program on the monitoring of structures, the Collaborative Research Center SFB 477, "Life Cycle Assessment of Structures via Innovative Monitoring". This program is divided into two parts: the monitoring of structures of constructional engineering and the monitoring of landfill sites, which is occupying us currently. Aiming at the determination of the structure's entire geometry, we divide our field of work into two lines. On the one hand, we adapt and optimize overground systems for the surveying of the surfaces, and on the other, we develop and adapt underground techniques to survey the drainage pipes (see Fig. 1) of landfill sites. The results of both monitoring techniques will yield data of vital importance for the FEM-models being developed by other participants in the research program.

One problem we are facing is caused by the fact that new landfill sites are not being put in service any more. So we had to give up our original hopes, to be able to conclude our project with recommending the installation of grooved pipes for monitoring. In this case, it would have been possible to take into account either a wider variety of measurement techniques and measurement systems, or simpler and smaller solutions.

Given the fact of having to survey only existing landfill sites, we will have to introduce our system into the drainage pipes of the sites. These are regularly being surveyed with hydrostatic levelling systems. An institute which is currently using such a system is the Institut für Wasserbau at the Leichtweiá-Institut, also at our university. Our colleagues at this institute have provided us with an impressive list of difficulties the surveying of such pipes involves:

[1]Institut for Geodesy and Photogrammetry, Technical University Braunschweig, Gaußstraße 22, D-38106 Braunschweig

1217-8977/2000/$ 5.00

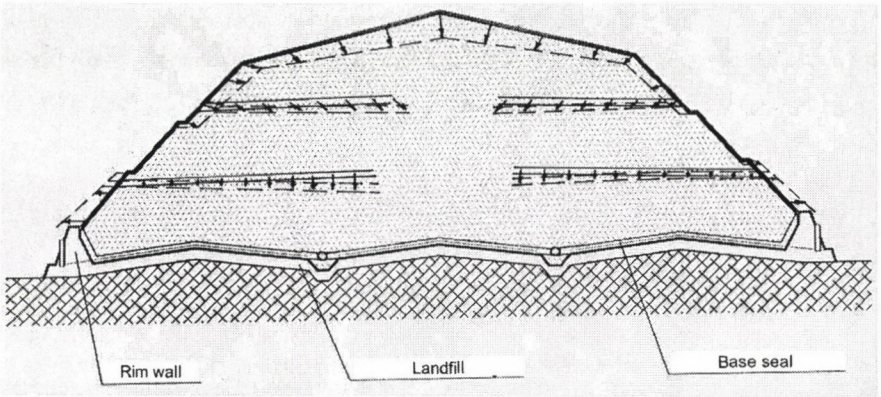


Fig. 1. The task: Determination of deformations of both the surface and the interior of landfill sites

— The interior of landfill sites makes the strictest possible demands on the firedamp proof equipment.

— The drainage pipes in use have standardized diameters of 300 mm, 250 mm or 200 mm. The planned system will only fit into pipes of the first two kinds, as the necessary reassembly of the system we are renting would have exceeded our budget. By assembling the inertial unit accordingly right from the start, these additional costs can easily be avoided.

— The entrance into the drainage pipes brings new problems. At the specific landfill sites the participants of our research program have chosen for monitoring, the drainage system can only be reached through a bend depicted in Fig. 2. To give an impression of the dimensions: the diameter of the pipe is 300 mm and the radius of the bend itself is one metre.

— Additional and quite fundamental difficulties will be caused by the junctions of the pipes, which are either socket joints or welded joints. In case of the first, we have to expect the pipes to have drifted apart within the socket joint, leaving a gap of some two centimeters depth. Moving the pig over this gap will cause severe shock to the equipment. In case of welded joints the seams can produce a similar, though smaller effect. Both cases require shock mounts, adaptive processing techniques and deterministic models of both free fall and shock, which will be fitted into the data by least squares techniques.

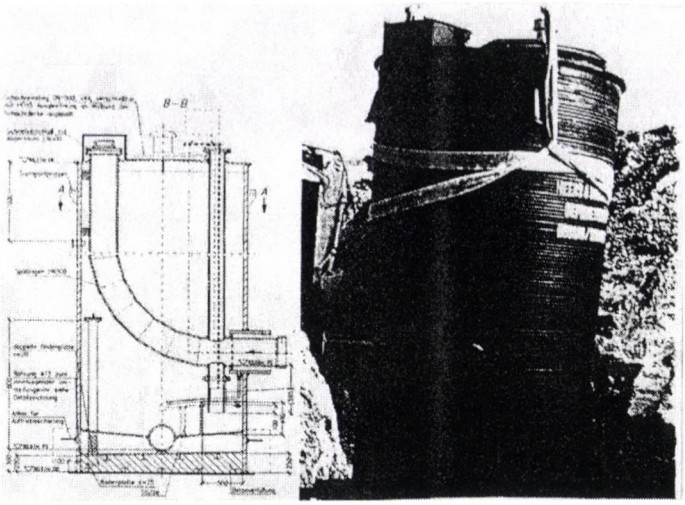


Fig. 2. Left: the bend we will have to pass. Right: mounting of the de-aeration segment

2. A study of different systems

In the application for the research program we had decided to use an inertial measurement system, the NAVIOOL by the Deutsche Montan Technologie. It had two mechanical gyros and the standard set of three accelerometers. It was planned to adapt and extend this system by additional sensors, the final choice of which being one of our aims of research. Unfortunately, this system was called from the market, thus forcing us to begin the study of different systems anew. This paper summarizes the forthgoing of this study.

2.1 The deflectometer

A very well known technique for surveying pipes in three dimensions is the "deflectometer", a variation of the inclinometer. Within our study we focussed on the deflectometer by Glötzl. It consists of two carriages, connected by a ball-and-socket joint, sketched in Fig. 3.

At least two tiltmeters determine the direction of the plumb line, given by the angles Az and ζ. In Fig. 3, these are placed in the rear carriage. Glötzl's deflectometer measures the two angles between both carriages by a system containing a CCD-Array in the rear carriage and a light source fixed to the front carriage at the ball-and-socket joint. The distance s measured along the axis of the pipe is given by the rod-length, by cable-length or by the length of the boring tools. The sixth, and last parameter necessary to navigate in three-space is introduced by constraints in motion. (These constraints are not depicted in Fig. 3.) The propagation of position is done in the following way. Both carriages have, in a well-defined respect, the same length. Given, the pig moves either in a grooved pipe or a similar system constraining its motion, protruding the pig for exactly the amount of the carriage's

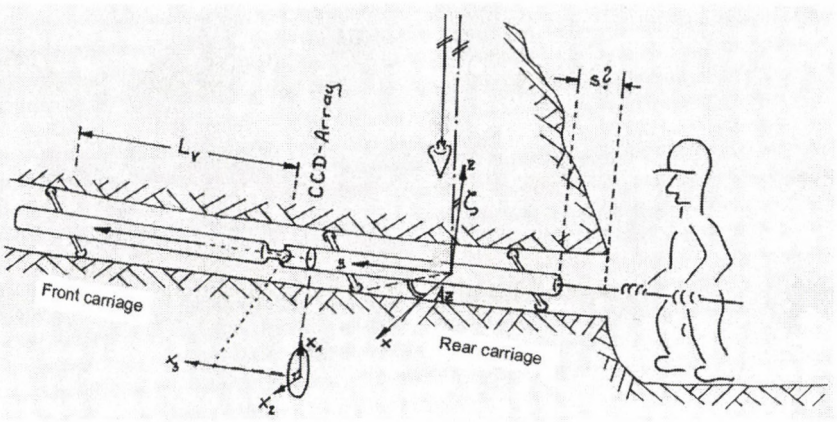


Fig. 3. Solution one – A "Deflectometer"

length will make the rear carriage assume the former position of the front carriage. The measurements of the CCD-Array and of the tilt-meters then yield nothing but a spatial traverse.

As a consequence, the quality of the navigation solution depends on how well the carriages fulfil this condition, i.e., how much the positions and, more important still, the spatial orientations of front carriage before and rear carriage after movement differ. As we have pointed out, good solutions can only be ensured in grooved pipes or pipes that contain a similar system. Both are out of question in our application. In addition to our doubts the system can pass the bend at the de-aeration segment at all, we rated this drawback as fundamental.

2.2 The Maxibor system

A solution for pipe surveying, which in its principle is quite similar to the deflectometer, is the "Maxibor" system by the Swedish company SolExperts. Developed from the preceding system "Fotobor", the processing algorithms have advanced to the point where automated processing is possible immediately after retrieval. The system propagates positions in a way similar to the deflectometer. As in the above case, starting position and orientation have to be known. The length measured along the axis of the pipe again has to be given by rod-, cable-length or by the length of the boring tools.

Differing from the deflectometer, the system contains a CCD-array, which views two illuminated rings with respect to a third ring at focal distance. In addition to that, the first ring carries a bulb, which gives the roll angle of the sond. The first illuminated ring at the fixed distance of 3 m yields the osculating plane and the curvature. The second ring at 6 m yields the derivation of curvature and the torsion. Similar to the principle of the deflectometer Maxibor has to be shifted by a defined length, here 3 m, so the rings swap their positions.

Though fascinating and quite inviting, especially if one toys with the thought

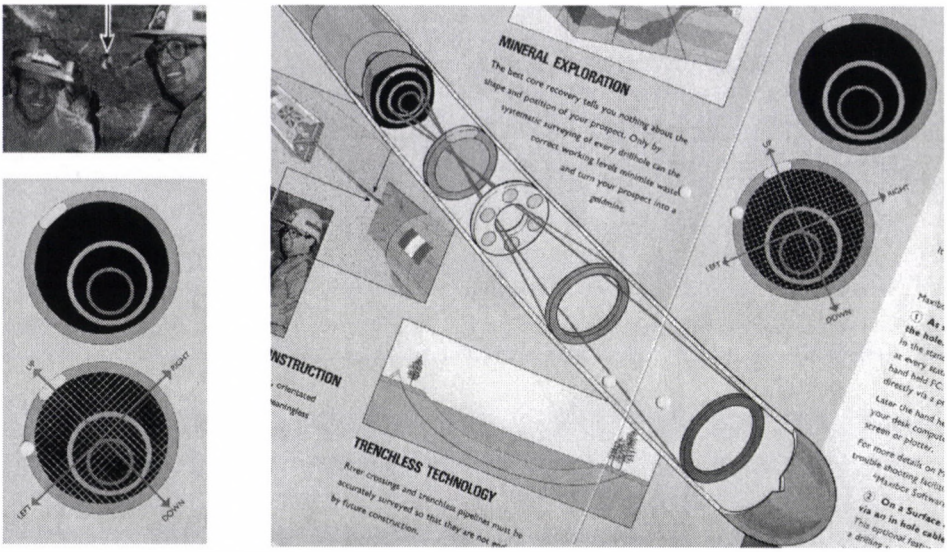


Fig. 4. Solution two – SolExpert's "Maxibor". Right: an overview over the whole system. Left, top: the Eurotunnel – the arrow points at the hole Maxibor went through. Left, bottom: the positioning technique (pictures from a SolExperts handout)

of retrieving and processing the data continuously, this system has disadvantages similar to the deflectometer. Here, the problem of passing the starting bend makes the sond plainly not to handle. In addition to that, the problem of friction could arise, as the sond in a whole is approximately eight metres long.

2.3 An inertial measurement unit

Finally, there is the measurement principle of our original design, inertial positioning. Here, the cost to precision ratio is commonly rated as disappointing. Actual purchase is out of question generally, even given generous budgeting, whilst renting faces the difficulties of a very restricted market, since the number of available high-precision inertial surveying systems is very small. In addition to that, cost increases as an inertial measurement unit cannot yield acceptable results without apt calibration techniques. Optimal calibration can only be done on a three-axial revolving table.

Fortunately, both an inertial unit for rental and a revolving table are available to us via the Deutsche Montan Technologie (DMT), Essen (see Fig. 6). Still, a firedamp-proof pig has to be developed, since security standards in the interior of landfills belong to the highest conceivable. To fit the IMU into a firedamp capsule, the electronic equipment will be remounted. This capsule will be extended by wheels, similar to existing pipe inspecting systems used by another Project of the collaborative research center. Since the latter are self-powered, they are going to drive our pig into the drainage system, while the inertial measurement unit is going to determine its position in stand-alone mode. So, the questions of feasibility the above measurement systems suffered from do not arise here.

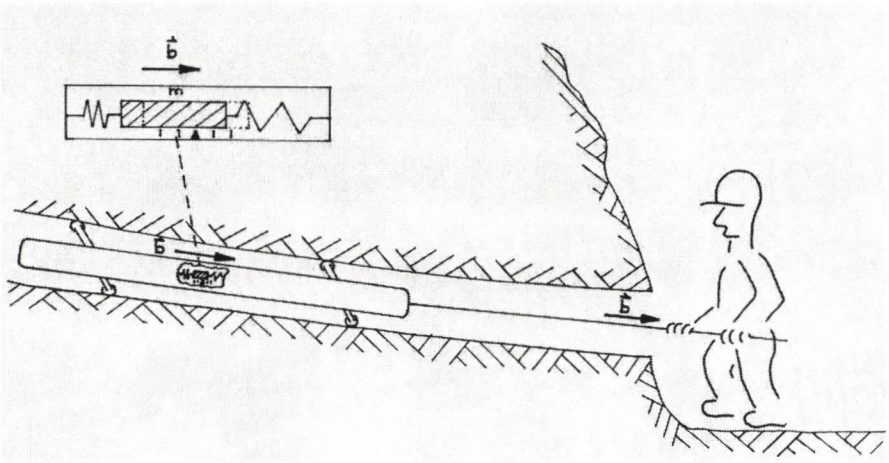


Fig. 5. Solution three – An inertial measurement unit (here only one accelerometer is sketched, for reasons of legibility)

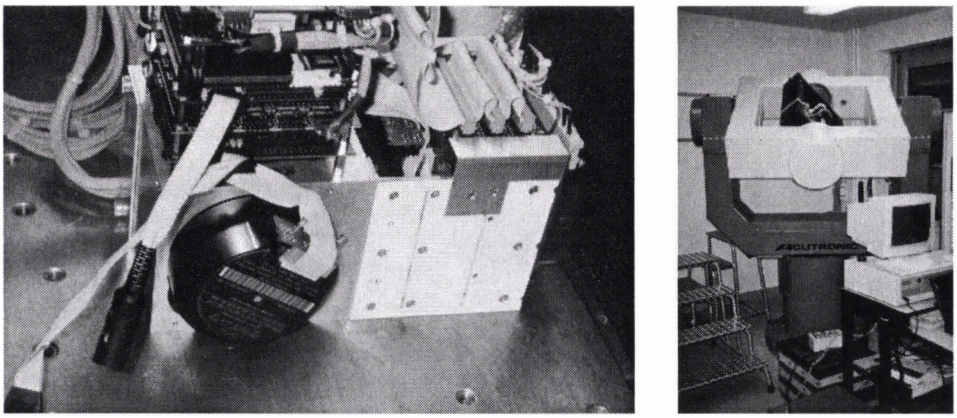


Fig. 6. Left: the inertial measurement unit IMU 5. Right: IMU 5 mounted on a rotating table

It has to be pointed out that precision yielded by state-of-the-art processing techniques will not be rated as satisfactory both by us and the other members of the research center. But, here we see our aims of research.

First, the way of motion in three-space obviously is restricted in a pipe. Thus, our model will postulate that the wheels, to put it in simple words, touch the ground, i.e. that they have contact with the jacket of the pipe. Common inertial navigation techniques assume free motion in three-space, where arbitrary rotations around all axes are admissible. Restricting this motion, or better, modelling it in a deterministic and, later on, in a stochastic way, can be rated as a feasible way to improve precision. The problems arising from possible free fall and subsequent shock at joint sockets or welded joints will be more severe.

Second, the pipe will be measured twice, both when moving in and when moving out. Here, the modelling of the trajectory as a continuous geometrical object will give us the chance of vastly improving filtering performance. This will be done by the postulation that both runs yield one and the same trajectory, of course only after reducing position from the sensor frame to the axis of the pipe.

Finally, one can make rigorous use of existing strategies for calibration both in the laboratory and in the field. The latter will be done with the help of a calibration table that allows for six-, twelve- or even 24-faces-calibration, i.e. a device on which the pig can be mounted in a specific number of orientations, where to each orientation there is a second one that can be obtained by a 180° turn around a horizontal axis. (In the laboratory this is done on the revolving table, but laboratory calibration can loose its validity during transport.) Comparing both values, before and after the 180°-turn, given by the accelerometer currently parallel to the plumb line will yield the offset of this accelerometer. In addition to that, one can of course put great stress on systematic corrections and reductions.

3. Summary

Three measurement systems for the surveying of pipes in existing landfills have been compared. The one system apt for this task was an inertial measurement unit. The remaining two systems cannot pass into the drainage system without violating their own measurement principles, since introducing either grooved pipes or comparable substitutes is out of question if one considers the difficulties the de-aeration segment brings. Still, the inertial system will have to be equipped with sophisticated, and partially not yet developed, processing techniques.

References

Deutsche Montan Technologie – Gesellschaft für Forschung und Prüfung mbH, Essen, Germany, Personal Consultations

Glötzl, Gesellschaft für Baumesstechnik mbH, Rheinstetten, Germany, Handout

Reflex Instrument AB, Vallentuna, Sweden, Handout

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 219–226 (2000)

COMBINATION OF GPS AND PRECISE LEVELLING IN MEASUREMENTS OF VERTICAL BRIDGE DEFORMATIONS

Z KAPOVIC[1], M RATKAJEC[1], S MASTELIC-IVIC[1]

This paper presents the research of the possibilities in parallel application of two methods, terrestrial and satellite, in measuring vertical deformations. It is the matter of a modified precise levelling and kinematic GPS method. Test loads have caused deformations of the bridge over the river Sava near Orasje. By means of precise levelling, all measuring sites on the bridge have been observed, and by means of GPS the middle of the bridge has been observed as related to the reference point. The most interesting comparative results of theoretical and measured deformations have been presented, and the comparative movements of the middle of the bridge measured by means of precise levelling, i.e. GPS in thirty phases of loads.

Keywords: bridge; deformation; kinematic GPS; measurement; precise levelling

1. Introduction

During the War in Croatia, the middle among the three spans of the bridge over the river Sava near Orasje was demolished. Since only the caissons of the river piers remained undamaged, new upper parts and span construction made of a steel continuous support with the opening 85 + 134 + 85 meters were made.

Before the bridge is opened to traffic, test loading is carried out according to the Book of Rules about testing the bridges through test loading. Testing of bridges can be: regular (according to the frequency of testing), normal (according to the size of load), static and dynamic (according to the nature of load), and transitory (according to loading duration).

During the test loading the construction of the bridge is loaded and deformed as the consequence of loading. In order to determine the amount of deformation, a few non-specific geodetic measuring procedures are applied. Since the parameters of loading are not changed in this testing through a certain period of time, the construction is considered stable. The research shows that the construction is still unstable because the bearings have become weaker or because of the influence of temperature (Churchward and Sokol 1981). Hence, in order to obtain relevant amount of deformation, the measurements have to be carried out with high accuracy, but in as short period of time as possible. In classical measurements, high accuracy does not go with the speed in measurements. The experiences prove that less accurate measurements made in a short period of time give higher quality indicators for making conclusions than more accurate measurements obtained in a longer period of time.

[1]University of Zagreb, Faculty of Geodesy, Institute of Engineering Geodesy, Kaciceva 26, HR-10000

1217-8977/2000/$ 5.00

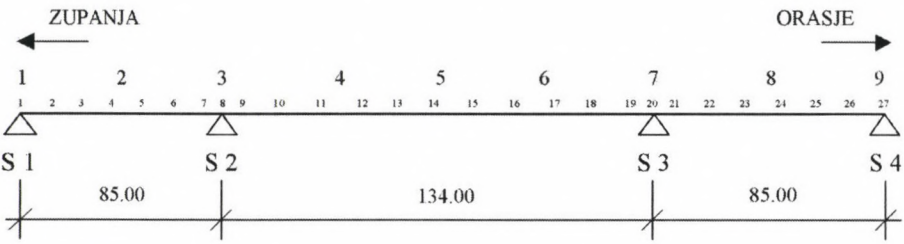


Fig. 1. Plane pole model with knots (1–27) and bench marks (1–9)

A single phase in classical procedures of measuring deformations lasts sometimes more than twenty minutes. In the course of that time, the construction may have some oscillations that cannot be encompassed by means of a classical levelling, and they can have negative influence on measuring accuracy. Depending on a sample rate, kinematic GPS method can encompass vibration, i.e. oscillations of the bridge during the test loading.

2. Calculation of theoretical amounts of vertical deformations and optimal measuring accuracy

A calculation model for determining theoretical values of vertical deformations is based on a method of final elements (Zienkiewicz 1977, Bathe 1990). The bridge construction is designed by segment pole elements of constant rigidity the connection of which between the deformation and knot powers in the plane is defined by the relation

$$[K]^e \cdot \{d\}^e = \{F\}^e \,, \tag{1}$$

where there are:
$[K]^e$ – matrix of the element rigidity,
$\{d\}^e$ – vector of element movement,
$\{F\}^e$ – vector of knot powers.

After connecting the elements into a unity, according to the plane pole model shown on Fig. 1, the system of algebraic equations is formed according to the expression (1) the solution of which yields the vector of element movement $\{d\}^e$. The calculation of movement has been carried out with the software "STAAD-III".

According to the limits (Kapovic 1993)

$$\frac{1}{20} < \frac{\sigma_f}{f} < \frac{1}{10} \,, \tag{2}$$

where there are:
σ_f – standard deviation,
f – measured vertical movement,
the optimal accuracy of measuring bridge deformations is defined. If the construction is well comprised by the calculation model, then the differences between the vector of measured movements $\{F\}^e$ and the vector of calculated movements from

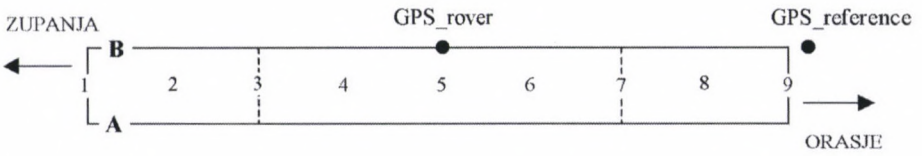


Fig. 2. Ground-plan presentation of measuring spots on the bridge over the river Sava near Orasje

the model $\{d\}^e$ will not be significant. In this case, "f" in the expression (2) can be replaced by "d" for the correct *a priori* assessment of accuracy in measuring the deformations.

3. Presentation of measuring vertical bridge deformations

At the specific places of the bridge (above piers, in the centres and quarters of the field) the measuring spots are marked in two lines along which the vertical movements have been determined (Fig. 2). Two precise levels Leica NA2 with invar rods have been used for measuring with the modified method of precise levelling. The measuring procedure is described more detailed in Kapovic et al. (1998). For kinematic GPS measurement between the reference point placed outside of the bridge and the point 5 on the line B placed in the middle of the bridge (see Fig. 3), two-frequency GPS receivers Turbo SII from the firm Topcon have been used. The method of kinematic relative positioning is described detailed in e.g. Hofmann-Wellenhof et al. (1997).

The measurement lasted more than 7 hours, and the initialisation during the kinematic GPS method lasted about twenty minutes. Sample rate ran up to 10 sec. For the purpose of processing kinematic measurement with GPS, the software GPSurvey, version 2.20 by Trimble was used. RMS values of kinematic relative positioning between the reference point and the point 5 on the line B are shown on Fig. 4.

During the measurement there were mostly more than 4 satellites, a RDOP as the precision measurement did not surpass 5, and it was mostly between 2 and 4 (see Fig. 5).

4. Some results of testing the bridge

Since the testing of the bridge by means of test loading has been carried even in 30 phases, only some interesting results are presented. The comparison between the measured and theoretical deformation in two characteristic phases is shown on Fig. 6.

Since the measured values run up from 81% to 95% of the calculated, one can speak about very well presupposed rigidity of the cross-section, but also about very well *a priori* determined accuracy in measuring vertical deformations.

The comparative results of processed GPS measurements and levelling for the point 5 on the line B for all 30 phases of testing are given on Fig. 7.

Fig. 3. Details in test loading and measuring of deformations

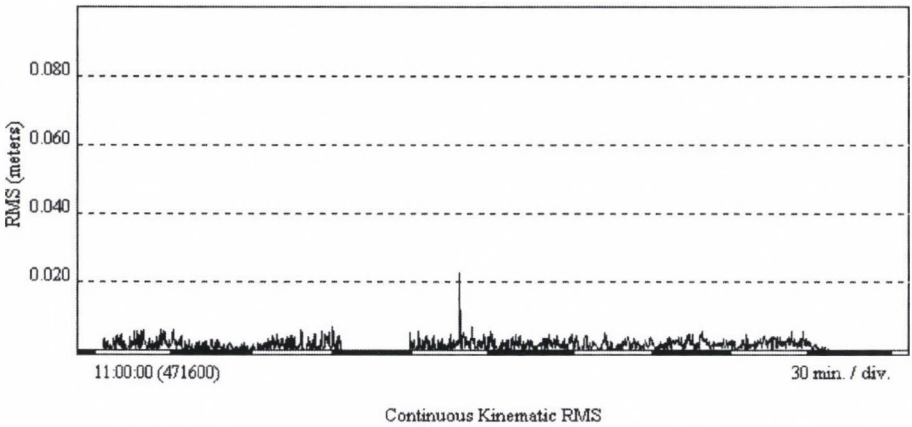


Fig. 4. RMS values of kinematic relative positioning

Standard deviation of measuring by precise levelling depends on a few factors. The lengths of sighting lines are, together with the instrumental errors, an important factor because we are dealing here with the modification of precise levelling. The lengths of the sighting line were from a few meters up to a few tens of meters because the bridge is pretty long. If we consider the expression (2) and look back on the Fig. 6, we can see that vertical movements are negligible on bearings, but run up in quarter to 50% of the complete amount of vertical movement in the middle of

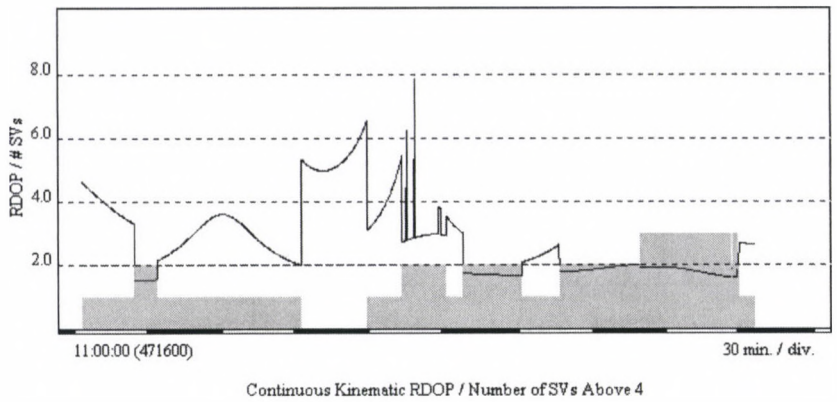


Fig. 5. Satellites and RDOP

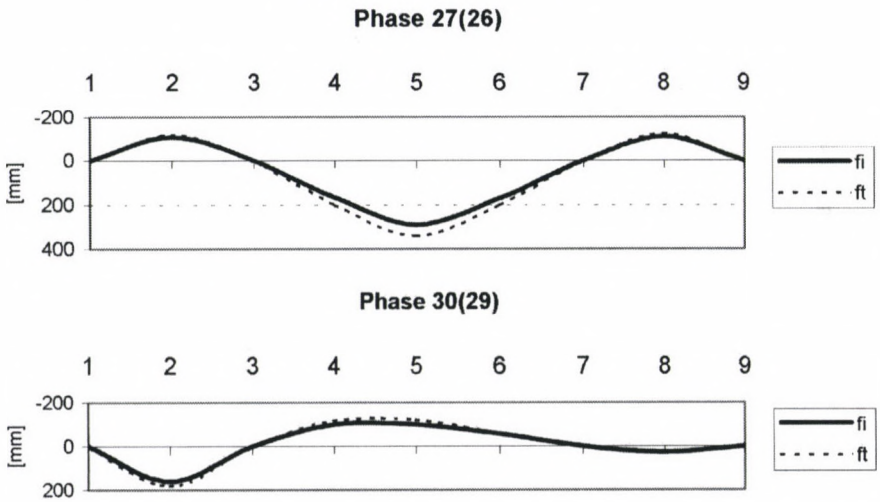


Fig. 6. Presentation of vertical bridge deformation in two phases of loading (fi = measured value, ft = theoretical value)

the span. The essential factor in it is the duration of measuring connected with weakening and vibrating of the construction, because the absolute amounts of the deformations are very big climbing up to almost thirty centimetres (see Fig. 7). The measurement is principally not repeated while testing the bridges through test loading for the purpose of obtaining redundant measurements, so they have not been carried out in this case either.

Considering all these factors, one cannot give the usual evaluation of accuracy in terrestrial measurements on the basis of redundant measurements. Starting from the fact that ten measured movements out of thirty in the point 5 on the line B are within the range of 40 mm to 150 mm and applying the expression (2) for this value in order to estimate the precision in measuring the deformations, we obtain the

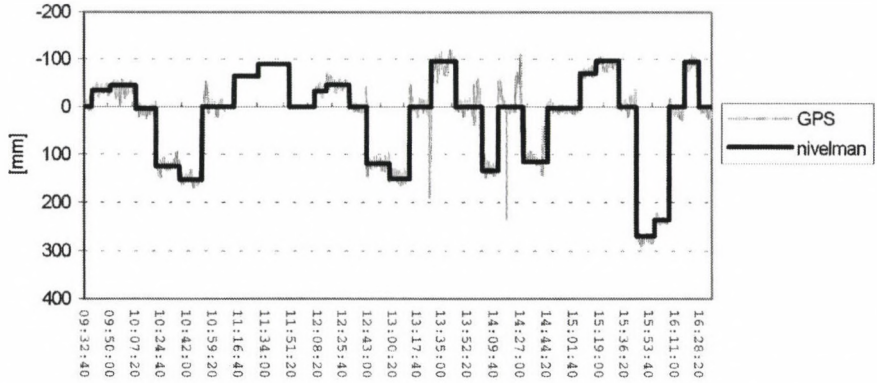


Fig. 7. Comparison of kinematic GPS measurement with the precise levelling in the point 5 on the line B

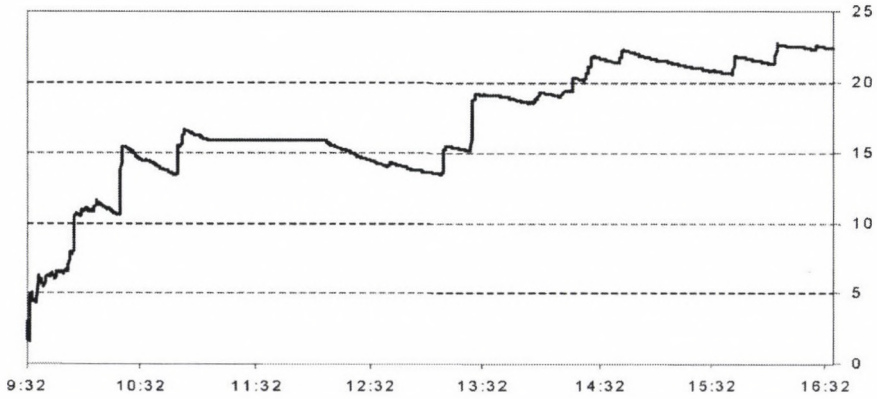


Fig. 8. Standard deviation of levelling from GPS in mm

amounts of RMS being 2 mm to 7 mm. It is assumed that the precision in measuring the deformations has been obtained within these limits, because the experienced and trained experts have made levelling considering all above mentioned factors.

Since the results of levelling a single phase are reduced to the middle of the time interval of this phase, during the transfer into the next phase the standard deviation of levelling data from GPS data jumps (Fig. 8). In comparative presentation of the results of both methods, terrestrial and satellite, one can see significant deviations in the transitions between phases.

This is why the standard deviation is increased in every phase of the change. The comparison has therefore been made with bias free standard deviation, where from the series of difference those having significant values have been excluded (Fig. 9). Thus we have obtained standard deviation of about 6 mm.

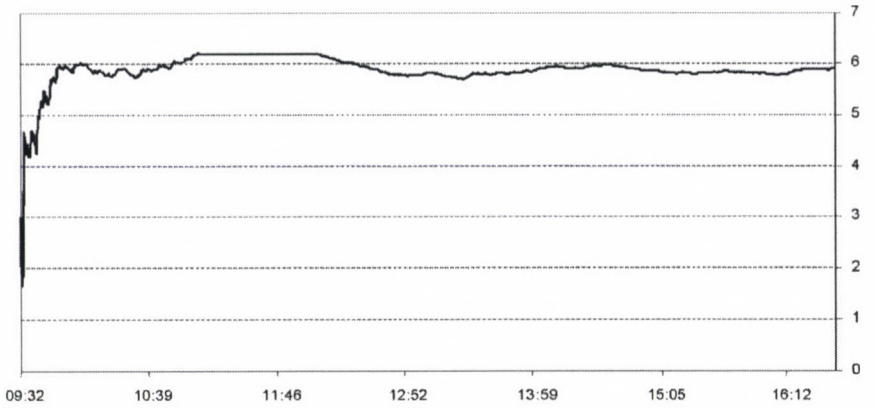


Fig. 9. Bias free standard deviation of the levelling from GPS in mm

5. Conclusion

Measured vertical deformations of the bridge over the river Sava near Orasje correspond very well with the theoretical values of deformations. Since the measured deformations are somewhat smaller than the theoretical, it is obvious that the construction is a bit more rigid than it was assumed through calculations.

It is also obvious that the measured deformations are in accordance with both methods, satellite and terrestrial, in the middle of the bridge. Each method has its own advantages, but also some disadvantages. The applied modified precise levelling is listed into classical methods of measuring vertical deformations. Applying this method it is possible to discredit the bridge that is being tested with test loading, with a sufficient number of points for the purpose of measuring deformations. In the comparison, kinematic GPS method does not make it possible mostly because of the price of GPS receivers. The advantage of kinematic method as compared to the precise levelling is permanent control of the bridge. The precise levelling has the advantage, compared to the kinematic method, in its accuracy. The kinematic method is above all applicable when the deformations are bigger, which is mostly the case with steel bridges.

The presented results of testing the bridge indicate that GPS can also, in the cases when bigger vertical movements are expected, be used to test the construction or to monitor over a larger period of time.

6. References

Bathe K J 1990: Finite-Elemente – Methoden. Berlin

Churchward A, Sokol Y 1981: *Journal of the Str. Div., ASCE*, 107, No. ST. 11, 2163–2167.

Hofmann-Wellenhof B, Lichtenegger H, Collins J 1997: Global Positioning System, Theory and Practice. Springer, Wien-New York

Kapovic Z 1993: Prilog odredjivanju i analizi pomaka i deformacija mostova s posebnim osvrtom na temperaturne utjecaje. Disertacija, Geodetski fakultet Sveucilista u Zagrebu, Zagreb

Kapovic Z, Ratkajec M, Roic M 1998: In: Proceedings of the 1st International Conference of Egineering Surveying, Stanek and Kopacik (eds), Slovak University of Technology Bratislava, Bratislava

Zienkiewicz O C 1977: The Finite Element Method. McGrow-Hill, London

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 227–238 (2000)

MOVEMENT CONTROL OF BUILDINGS UNDER CONSTRUCTION IN SZÉKESFEHÉRVÁR

Sz Csepregi[1]

The paper deals with the monitoring of movements and deformations of buildings under construction. Vertical movements, tilts and deformations of buildings are defined. Different geodetic methods and the aspects of the establishment of the control network and the benchmarks are given. Determination of tilt by means of an osculating plane is detailed described.

Keywords: building; movement; deformation; tilt; levelling; osculating plate; control network

1. About the movements of buildings

The aim of the movement control is to determine:

— the vertical movements,

— the tilt,

— the deformation

of the building.

The movements are represented in Fig. 1.

The vertical movements depend on the increasing load during the construction, the load-bearing capacity of the soil, the type of the foundations and besides, the groundwater, the different mechanical effects (vibration, concussion) are influential factors. If the vertical movement of the building is even and moderate, it causes no damage. According to the empirical measurements the amount of the vertical movement of a ten-storied building is 2–4 cm. In order to measure the vertical movements benchmarks are placed in the walls of the buildings. We have to determine the change of the elevation of the benchmarks compared to the surrounding fixed vertical control points. The procession of the vertical movements is similar to a reciprocal function.

The tilt of the building happens because of the unequal load and the unequal load-bearing capacity of the soil. The amount of the tilt is demonstrable by the unequal change of elevation of the benchmarks. The movements of the benchmarks compared to one another can be a few millimetres or one cm. The tilt makes the stress of the construction unequal. In case of slight tilt the construction takes the unequal stress without any deformation, but a heavy tilt may be damaging.

If the amount of the tilt is so high that the construction cannot take it without deviation or deformation, that causes the distortion of the building. This is the

[1]University of Sopron, College of Surveying and Land Management, Department of Geodesy, Székesfehérvár

1217-8977/2000/$ 5.00

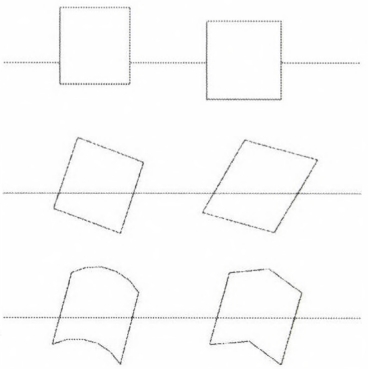


Fig. 1

most dangerous kind of movement. In the worst case it may lead to brake of some parts of the structure. In the course of the vertical control measurements the most significant object is to discover this dangerous kind of movement. Similarly to the tilt measures, we can determine the distortion by measuring the changes of the difference in elevation of the benchmarks. According to the experiences the general amount of the distortion is 2–4 mm.

2. The control net

When we planning the vertical control net, it is useful to consider that it will support the determination of the tilt and the distortion which is very important for the static safety of the building since the rate of these factors is low, the determination needs higher accuracy. The vertical movement is less harmful for the building, so in this case the margin of the vertical movement can be higher.

Considering this, we think that the network of the benchmarks on the walls should be determined as an independent free net. Analysing the independent net we can draw conclusions about the tilt and the distortion. To determine the vertical movement of the building, we have to connect the independent net to the surrounding fixed net (Fig. 2). Monumentation of the benchmarks should be carried out very carefully. Besides the geodetic rules, we should respect the architects' opinion. It is necessary to place benchmarks at the break-lines of the walls and at the dilatational interspaces. It is very important to place benchmarks inside the building. The levelling should be carried out with few turning points. The benchmarks should placed in the foundations, or in the ground floor.

It is also very important to save the benchmarks from damage. Placed at the height of 0.8–1.0 m can remain for a long time. When we replace a destroyed control point, we should consider it as a new one.

The permanent marking of the control point should be done with great care. Only the precise levelling benchmarks are suitable. Their surface — where the rod is put on — should be ground. The benchmarks we used in Székesfehérvár were 4 cm in diameter. They were put in a 12–15 cm long borehole and fixed with concrete-

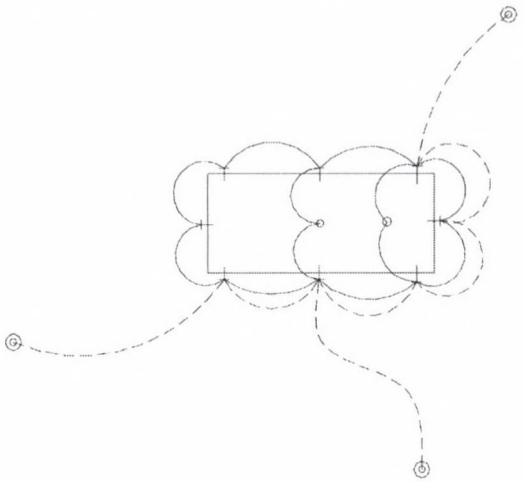


Fig. 2

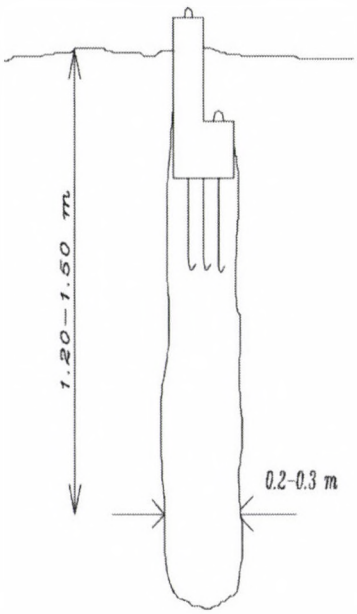


Fig. 3

glue. Screws are also useable, but their head should be grind round. Iron pipes are used only in special cases. The HILTI nails are unsuitable for control measurements. They spring when the rod is put on them, and they are often destroyed after a short time. We think that the screwed-in marks are not suitable.

The fixed control points should be placed out of the building area. It is not recommended to place them in old buildings, except if the building is not to be

demolished. We should use at least 3 fixed points, but they should not be in the same building. In unbuilt areas we can apply a concrete pile buried in the ground, with a prefabricated rock head (Fig. 3). The points should be out of the movement zone. The distance between the building and the point should be three times longer than the height of the building. It is recommended marking the turning points of the levelling lines. It is also useful to mark the stations.

3. About the measurements

In case of control measurements the standard deviation of the determination should be better then the 1/10 part of the movements. As written above all types of movements have different amounts so there are different requirements of accuracy. In case of the tilt and the deformation the standard deviation of the determination should be about 0.2–0.3 mm. The determination of the tilt needs a reliability of 1 mm.

To satisfy these opposing requirements economically, we determine the tilt and the distortion control net with high accuracy, and we connect this independent net to the fixed net with lower accuracy. Determination of tilt and distortion requires precise levelling. For our measurements we used MOM A-31 precise level, to our great satisfaction. Some buildings were measured with Zeiss Ni 002 instrument. This level has a "semi-absolute" instrument horizon and a reversible eyepiece which were very useful during the measurements. We measure out of working time, generally on Saturday and Sunday or tried to avoid the fitting vibrations.

Generally the distance between the instrument and the rod was less than 10 m, often 2–3 m. Because of the short distance we noticed no atmospheric vibration, even around noon. We avoided using turning points except when it was absolutely necessary.

We used only one rod to eliminate the zero error. At the beginning we had not got a 1 m long precise rod, so we used a plotting ruler instead of a rod. We aimed with the wedge-formed wires the graduation lines of the ruler, which we had highlighted with a sticker. In case of so short distances the graduation lines of the 3-m long precise rod are too thick so we can not aim them accurately. The order of the reading was 'backsight-foresight foresight-backsight' in order to eliminate the settlement of the tripod. If we have a 0.5 m long invar rod, we can use it.

The most dangerous source of error is the error of the instrument adjustment. This error can be eliminated when the distances at the backsight and the foresight reading are equal. The difference between the goal of the traditional levelling and the movement control gives another solution. In the first case we determine the elevation of the control points. At the movement control we have the movements of the control points. So if we set up the instrument on the same stations and the error of adjustment is unchanging, it will have no effect on the measurements. This means that it is useful to use the same control net, the same stations and the same instrument during the measurements. Nevertheless, it is necessary to check and adjust the instrument before measuring.

We placed the benchmarks at the height of 0.8–1.0 m so that the 1-m long

ruler should be aimed comfortably. Because of the earthwork sometimes we had to lengthen the tripod. In these cases the height of the transit axis was increased to 2–2.5 m with decreasing the stability.

We connected the independent net to the control net by the rules of precise levelling. We kept the distance between the instrument and the rod under 30 m. To eliminate the effect of refraction we worked in the morning and the evening hours.

During the connecting measurements we used a precise invar rod, which was supported with sticks. These measurements are less accurate than the normal measurements. At the measurements around the building the reliability of a difference in elevation was 0.1–0.2 mm, white at the connection measurements it was 0.5 mm.

First time (at the so-called basic measurements) it is useful to measure twice. Unfortunately it was not always possible.

4. Computation of the net and determination of the movements

We determine the elevation of the benchmarks after every measurement. The method of calculation must be the same during the measurement.

In most cases the precise adjustment is not necessary, a simpler method is also enough. However, the method of computation should be the same during all measurement control. The net of the benchmarks in the building should be computed as an independent net, so that the errors of the connecting measurement should not influence.

To connect the independent net to the surrounding control net, we compute the elevation of only one benchmark, but including several fixed control points. We should leave only those points, which had undoubtedly moved.

The next step is the determination of the movements. Generally it is enough to compute the movement from the basic and the previous measurement.

We often represent the movements by a movement chart, which is constructed by the points of time and the movements.

The axonometric representation contains the amount of the movements, the tilt and the distortion at the same time.

The amount of distortion is well represented by arcs, which are computed along some lines of building. They can be computed from movements of three points, which are on the same horizontal line.

5. Determination of tilt with an osculating plane

The tilt and the distortion of the building can be determined with on osculating plane. The points' movement from the basic position is considered as vertical movement of the points of a horizontal plane.

We fit another plane on the moved points with the best fitting. We determine the tilt and the vertical movement of this plane in the direction of the axes of coordinates. This represents the tilt, the distortion and the vertical movements of the whole building (Fig. 4).

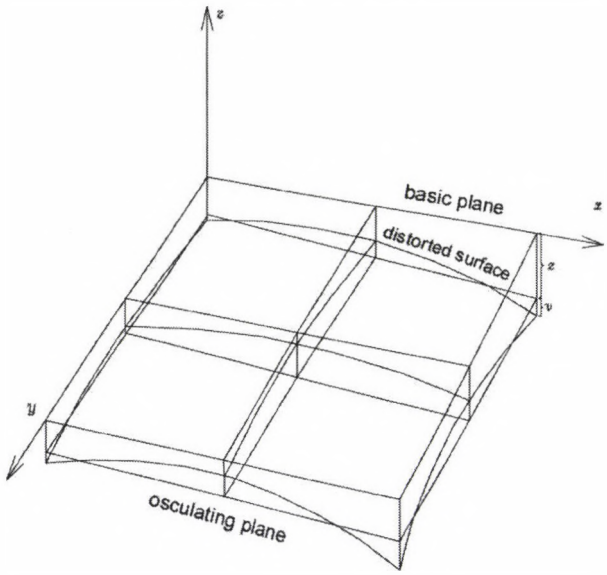


Fig. 4

We have to know the horizontal x, y co-ordinates of the points, which are supposed to be correct. They can be determined with approaching accuracy. Their dimension is the same. The dimension of the incorrect 2 co-ordinates can be different.

We can write the observation equation from the general equation of the plane:

$$Ax_i + By_i + Cz_i + D = 0\,. \tag{1}$$

We divide the equation by C, and mark the quotients with

$$\frac{A}{C} = a \qquad \frac{B}{C} = b \qquad \frac{D}{C} = d \tag{2}$$

the equation of the plane is

$$ax_i + by_i + z_i + d = 0\,. \tag{3}$$

The measured points do not fit on a plane exactly, we mark their distance from the plane with v_i so when we determine the plane, we suppose that the x and y co-ordinates are correct, and we add residuals only to the vertical movements. The sign of the residual is positive when the point is under the plane, so the co-ordinates of the osculating plane can be computed by the following relation: $z + v$.

So the observation equation is

$$-v_i = ax_i + by_i + z_i + d\,, \tag{4}$$

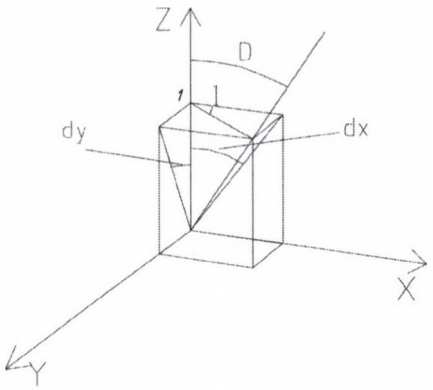


Fig. 5

where a and b are the tangents of the tilt of the plane in the directions of the x and y axes, and d is the shift of plane (Fig. 5). The coefficient matrix and the constant term vector of the normal equations are

$$\mathbf{N} = \begin{bmatrix} \Sigma x_i x_i & \Sigma x_i y_i & \Sigma x_i \\ & \Sigma y_i y_i & \Sigma y_i \\ & & n \end{bmatrix} \qquad n = \begin{bmatrix} \Sigma x_i z_i \\ \Sigma y_i z_i \\ \Sigma z_i \end{bmatrix}, \tag{5}$$

where n is the number of the measured points. If we enter the co-ordinates of the barycentre

$$x_{sp} = \frac{\Sigma x_i}{n} \quad y_{sp} = \frac{\Sigma y_i}{n} \quad z_{sp} = \frac{\Sigma z_i}{n} \tag{6}$$

and the barycentric co-ordinates

$$x'_i = x_i - x_{sp} \quad y'_i = y_i - y_{sp} \quad z'_i = z_i - z_{sp} \tag{7}$$

then in the normal Eqs (5) Σx, Σy and Σz will be zero.

The normal equations are divided to a system of equation with two unknowns and an equation with one unknown.

$$\begin{gathered} \Sigma x'_i x'_i \cdot a + \Sigma x'_i y'_i \cdot b + \Sigma x'_i z'_i = 0 \\ \Sigma x'_i y'_i \cdot a + \Sigma y'_i y'_i \cdot b + \Sigma y'_i z'_i = 0 nd + \Sigma z'_i = 0 \\ nd + \sum z'_i = 0 \end{gathered} \tag{8}$$

invert the coefficient matrix of (8)

$$\mathbf{N}_{-1} = \mathbf{Q} = \begin{bmatrix} \dfrac{\Sigma y'_i y'_i}{det} & -\dfrac{\Sigma x'_i y'_i}{det} & 0 \\ -\dfrac{\Sigma x'_i y'_i}{det} & \dfrac{\Sigma x'_i x'_i}{det} & 0 \\ 0 & 0 & \frac{1}{n} \end{bmatrix} \qquad \mathbf{n} = \begin{bmatrix} \Sigma x' z' \\ \Sigma y' z' \\ \Sigma z' \end{bmatrix}, \tag{9}$$

where

$$det = \Sigma x_i' x_i' \Sigma y_i' y_i' - \Sigma x_i' y_i' \Sigma x_i' y_i'$$

is the determinant of the first four elements. The solution of the system of equations is

$$a = \frac{\Sigma x_i' y_i' \Sigma y_i' z_i' - \Sigma y_i' y_i' \Sigma x_i' z_i'}{\Sigma x_i' x_i' \Sigma y_i' y_i' - \Sigma x_i' y_i' \Sigma x_i' y_i'}, \quad b = \frac{\Sigma x_i' y_i' \Sigma x_i' z_i' - \Sigma x_i' x_i' \Sigma y_i' z_i'}{\Sigma x_i' x_i' \Sigma y_i' y_i' - \Sigma x_i' y_i' \Sigma x_i' y_i'}, \quad d = -\frac{\Sigma z_i'}{n} = 0\,, \tag{10}$$

where a and b are the tangents of the tilt of the plane, and d is the average shift of the plane. We determine the sum of squares of the residuals from the complementary system of equation:

$$\Sigma pvv = a\Sigma x_i' z_i' + b\Sigma y_i' z_i' + \Sigma z_i' z_i'\,. \tag{11}$$

The standard error of unit weight is

$$m_0 = \sqrt{\frac{\Sigma pvv}{n-3}}\,. \tag{12}$$

As we have determined a and b, we can compute the amount (D) and the direction (I) of tilt of the plane.

$$\begin{aligned} \tan D &= \sqrt{a^2 + b^2} \\ \tan I &= \tfrac{b}{a} \end{aligned}\,.$$

After polar conversion we get

$$\begin{matrix} (y) = a & & (t) = \tan D \\ & \to P & \\ (x) = b & & (\delta) = I \end{matrix}\,. \tag{13}$$

We can compute the data of reliability from the matrix of the weighting coefficients

$$m_a = m_o\sqrt{Q_{aa}} \qquad m_b = m_0\sqrt{Q_{bb}} \tag{14}$$

we obtain the reliability of the amount and the direction of tilt by applying the law of the propagation of errors. The weighting coefficients of the amount and the direction of tilt are marked with Q_{DD} and Q_{II}.

$$\begin{aligned} &Q_{DD} = \cos^4 D \left(\frac{Q_{aa} + Q_{bb}}{2} + W\cos 2(I-F)\right) \\ &\text{and} \\ &Q_{II} = \tfrac{1}{t\tan^2 D}\left(\frac{Q_{aa} + Q_{bb}}{2} - W\cos 2(I-F)\right) \end{aligned}\,, \tag{15}$$

where F and W can be computed with the following relations:

$$\tan^2 F = \frac{2Q_{ab}}{Q_{aa} - Q_{bb}} \qquad W = \frac{1}{2}\sqrt{(Q_{aa} - Q_{bb})^2 - (2Q_{ab})^2}\,.$$

It is useful to apply the polar conversion here:

$$\begin{matrix} (y) = 2Q_{ab} & & (t) = 2W \\ & \rightarrow P & \\ (x) = Q_{aa} - Q_{bb} & & (\delta) = 2F \end{matrix} .$$

We have to calculate after every measurement, so it is useful to separate the relations, which do not depend on the amount of movements.

These are the normal Eqs (8) and its inverse (9) and (15) relations. It makes simpler the computation of tilt.

6. Determination of distortion

It is useful to calculate the residuals of all points after the adjustment computation. They can be calculated with the observation Eq. (4).

In case of distortion the amount of residuals depend on the errors of measurement, and also on the deformation of the building.

If the sum of squares of the residuals increases evenly, the building is being deformed. If it does not change, there is no distortion.

Knowing the residuals we can draw an isoline diagram on the ground plan. It is a good way of representing the distortion. In case of distortion the isolines drawn from different measurements change systematically. A growing 'terrain' indicates surely distortion.

The calculations of the osculating plane and the residuals require more work than the recent methods. Nowadays we have programmable calculators, which we can

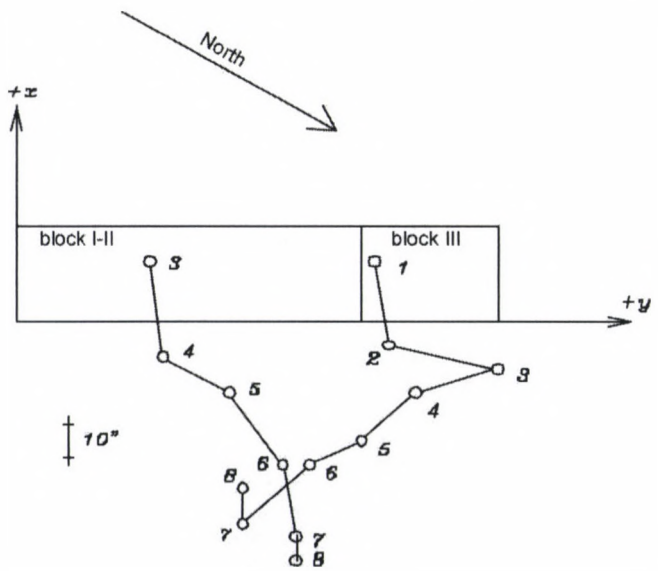


Fig. 6

Table I. The movements of the buildings

Point-Nr.	Coordinates		Movements (mm)				
	Y (m)	X (m)	4. Measur.	5. Measur.	6. Measur.	7. Measur.	8. Measur.
1	39.8	0	–4.01	–10.79	–16.53	–20.18	–21.73
2	30.0	0.0	–6.94	–12.76	–18.56	–22.68	–24.24
3	19.8	0.0	–6.30	–12.66	–18.18	–22.18	–23.97
4	10.3	0.0	–6.10	–10.31	–15.08	–18.46	–20.13
5	0.0	0.0	–3.78	–8.00	–11.26	–14.75	–16.50
6	39.8	5.7	–3.97	–10.65	–16.15	–19.63	–20.98
7	30.0	6.0	–6.12	–11.77	–16.87	–20.69	–22.57
8	19.8	6.1	–6.78	–11.76	–17.05	–20.62	–22.36
9	10.3	5.9	–4.46	–9.72	–13.64	–16.81	–18.10
10	0.0	5.7	–3.01	–6.92	–10.03	–12.97	–14.75
11	39.8	11.1	–3.41	–9.83	–14.61	–17.99	–19.64
12	30.0	10.6	–4.66	–11.01	–16.09	–19.48	–21.12
13	19.8	10.8	–4.84	–10.74	–15.46	–18.66	–20.40
14	10.3	10.6	–3.64	–8.66	–12.23	–14.97	–16.32
15	0.0	11.1	–2.12	–10.67	–8.16	–10.25	–11.87
z			–4.41	–10.08	–14.66	–18.01	–19.64
I			168.3°	149.8°	147.6°	150.8°	151.7°
D			26”	39”	1′00”	1′14”	1′16”
Dx			–26”	–34”	–51”	1′06”	–1′07”
Dy			5”	20”	32”	36”	36”

Table II. The residuals of the osculating plane

Point-Nr.	Residuals (mm)				
	4. Measur.	5. Measur.	6. Measur.	7. Measur.	8. Measur.
1	–1.58	–2.10	–2.60	–3.06	–3.18
2	+0.60	+0.80	+0.96	+0.96	+1.04
3	+1.22	+1.68	+2.17	2.45	+2.55
4	+0.26	+0.24	+0.55	+0.40	+0.37
5	–0.80	–1.08	–1.66	–1.50	–1.45
6	–0.91	–1.26	–1.51	–1.74	–2.02
7	+0.52	+0.81	+0.75	+1.06	+1.31
8	+1.32	+1.61	+2.30	+2.50	+2.60
9	+0.35	+0.67	+0.65	+0.69	+0.34
10	–0.87	–1.22	–1.48	–1.49	–1.36
11	–0.80	–1.20	–1.71	–1.69	–1.61
12	+0.60	+0.79	+1.08	+1.26	+1.32
13	+1.09	+1.54	+2.12	+2.33	+2.48
14	+0.12	+0.39	+0.40	+0.33	+0.08
15	–1.09	–1.68	–2.02	–2.51	–2.49
Σpvv	12.16	23.31	39.04	48.09	51.18
m_0	1.01	1.39	1.80	2.00	2.07
m_D	12”	15”	19”	21”	22”
m_I	9.9°	14.0°	12.4°	10.5°	10.3°

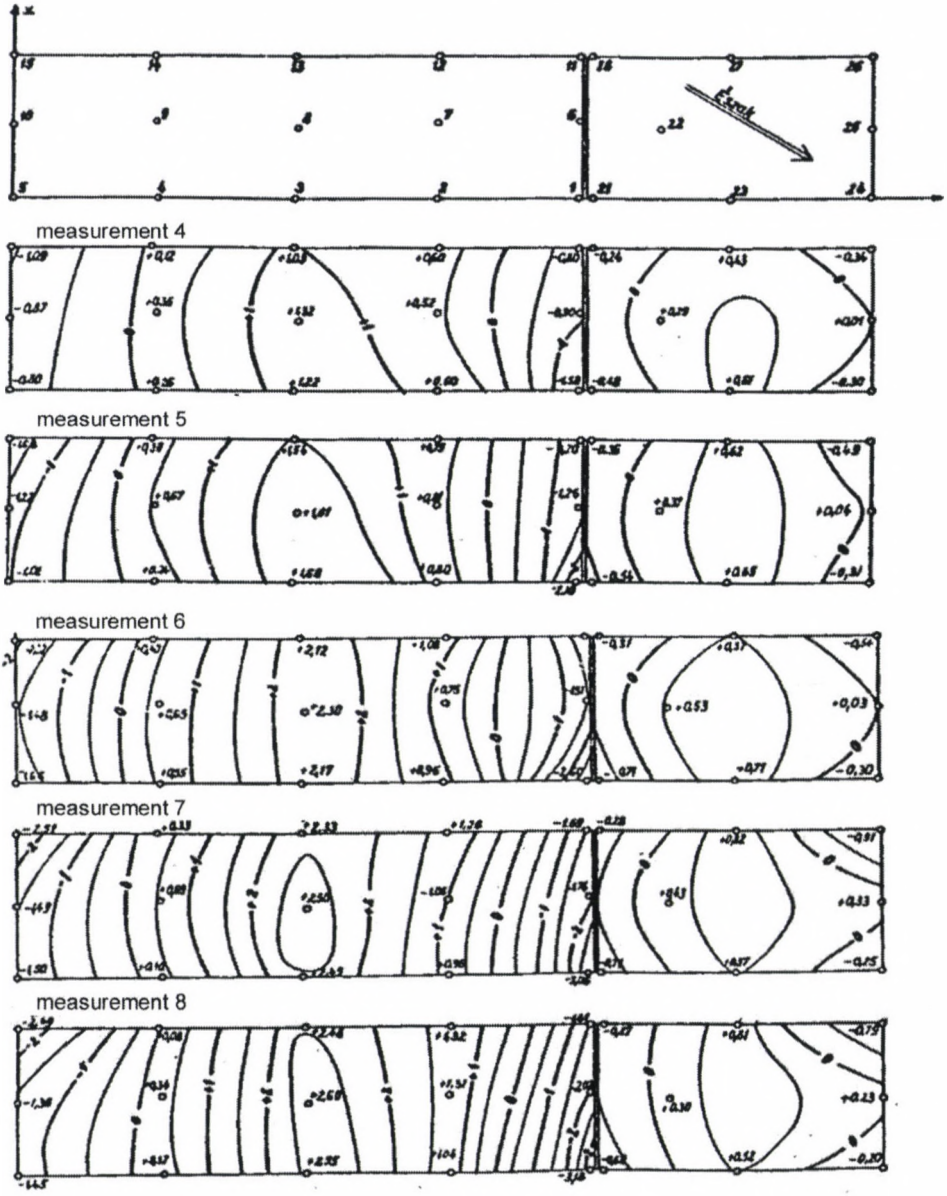


Fig. 7

process the date with directly after the measurement. It means that the movement and the distortion are described in a better, more graphical way.

The residuals after the adjustment may contain systematic errors, which can be determined only by analysing the residuals. We analysed the residuals graphically, but it can be also solved by mathematical analysis. This solution is after used

in linear transformations. After the transformation the residuals contain the local distortion of the points. We correct the residuals with subsequent corrections, which are determined graphically or numerically.

7. Example

Here is the description of movement control of the building No. 1 in Palotaváros, Székesfehérvár.

This building lies on the north of the first section of the housing estate. The surrounding area was deep moorland with inland maters. In the 1960's the most of the water was drained, but some parts of the region have kept their original quality. The upper 3–4 m thick layer of the soil is grey silt, which is unsuitable for building on. At the depth of 5–6 m there lies Pannonian soil, which is more solid than the silt.

The building is 60 m long, 11 m wide and 33 m high. It consists of 3 blocks. The first two and the third blocks are separated with a dilatational interspace. The load of the foundations is 34–45 Mp per meter, depending on the height. The building stands on 97 iron covered foundation column, which are 7.51 m long and 14 m in diameter. The foundation columns are linked with ferroconcrete grid.

The walls inside the ground floor were built later. The ground floor consists of ferroconcrete walls and floor.

We placed the benchmarks in the ground floor. We began the measurements when the ground floor of the third block was finished. The ground floor of the first and the second block and placing of the benchmarks was finished by the third session of measurements. It was the basic measurement of the first and the second blocks, During the measurements the third block was 2–4 levels higher.

In Table I we represent the calculations of the osculating plane in case of the first and the second block. Table I contains the horizontal co-ordinates of the benchmarks and the determined movements. D is the amount of tilt, Dx and Dy are the amounts of tilt in the direction of the x and y axes.

Table II contains the residuals of the osculating plane. On the foot of the table we can see the sum of squares of the residuals (Σpvv), the standard error of unit weight (m_0), the standard error of the tilt (m_D) and the direction of the tilt (m_I). The data represent the increase of the sum of squares and the standard error of the unit weight against the time. The distortion increases systematically.

Figure 6 represents the tilt of the two parts of the building. (The first part contains the first and the second blocks, the second point is the third block.) The tilt of the first and the second part was calculated independently. The two parts tilt towards each other and across, which is caused by the movements of the crane.

Figure 7 represents the isoline diagram at the distortion.

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 239–243 (2000)

DETERMINATION AND MODELLING OF DUMPING SURFACES WITH REAL-TIME-GPS

M SCHÄFER[1]

The deposit of waste on dumps will play a big role at the beginning the next millennium in the disposal concept of domestic refuse and industrial waste. A dumping site can be considered as a body is on which on long-term basis considerable deformations occur, so it is necessary to register and document this. By using Global Positioning System (GPS) in the Real-Time-Kinematic (RTK) mode we have the possibility to obtain data about surface of the dump with high accuracy without influencing the dumping process considerably. These data can then be used for modelling the geometry of the surface build up in post-processing.

Keywords: dumping surface; GPS; volume determination; waste deposits

1. Dump measurement with RTK-GPS

The availability of GPS nowadays provides a measuring system which is characterised by high efficiency, short measuring times on the points, and the advantage of no requirement of direct lines of sight between measurement points. In the Federal Republic of Germany, measurements on dumps are undertaken under recommendations and rules from the Waste Water Technical Union e.V. (ATV 1993). In the use of the GPS for RTK applications some basic requirements must be considered, in addition to the range corrections, carrier phase corrections for all satellites which are offered by the GPS system vendors in own and locally applicable constellations must also be provided in real time via an appropriate communication solution.

The use of RTK-GPS offers new and almost unlimited possibilities for use in the measurement of dumping sites, in general the relatively shadow free locations of the dumping grounds and the independence of the measuring system from weather conditions allow for its extensive use in measurements (Niemeier and Homann 1996).

1.1 Surface and volume determination

By the use of a mobile receiver (Rover) discrete points on the dumping surface could be registered (Fig. 1). As a base station for conveyance of the correct source data, the RTK-capable GPS reference receiver is fitted as far above the dumping surface as possible, in order to guarantee a good inter station communication without interference (Niemeier et al. 1999).

During the measurement of the discrete points no serious Raster is required (Fig. 2). It is however required by the modelling of the geometry as described in the following chapter, to obtain as far as possible a regular point coverage of the

[1]Institut for Geodesy and Photogrammetry, Technical University Braunschweig, Gaußstraße 22, D-38106 Braunschweig, e-mail: ma.schaefer@tu-bs.de

1217-8977/2000/$ 5.00

Fig. 1. RTK measurements on dumping ground (Rover)

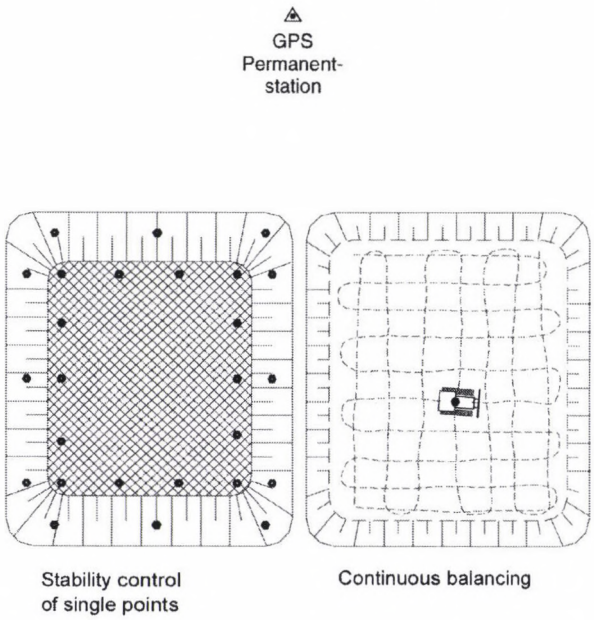


Fig. 2. Means for the surface determination by means of RTK-GPS

dumping ground in order to avoid possible extrapolations and other densification processes.

In addition to undertaking measurements on foot there is the consideration of using vehicle (compactor) movement on a dumping site to obtain measurements useful for modelling the geometry of the dumping site (Fig. 3). An investigation project on this subject, is at moment being undertaken at the Institute for Geodesy and Photogrammetry, TU Braunschweig under the title "The use of GPS receivers on landfill compactors for the measurement of dumping surfaces".

Compared to the use of a Rover on foot, there are some difficulties to be considered in the use of the Rover on a landfill compactor. On the one hand, inclination must be registered and it must be combined with the GPS data in order to correct the erroneous rotational position determination because of the misalignment of the vehicle. For this purpose, an inclination sensor is fitted in addition to the GPS sensor on the vehicle. A problem in the case of the combination of different sensors is the time synchronisation. The connection of both sensors and the time synchronisation can implemented via a control computer (Niemeier et al. 1999).

Parallel to the determination of geometry, it is also possible to determine the volume of the dumping body because of the almost continuous data recording (with 10 Hz), during the compression rides by means of GPS. The determination of volume via the formulation described in the following section can also be made computationally from digital terrain models ground (DGM).

2. Modelling of the dumping surface geometry

Modelling the dumping surface can be realised through the formation of a digital terrain model (DTM). Two different approaches are used to achieve this, one procedure involves the approximation of the ground surface with the aid of surface functions. Whereby the height of a grid point is determined as a distance between the approximating surface polynomial and the height reference surface. Surface functions come as level or sloping planes in the simplest case, up to complex polynomials of higher degree. A further procedure is the triangulation of relief points in the so-called triangular intermeshing which is used in commercial software.

In comparison to a point by point survey, were the discrete points are picked at a defined distance, the points determined by means of landfill compactor lie closely together in some lines while at the same time some of the neighbouring lines are further apart as they are controlled by the motion of the compactor (see Fig. 3).

This fact forms an unfavourable condition for triangular intermeshing from the measured points. Therefore, it appears more favourable to assign points to a firm raster along the trajectories. Consequently, a DTM generation is possible via interpolation procedures from intermeshing the pixels (Fig. 4). Changes on the dumping surface can be determined by different formations of ground descriptions succeeding each other within a given time. The significance of determination of surface areas in part or in total is still to be discussed.

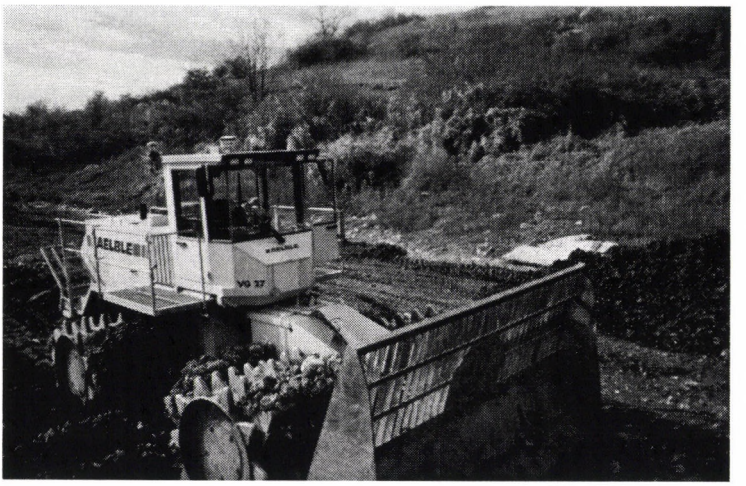


Fig. 3. RTK-GPS on a Landfill compactor

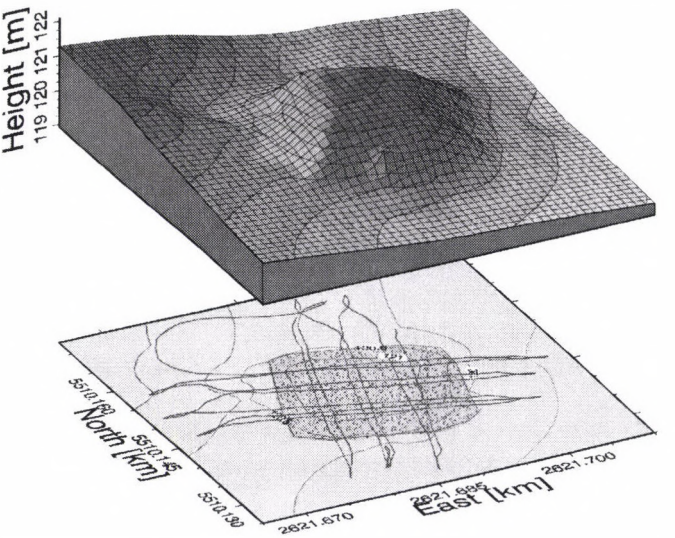


Fig. 4. Modelling of geometry from GPS data

3. Outlook

In conclusion one can say that the use of GPS represents a future pointing technology, with its accuracy potential and the possible applications for the dump measurement. In addition to the geometric information, volume changes and even control information concerning the compression degree, can be derived without additional measurements.

GPS combined with other sensors therefore represents for the future, the basis

of a multi-sensor system for the optimization of the crossings of a landfill compactor i.e. of the modular packaging system and therefore a more efficient utilization of the capacity of the vehicle fleet.

4. References

ATV, Abwassertechnische Vereinigung e.V. 1993: Durchführung und Auswertung von Vermessungsarbeiten auf Deponien im Rahmen der Eigenkontrolle des Deponiebetreibers. In: Regelwerk Abwasser-Abfall, Hinweis H 359.

Niemeier W, Homann Ch 1996: Überwachung und Steuerung auf Deponien. In: Ingenieurvermessung '96, Beiträge zum XII. Internationalen Kurs für Ingenieurvermessung, Bd. 1, Hrsg.: Brandstätter, Brunner, Schelling, Dümmler-Verlag, Bonn

Niemeier W, Schäfer M, Thomsen S, Homann Ch, Schaller M-B 1999: *Z. für Vermessungswesen*, 124, 257–264.

Acta Geod. Geoph. Hung., Vol. 35(2), pp. 245–251 (2000)

INCREASING THE RESOLUTION OF CCD LINE SENSORS BY IMAGE MATCHING

GY MENTES[1]

The CCD image line sensors have very good mechanical and electrical properties to apply them as position sensor of high stability. Their only disadvantage is the low resolution arising from the distance between two adjacent pixels (7–13 μm) determined by the manufacturing technology. Because of it, they cannot be used directly for measurements of very small linear and angular displacements needed in geodynamics and engineering geodesy. In the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences a CCD displacement transducer of subpixel accuracy was developed. A microprocessor is built into the new transducer which determines the displacement by fitting the dark-bright transfer functions of the CCD sensor at different positions by means of the least squares method. The accuracy of the displacement transducer is ± 0.02 pixel.

Keywords: CCD sensor; displacement transducer; engineering geodesy; geodynamics; greyvalue; image processing; microprocessor

1. Introduction

The requirements against the accuracy of measurements in geodynamical research and engineering geodesy are steadily increasing. The conventional displacement and angle transducers e.g. capacitive, inductive transducers (LVDT, LVRT) most frequently used in the deformation measuring instruments have a lot of disadvantages when applying them in modern measuring systems. These transducers provide an analogous output voltage at a level of some millivolts and are sensitive to temperature variations, disturbing electromagnetic fields and especially the capacitive transducers are sensitive to humidity variations. This means that the analogous output signal of these transducers considerably varies due to the variation of the environmental parameters. This effect is called "drift" of the transducer. The instrumental drift is very disturbing in case of long-term deformation measurements needed for monitoring of the movements of tectonic plates in geodynamics or for controlling industrial establishments e.g. dams, cooling towers, hydro and atomic power stations, etc. in engineering geodesy.

Today, the development and application of modern transducers play an important role in achieving the required high accuracy and stability of deformation measuring instruments. A possible solution for elimination of the above mentioned error sources of the conventional transducers is the application of CCD sensors in displacement transducers. The CCD image line sensors have a great advantage, namely their basic material is silicon, which is mechanically very stable and has a low coefficient of thermal expansion. Therefore the CCD image line sensors represent a very stable and accurate reference scale. This feature is very important

[1]Geodetic and Geophysical Research Institute of the Hung. Acad. Sci., H–9401 Sopron, POB 5, Hungary

1217-8977/2000/$ 5.00

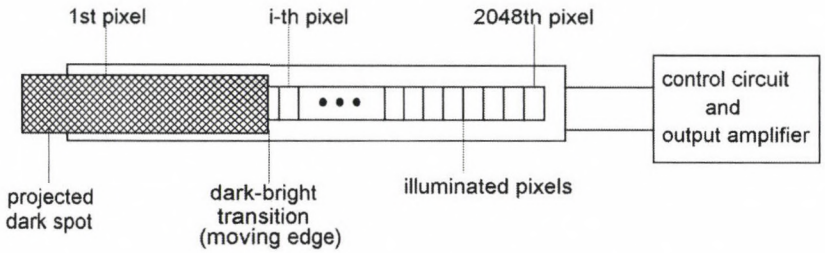


Fig. 1. The principle of the displacement transducer built with CCD image line sensor

especially in engineering geodesy because the measurements in industrial establishments have to be carried out in a wide temperature range. On the other hand the CCD image line sensors have a relative large distance between two adjacent pixels and as a result of this they have a very limited resolution as compared to the commercial inductive and capacitive transducers. To overcome this problem the properties of the CCD image line sensors were investigated in the frame of a co-operation between the Institute of Engineering Geodesy of the TU Wien and the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences (Kahmen and Mentes 1993a, 1993b). On the base of this research a new method for increasing the resolution of the CCD sensors was developed in the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences.

2. The principle of CCD displacement transducers

A CCD image line sensor consists of a row of image sensor elements (256–4096) separated from each other. The photon generated electrons are accumulated in the sensor elements. The amount of charge accumulated in each element (pixel) is a linear function of the incident illumination intensity and the integration period. The output signal will vary in an analogous manner from a thermally generated background level at zero illumination to a maximum at saturation under bright illumination. At position sensors, the displacement of a moving edge is proportional to the displacement to be measured. The moving edge divides the row of the sensor elements into dark and bright areas as it is shown in Fig. 1. Figure 2 shows the output signal of the CCD sensor at two edge positions. At the pixels under the moving edge the output signal of the CCD sensor changes very abruptly. If we can determine the serial number of the pixel where the dark-bright transition takes place, we can determine the position of the moving edge. The difference of the two serial numbers of the pixels at the two edge positions gives the displacement of the moving edge as a number of pixels. If we know the distance between two adjacent pixels we can determine the displacement in micrometers.

A simple method to determine the position of the moving edge is the following: a control circuit shifts the voltage of each pixel serially to the output of the CCD image line sensor. A voltage comparator compares the analogous output voltage of the CCD sensor with a constant threshold voltage whose value is the mean value of the minimum and maximum output voltages of the CCD sensor (Fig. 3). The

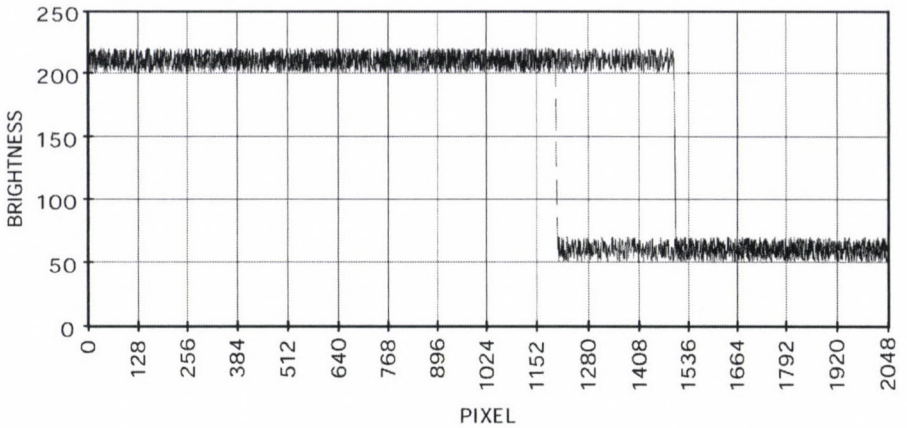


Fig. 2. The output signal of the CCD sensor at two edge positions

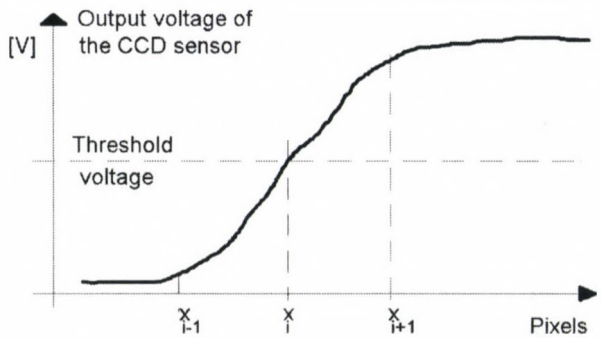


Fig. 3. The principle of the direct determination of the position of the moving edge

position of the moving edge is given by the serial number of the pixel at which the output voltage is equal or greater than the threshold voltage. The block diagram of the transducer is shown in Fig. 4. A counter counts the impulses shifting the pixel voltages to the output. If the pixel voltage is equal or greater than the threshold voltage the comparator loads the counter value to a storage register. The contents of the register gives the position of the moving edge and it can be displayed or transferred to a computer or a data logger.

The accuracy achieved by this method is ± 1 pixel in the best case. The accuracy very strongly depends on the temperature stability of the threshold voltage. This transducer was built with the CCD image line sensor "Fairchild CCD 143". The distance between two adjacent sensor elements is 13 μm at this type. It means that the accuracy is worse than ± 13 μm. In most cases, especially in geodynamics this accuracy is not sufficient. Therefore, in the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences a new displacement transducer was developed. A microprocessor is built into the transducer which determines the displacement of the moving edge with a subpixel accuracy.

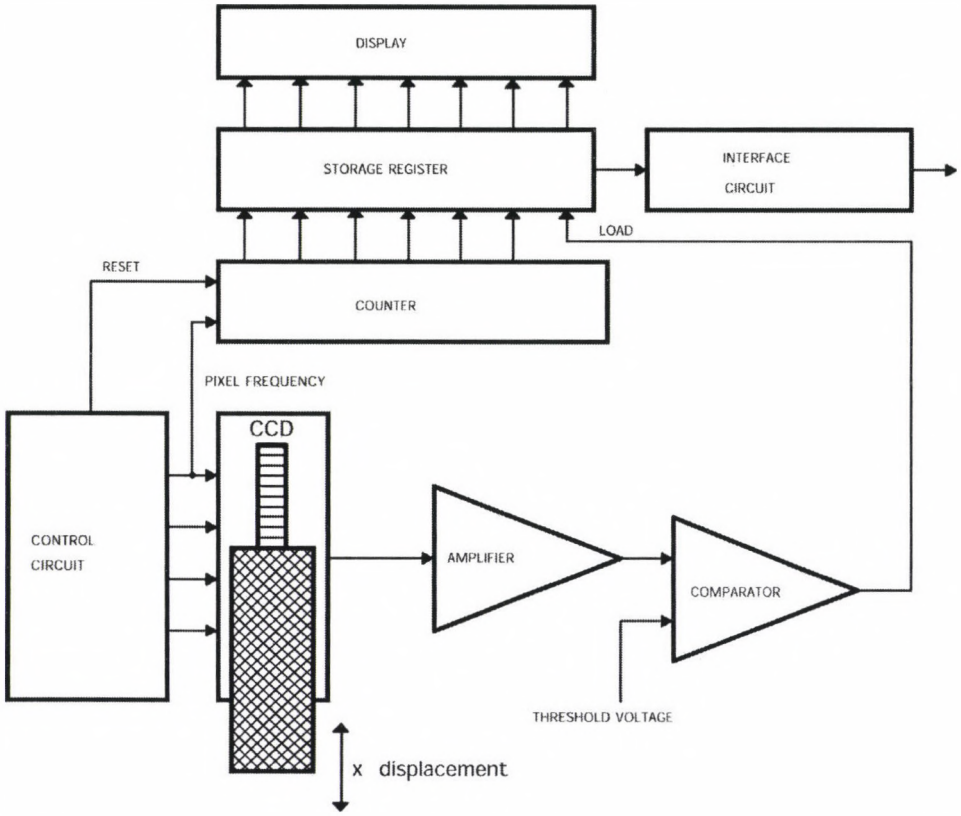


Fig. 4. The block diagram of the CCD displacement transducer with direct determination of the position of the moving edge

3. CCD displacement transducer with increased resolution

The distance between two dark-bright transition curves $f(x)$ and $g(x + \Delta x)$ gives in this transducer the displacement of the moving edge (Fig. 5). The output signal of the CCD sensor is digitized and the built-in microprocessor determines the distance between the two curves fitting them by the least squares method. The shift Δx giving the best fitting is the displacement of the moving edge and is determined from two values k and λ.

The rough distance between the two curves is given by the integer number of pixels k lying between the points of inflexion of the two curves. The dark-bright transition at the moving edge is enacted very abruptly within 4 pixels as it is proved by measurements. Figure 6 shows a bright-dark transition enlarged. The curve has a small noise at the dark and bright pixels caused by the somewhat different sensitivities of the individual pixels and by the errors of the A/D converter. These features of the transition curve were used to simplify the determination of the places of the inflexion points.

The digitized dark-bright transition curve is given in the 1st position of the

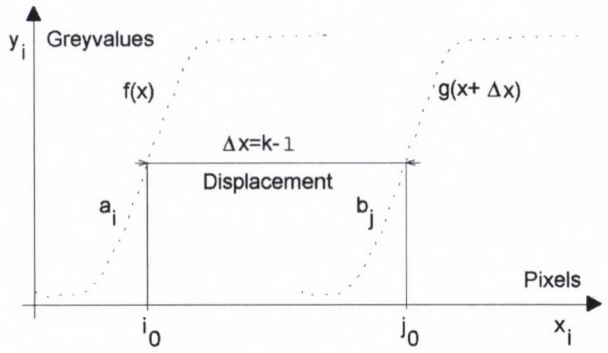


Fig. 5. The principle of the displacement of the moving edge by fitting the dark-bright functions of the CCD sensor at two moving edge positions

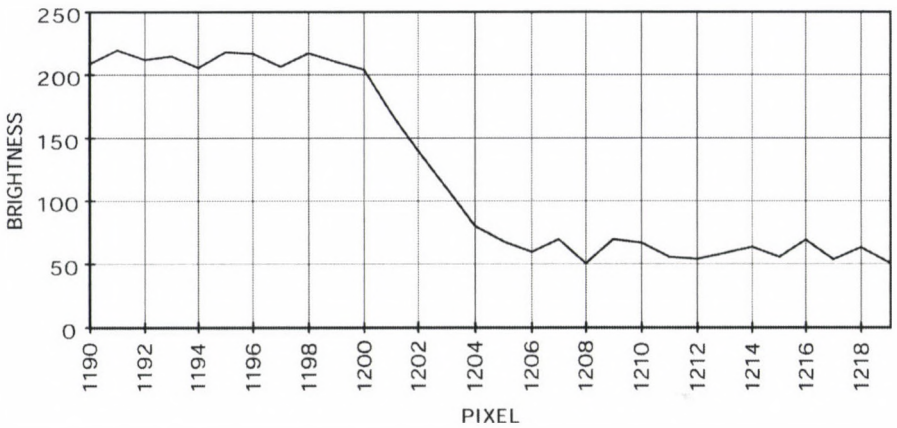


Fig. 6. The dark-bright transition of a CCD image line sensor at the moving edge

moving edge by the digitized values $\{a_i\}$ and in the 2nd position $\{b_j\}$, where $i = 1 \ldots N$ and $j = 1 \ldots N$. N is the number of the pixels of the CCD image line sensor, which is 2048 at the sensor used by us.

For the determination of the places of the inflexion points two moving sums of five terms are calculated for both transition curves:

$$c_i = a_i + a_{i+1} + a_{i+2} + a_{i+3} + a_{i+4} \qquad i = 1 \ldots N - 12$$

$$c_{i+8} = a_{i+8} + a_{i+9} + a_{i+10} + a_{i+11} + a_{i+12} \qquad i = 1 \ldots N - 12$$

$$d_j = b_j + b_{j+1} + b_{j+2} + b_{j+3} + b_{j+4} \qquad j = 1 \ldots N - 12$$

$$d_{j+8} = b_{j+8} + b_{j+9} + b_{j+10} + b_{j+11} + b_{j+12} \qquad j = 1 \ldots N - 12$$

$$i_0 = i + 6, \text{ when } |c_i - c_{i+8}| \text{ is maximum} \qquad i = 1 \ldots N - 12$$

and

$$j_0 = j + 6, \text{ when } |d_j = d_{j+8}| \text{ is maximum.} \qquad j = 1 \ldots N - 12.$$

The rough distance between the curves in pixel:

$$k = j_0 - i_0 \,.$$

The fractional value of the distance within the pixel λ is searched in the form of $k - \lambda$, where $0 \leq \lambda \leq 1$. To determine λ the first dark-bright curve must be shifted to the second by $k - \lambda$. The best fitting is obtained if the square sum of the deviations

$$\sum_i \left[a_{i+k} - (b_i + \lambda\,(b_{i+1} - b_i))\right]^2 \qquad i = 1 \ldots N - 12$$

is minimum. From this condition

$$\lambda = \frac{\sum\limits_i (b_i - b_{i+1})\,(b_i - a_{i+k})}{\sum\limits_i (b_i - b_{i+1})^2} \qquad i = i_0 - 3,\, i_0 - 2, \ldots i_0 + 3$$

can be calculated and the displacement of the moving edge: $\Delta x = k - \lambda$.

If $|\lambda| > 1$ then k must be incremented or decremented by one and the calculation of λ must be repeated till $|\lambda| \leq 1$.

Figure 7 shows the block diagram of a displacement transducer constructed according to this idea. The greyvalues of the 2048 pixels of the CCD image line sensor are digitized during the readout and stored in a separated image memory with a capacity of 2048 byte. The pixel frequency is high (100 kHz–1 MHz) and therefore the A/D converter is controlled not by the built-in microprocessor but by a separated control circuit. A counter counts the pixels readout from the CCD sensor and addresses the image memory during the readout. Between two readouts, during the integration time of the CCD image line sensor the built-in microprocessor transfers the greyvalues from the image memory to its own, makes the image matching and displays or transmits the calculated displacements.

The characteristics of the transducer was tested by means of a laser interferometer. The linearity errors of the transducer are less than 0.27 μm in the whole measuring range (25.4 mm). This value corresponds to about 0.02 pixel in the case of the applied CCD image line sensor. According to our investigations the accuracy of the transducer does not depend significantly on the variations of the illumination because the bright-dark transition function (Fig. 6) is very steep and it is filtered during the calculations.

4. Conclusion

Due to their high mechanical and electrical stability the CCD image line sensors with increased resolution can be widely used for displacement measurements in the engineering geodesy and geodynamics. There are two possibilities to increase the resolution power of the displacement transducers built by CCD image line sensors. In the first case the displacement of the moving edge can be enlarged to a multiple

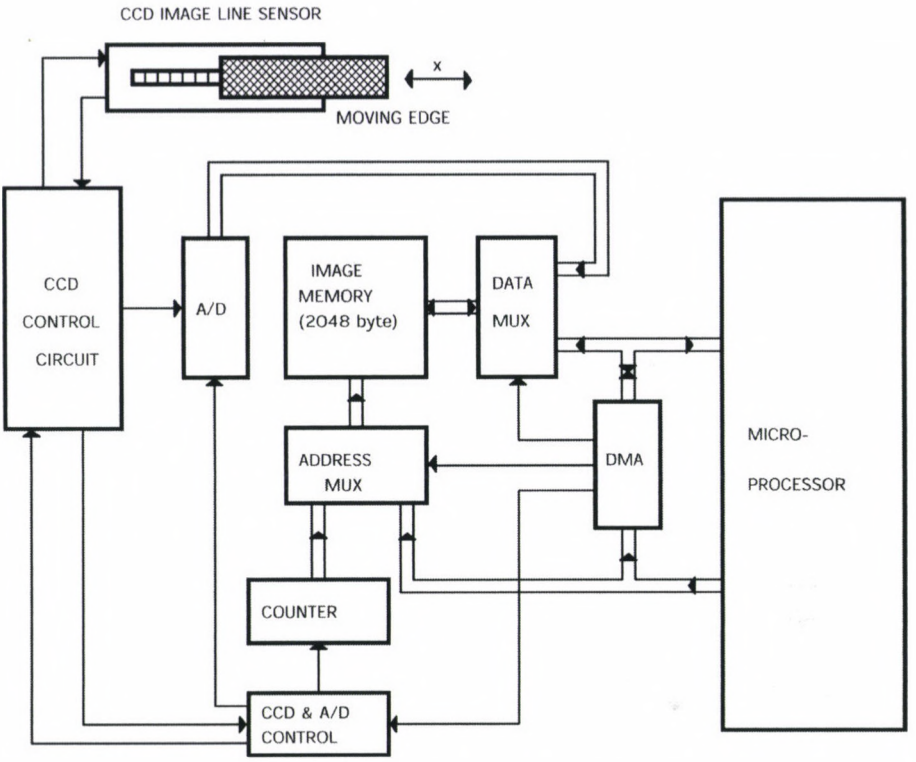


Fig. 7. The block diagram of the CCD displacement transducer using microprocessor for the determination of the position of the moving edge with an increased accuracy

of the displacement to be measured by means of a mechanical or optical transmission. This solution causes additional errors in the transducer. In the second case the position of the moving edge is determined with a subpixel accuracy by means of a built in microprocessor as it is described above. The two methods can be combined in order to achieve a higher resolution power especially for high accuracy geodynamical measurements.

Acknowledgement

The presented research work was financially supported by the National Scientific Research Foundation (OTKA) in the frame of the project T 019030.

References

Kahmen H, Mentes Gy 1993a: In: Optical 3D Measurement Techniques II, Gruen and Kahmen (eds), Herbert Wichmann Verlag GmbH, Karlsruhe, 179–185.

Kahmen H, Mentes Gy 1993b: In: Optical 3D Measurement Tecniques II, Gruen and Kahmen (eds), Herbert Wichmann Verlag GmbH, Karlsruhe, 315–321.

Acta Geod. Geoph. Hung., Vol. 35(2), p. 253 (2000)

BOOK REVIEW

A CIVIDINI ed.: *Application of Numerical Methods to Geotechnical Problems.* Proceedings of the Fourth European Conference on Numerical Methods in Geotechnical Engineering – NUMGE98, Udine, Italy, 1998, CISM Courses and Lectures, No. 397, Springer-Verlag, Wien New York, 1998, 765 pp., 135 USD

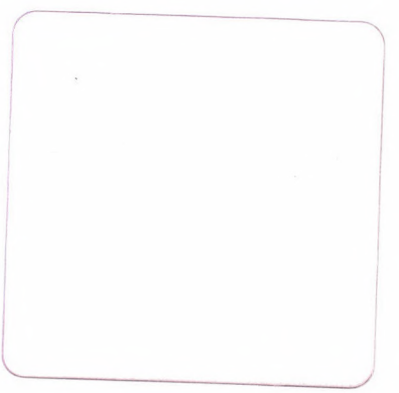
The NUMGE98 Conference brought together senior and young researchers, scientists and practicing engineers from European and overseas countries, to share their knowledge and experience on the various aspects of the analysis of Geotechnical Problems through Numerical Methods. The papers address a broad spectrum of geotechnical problemes, including tunnels and undergound openings, shallow and deep foundations, slope stability, seepage and consolidation, partially saturated soils, geothermal effects, constitutive modelling, etc. Most of the problems are solved with the Finite Element Method, but there are presented also some other methods as e.g. Artificial Neural Network Analysis or Fractal Analysis to highlight geotechnical problems. The results calculated from Numerical Modelling are compared with the results of Field Measurements and of Small Scale Tests in the first chapter. In the second chapter the results obtained from the Numerical Modelling are verified by Laboratory Tests and Constitutive Modellings. In the third one the Theoretical Aspects of Numerical Algorithms are studied.

Main Sessions: Invited Lectures; Interpretation of Field Measurements and of Small Scale Tests; Laboratory Test and Constitutive Modelling; Numerical Algorithms and Theoretical Aspects; Miscellaneous; Late Papers.

This book is very useful for both computer experts and engineers who are dealing with the theory of the modelling of different geotechnical problems and with their practical use, respectively.

S Szalai, J Benedek

Akadémiai Kiadó, Budapest



INSTRUCTIONS TO AUTHORS

Manuscripts should be sent to the Editorial Office (address see above). Articles not published or submitted for publication elsewhere are only accepted.

Manuscripts should be preferably on disc (LATEX *.tex files or Word for Windows documents or text-only ASCII files) with a written copy, or via e-mail (actagg@ggki.hu). Files will be accepted, after previous agreement, via *anonymous ftp*, too. Typewritten manuscripts (in duplicate) are also accepted.

Manuscripts should include the following components which should be presented in the order listed (tables and illustrations should be separated in case of manuscripts on disc, too).

1. Title, name(s) of the author(s), affiliation, dateline, abstract, keywords
2. Text, acknowledgements
3. References
4. Footnotes
5. Legends
6. Tables
7. Illustrations

1. The *affiliation* should be as concise as possible and should include the complete mailing address of the authors. The *date of receipt* will be supplied by the editors. The *abstract* should not exceed 250 words and should clearly and simply summarize the most important methods and results. 5-10 significant expressions describing the content are used as *keywords*. Authors may recommend these keywords.

2. The *text* should be in English and as short and clear as possible. In case of typewritten manuscripts, please note the following:

- avoid possible confusion between o, O (letters) and 0 (zero), I (letter) and 1 (one), ν (Greek nu) and u, v (letters), etc.
- explain ambiguous and uncommon symbols by making marginal notes in pencil
- formulas are to be numbered consecutively with the number in parentheses to the right of the formula.
  Text references to the equations may then be made by the number in parenthesis. The word equation in this context is to be abbreviated to Eq. and Eqs in the plural
- the International System of Units (SI) should be used.

3. *References* are accepted only in the Harvard system. Citations in the text should be as:

... (Bomford 1971) ... or Bomford (1971) ...
... (Brosche and Sündermann 1976) ...
... (Gibbs et al. 1976b) ...

The list of references should contain names and initials of all authors (the abbreviation et al. is not accepted here); for journal articles, year of publication, title of the journal in abbreviated form, volume number, first and last page. For books or chapters in books, the title is followed by the publisher and place of publication.

All items must appear both in the text and in the list of references.

4. *Footnotes* should be typed on separate sheets.

5. *Legends* should be short and clear. The place of tables and figures should be indicated in the text on the margin.

6. *Tables* should be numbered serially with Roman numerals. Vertical lines are not used.

7. All the *illustrations* should contain the figure number and author's name in pencil on the reverse. The most important point with figures is clearness. Photographs and half-tone illustrations should be sharp and well contrasted. Colour photographs will be accepted, but the extra cost of reproduction in colour must be borne by the authors. Encapsulated postscript (EPS) files will be accepted on disk or via e-mail, too.

Only original papers will be published and a copy of the Publishing Agreement will be sent to the authors of papers accepted for publication. Manuscripts will be processed only after receiving the signed copy of the agreement. Information is sent to the first author if no other wish is expressed.

Acta Geodaetica et Geophysica Hungarica

VOLUME 35, NUMBER 3, 2000

Akadémiai Kiadó, Budapest

AGG 35 (3) 255–381 (2000) HU ISSN 1217-8977

ACTA GEODAETICA ET GEOPHYSICA HUNGARICA

A Quarterly Journal of the Hungarian Academy of Sciences

Acta Geodaetica et Geophysica Hungarica (AGG) publishes original reports and reviews on geodesy and geophysics in English.

AGG is published in yearly volumes of four issues by

AKADÉMIAI KIADÓ
Prielle K. u. 4, H-1117 Budapest, Hungary
http://www.akkrt.hu

Manuscripts and editorial correspondence should be addressed to

AGG Editorial Office
Geodetical and Geophysical Research Institute
P.O. Box 5, H-9401 Sopron, Hungary
Phone: (36-99) 508-340
Fax: (36-99) 508-355
E-mail: actagg@ggki.hu
http://www.ggki.hu/agg/

Subscription information

Orders should be addressed to

AKADÉMIAI KIADÓ
P.O. Box 245, H-1519 Budapest, Hungary
Fax: (36-1) 464-8221
E-mail: kiss.s@akkrt.hu

Subscription price for Volume 35 (2000) in 4 issues US$ 180.00, including normal postage, airmail delivery US$ 20.00.

Printed in Hungary
PXP Ltd., Budapest

AGeod 35 (2000) 3

CONTENTS

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 255–264 (2000)

SOME MAGNETIC PROPERTIES OF SEDIMENTS FROM LAKE FERTŐ (NEUSIEDLERSEE) REGION, AUSTRIA-HUNGARY

A JELINOWSKA[1], C CARVALLO[1], V WESZTERGOM[2], P TUCHOLKA[1], M MENVIELLE[3], L SZARKA[2], F KOHLBECK[4], J J SCHOTT[5]

[Manuscript received January 29, 1999]

This paper presents the results of a magnetic susceptibility and thermomagnetic behaviour study from 6 sediment cores from the Lake Fertő area (Hungary/Austria boundary). These data indicate a very low content of magnetic particles in the sediment, which represents late Holocene. In studied bulk sediment samples we observed four characteristic types of thermomagnetic behaviour. Some of these behaviours indicate chemical transformations of non magnetic iron minerals: siderite and iron and sulphur bearing mineral into magnetite and magnetic iron monosulphide respectively during heating. The observed minerals are independent of lithology, which suggests their post depositional origin related to the anoxic early diagenetic conditions in the sediments. The presence of siderite and iron and sulphur bearing minerals in the sediment can reflect variation in the salinity of waters in the lake and thus can give information about the development of the basin in terms of the hydrological system (relation between evaporation and alimentation) of the lake.

Keywords: core sample; Lake Fertő; magnetic properties; palaeoclimate

Introduction

The presence of magnetic minerals, their concentration and size in lake sediments depend on the circumstances surrounding the deposition of this sediment and post-depositional processes. The allogenic fraction is supplied from the catchment area by waters and wind and depends on the detrital input to the lake, which carries varying quantities of magnetic minerals (e.g. iron oxides: magnetite Fe_3O_4, maghemite γFe_2O_3, hematite αFe_2O_3, and iron hydroxides: goethite FeOOH). It also carries paramagnetic (e.g. iron-bearing clay) and diamagnetic (e.g. Ca/Mg carbonates, quartz, organic matter ...) minerals and materials. The post-depositional processes include the dissolution of detrital iron minerals, the authigenic production of iron sulphides (e.g. greigite Fe_3S_4, pyrite FeS_2), oxides (e.g. magnetite), carbonates (e.g. siderite $FeCO_3$) and phosphates (e.g. vivianite $Fe_3(PO_4)_2 8H_2O$) minerals, and biogenic mineralisation (e.g. magnetite, greigite). These processes can

[1]Laboratoire de Géophysique, Université Paris Sud, Bat. 504, F–91405 Orsay Cedex, France
[2]Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H–4901 Sopron, POB 5, Hungary
[3]Centre d'Etude des Environments, Terrestre et Planetaires, F–94107 Saint Maur des Fosses, 4 Avenue de Neptune, France
[4]Technische Universität Wien, Abt. Geophysik, A–1040 Wien, Gusshausstrasse 27–29, Austria
[5]Ecole et Observatoire de Physique du Globe de Strasbourg, France

1217-8977/2000/$ 5.00

alter the mineral magnetic assemblage of detrital origin in the sediments (Berner 1980, Berner 1981, Curtis 1987, Mann et al. 1990). Thus investigations of magnetic minerals, their concentration and size are extremely useful in studies of lake sediments: their fluctuations depend on the environmental and climatic changes in the cachment area of the lake.

The composition, grain size and concentration of magnetic minerals in the sediment can be detected using different magnetic parameters of the bulk sediment (Thompson and Oldfield 1986, Verosub and Roberts 1995). Several parameters (magnetic susceptibility, various remanent magnetisations, magnetic hysteresis parameters) depend mainly on the concentration and grain size of ferromagnetic (*sensu lato*) minerals. Low field magnetic susceptibility depends also on the concentration of paramagnetic and diamagnetic materials. Analysis of the thermomagnetic behaviour with determination of the Curie temperature of magnetic minerals allows their identification (Curie temperature is a critical temperature below which ferromagnetic materials carry a remanent magnetisation; above this temperature they are paramagnetic; Curie temperature is unique for each specific magnetic material). It also gives information about the possible transformations with respect to temperature of the various iron-bearing materials. Combined interpretation of all these parameters provides valid information about the magnetic fraction, even when the concentration of magnetic material in the sediment is very small, to the point that it is undetectable by other methods such as X-ray diffraction (XRD).

In 1997, in the Lake Fertő region, physical properties of lake sediments measured from the earth's surface and directly from sediment cores were investigated by a Hungarian-Austrian-French geophysical team. Under this project 6 sediment cores were collected and geoelectrical soundings were carried out from the surface around the drill holes. Electric properties of sediments deduced from soundings were compared to those measured directly on cores (Kohlbeck et al. 1998, 2000). In this paper we present some magnetic parameters studied in sediment cores. These parameters provide information on changes in the lake system during the late Quaternary period covered by this sediment sequence.

Lake Fertő (Neusiedlersee) (309 km^2) is located on the Austro-Hungarian border south (20 km) of the Danube. It is a shallow (average water depth 50–60 cm) water body with large water level fluctuations over the historical period. More than one-third of its surface is overgrown with potuberant ulignant plants. Lake Fertő is the terminal lake of the Vulka and Rákos rivers and without an outlet in its natural condition. Water can be discharged through the Hanság main-channel. The lake is supplied by local rainfall (78%) and the Vulka and Rákos rivers (20%); groundwater feed represents only 2%. The age of the lake is estimated to be nearly 20 kyrs. Its origin is related to the simultaneous formation of two basins (Hanság and Fertő). The gravel barrier separating these basins was formed by large depositions from the Danube and Rába rivers at the end of Pleistocene. The two basins were frequently connected during high water periods.

Core collection and sediment description

Six cores were taken from the southern part of the Fertő-Hanság area using a Wright piston corer. The length of the cores varies from 100 cm to 350 cm. Hard sediments or liquified sands prevented collection of longer cores using this method. One core was taken from Lake Fertő at Fertőrákos, and the following five, from the dried out part of the area eastward, in order to obtain sediments from the Fertő and Hanság basins (Fig. 1). The collected sediment presents the youngest part of the Quaternary, most probably late Holocene. The Lake Fertő sediment in the core Fertőrákos (Fig. 2) consists of brown, downwards grey clay, overlying grey, fine grained sand. Below the sand, sandy clay and clay occur, with oxidation traces in its upper part, and grey colour downwards. Sediments from five other cores are characterised by similar lithological units (clay, sand and intermediate phases) of brown, grey or yellow (rarely) colour and of different quantities in each core. There are also gravel levels. As these sediments were taken from the dried out part of the Fertő-Hanság area, their upper parts consist of the organic matter accumulation due to the soil development. Variations of the lithology observed in studied cores show the stages of the evolution of each studied part of the basin: 1. clay sediments indicate the presence of the water body with low transport energy, most probably in lacustrine conductions, 2. sand and gravel suggest high transport energy and are attributed to the fluvial deposits, 3. soil development indicates dry episodes.

Magnetic mineral sampling and measurements

Standard sampling methods were employed to obtain semi-continuous coverage of each core. The cores are preserved in plastic tubes which were sliced into two parts (one half kept as an archive and the second used for sampling). We pushed standard $22 \times 22 \times 22$ mm perspex cubes into the sediment, after cleaning the surface of a half-core. 260 samples were obtained from all studied cores.

Low field susceptibility ($\chi = M/H$, where M is the magnetisation induced in a material by an applied field H, of a strength less than 1 mT) was measured with a Bartington Instruments MS-2 susceptibility bridge. The thermomagnetic behaviour, with the Curie point of the magnetic components of the sediment was determined on a horizontal force translation balance (Curie balance) in air atmosphere. Mineralogy of sediments was determined by X-ray diffraction (XRD) and microprobe.

Results

Magnetic low field susceptibility χ (Fig. 2) presents low values in all measured samples. This is due to the very low magnetic particle content in Lake Fertő sediments, to the point that their presence is not reflected by χ, which is also sensitive tot he para- and diamagnetic particles. The most often obtained low and positive values of χ indicate the dominance of paramagnetic materials in sediments (such as iron-bearing clay minerals). Low and negative values of χ occasionally obtained

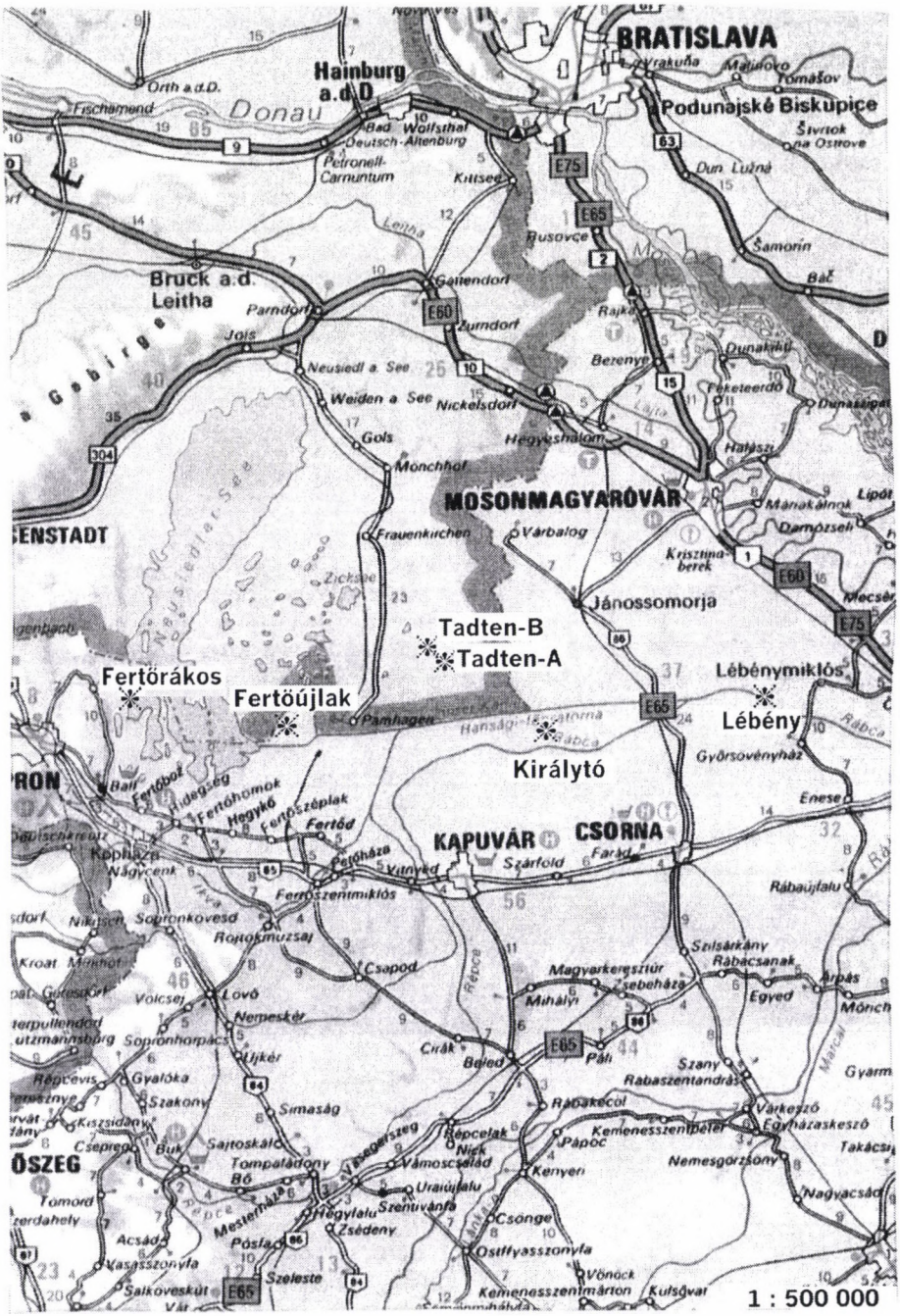


Fig. 1. Map showing the area of investigation and locations of coring sites

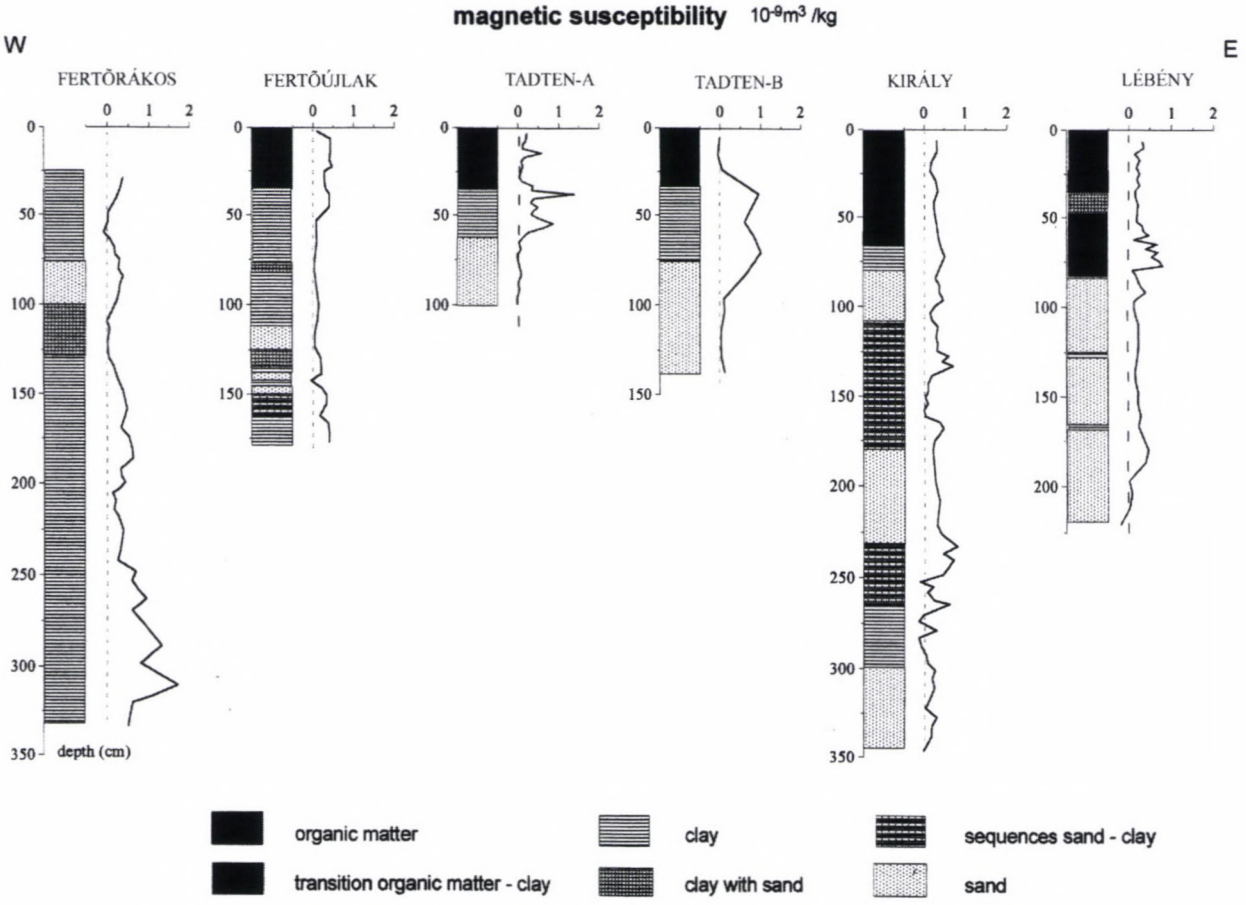


Fig. 2. Down-core variation in lithofacies and magnetic susceptibility in studied sedimentary cores

are due to the presence of diamagnetic materials. The sediment diamagnetic properties are usually accounted for by a high quartz, carbonate and/or organic matter content. In our study, the dia- and paramagnetic properties are not clearly related to one type of lithology, such as: clay — para, and sand and/or organic matter — diamagnetic properties. It means that different components are present in all lithological units. Therefore it is impossible to distinguish the periods of the lake evolution on the basis of magnetic susceptibility.

As χ does not show variations in magnetic fraction, we have chosen samples for analyses of the thermomagnetic behaviour on the basis of lithological changes for each core. In studied bulk sediment samples, we observed four characteristic types of behaviour during Curie balance experiments, when heated in air atmosphere.

1. The first thermomagnetic behaviour shows the stepwise decrease in magnetisation in the sample on heating to 650°C, and stepwise remagnetisation during cooling to room temperature, without a clearly defined Curie temperature of any mineral (Fig. 3a). This behaviour indicates that paramagnetic minerals dominate in the sample, and these minerals, as the process is reversible, do not undergo any chemical transformation. This behaviour was observed in sand (Tadten-A 70 cm depth), clay (Király 262 cm; Fertőrákos 139 cm, 199 cm depth), sequence sand-clay (Fertőrákos 113 cm depth) and in organic matter (Király 36 cm depth).

2. The second thermomagnetic behaviour shows the stepwise decrease in magnetisation during heating, as in the previous case, but during cooling we observe an increase in the magnetisation from 580°C (Fig. 3b), which is the Curie temperature of magnetite. When the sample is heated once more, we observe the behaviour of magnetite. This indicates that the paramagnetic material present in the analysed samples undergoes chemical changes during the experiment (at high temperature), which favour the creation of magnetite. This behaviour was observed in organic matter (Tadten-A 13 cm, 23 cm depth) and in the sequence of the transition organic matter-clay (Lébény 32 cm depth).

3. The next thermomagnetic behaviour is characteristic of paramagnetic materials as previously until about 450°C during heating. At about 450°C, the magnetisation increases with the maximum at 500°C and decreases to 0 at 580°C (Fig. 3c). It suggests that the paramagnetic material present in the sample was transformed into magnetite during heating. In order to confirm the formation of magnetite, we heated another fresh sample to 500°C and after cooling we repeated the Curie balance experiment on heating it to 650°C. This experiment shows the thermomagnetic behaviour expected for magnetite (Fig. 3d). The observed transformation of the paramagnetic material to the magnetite is similar to that known for siderite if heated to 500°C (Hus 1990). In order to verify this hypothesis we made XRD analyses of the bulk sediment material (i) before and (ii) after heating to 500°C. These experiments show clearly the presence of the siderite in the unheated sample and this siderite disappears in the heated one. We also made XRD analysis of the magnetic

extract obtained from the bulk sediment sample heated to 500°C; the presence of the magnetite is difficult to identify probably because it is poorly crystallised. This thermomagnetic behaviour, which permits identification of siderite in sediment samples, was observed in sand (Tadten-A 96.5 cm; Király 346 cm; Lébény 150 cm, 176 cm; Fertőrákos 92 cm depth), clay (Tadten (a) 44 cm, 53 cm; Kiraly 236 cm; Fertőrákos 35 cm, 68 cm, 263 cm, 310 cm depth) and organic matter (Lébény 72 cm depth).

4. The last thermomagnetic behaviour is characteristic of paramagnetic minerals until about 300°C. At about 300°C, an increase in magnetisation is seen (Fig. 3e), followed by a decrease at 400°C and a second decrease to zero at 580°C. During cooling the magnetisation increases from 580°C. If we repeat the experiment on this material, we observe clearly the characteristic behaviour of magnetite. Therefore we observed the transformation of the initially paramagnetic material to ferromagnetic material at about 300°C, which afterwards is transformed to magnetite. In order to identify the magnetic phase created at 300°C, we heated a sample to 300°C. The created mineral was subjected to a Curie balance experiment. This experiment indicates the presence of a mineral that undergoes a major decrease in magnetisation between 300 and 400°C and afterwards is transformed to magnetite. Such behaviour is similar to that observed for iron monosulphides: (i) greigite (Tric et al. 1991, Snowball 1991, Hoffman 1992, Roberts and Turner 1993, Reynolds et al. 1994, Jelinowska et al. 1995) and (ii) smythite (Fe_9S_{11}) (Krs et al. 1992, Hoffman 1993). Analyses with microprobe of a magnetic extract obtained from the sample heated until 300°C shows iron and sulphur ions in a ratio which confirms the presence of the iron monosulphide. Identification of this iron monosulphide by XRD was not possible most probably because, as in the previous case for magnetite, of the non crystalline form of the created mineral. This experiment shows the transformation of the initially paramagnetic material into magnetic iron monosulphide with temperature (about 300°C), which is unstable and transforms afterwards into magnetite. At the present state of the study we can not identify this non-magnetic material which contains iron and sulphur ions. The thermomagnetic behaviour described here was observed in clay sediments (Tadten-A 36.5 cm depth), sand (Király 103 cm depth) and organic matter (Tadten-A 29.5 cm, Lébény 8 cm).

Discussion and conclusions

Sediment from Lake Fertő are dominated by a paramagnetic fraction. With Curie balance experiments we identified 4 different types of thermomagnetic behaviour: (i) the paramagnetic material remains stable during heating and cooling; (ii) the paramagnetic material is transformed into magnetite at high temperature; (iii) the paramagnetic material undergoes transformation into magnetite during heating (at 500°C) and is identified as siderite; (iv) the paramagnetic material undergoes transformation into magnetic iron monosulphide during heating (at 300°C)

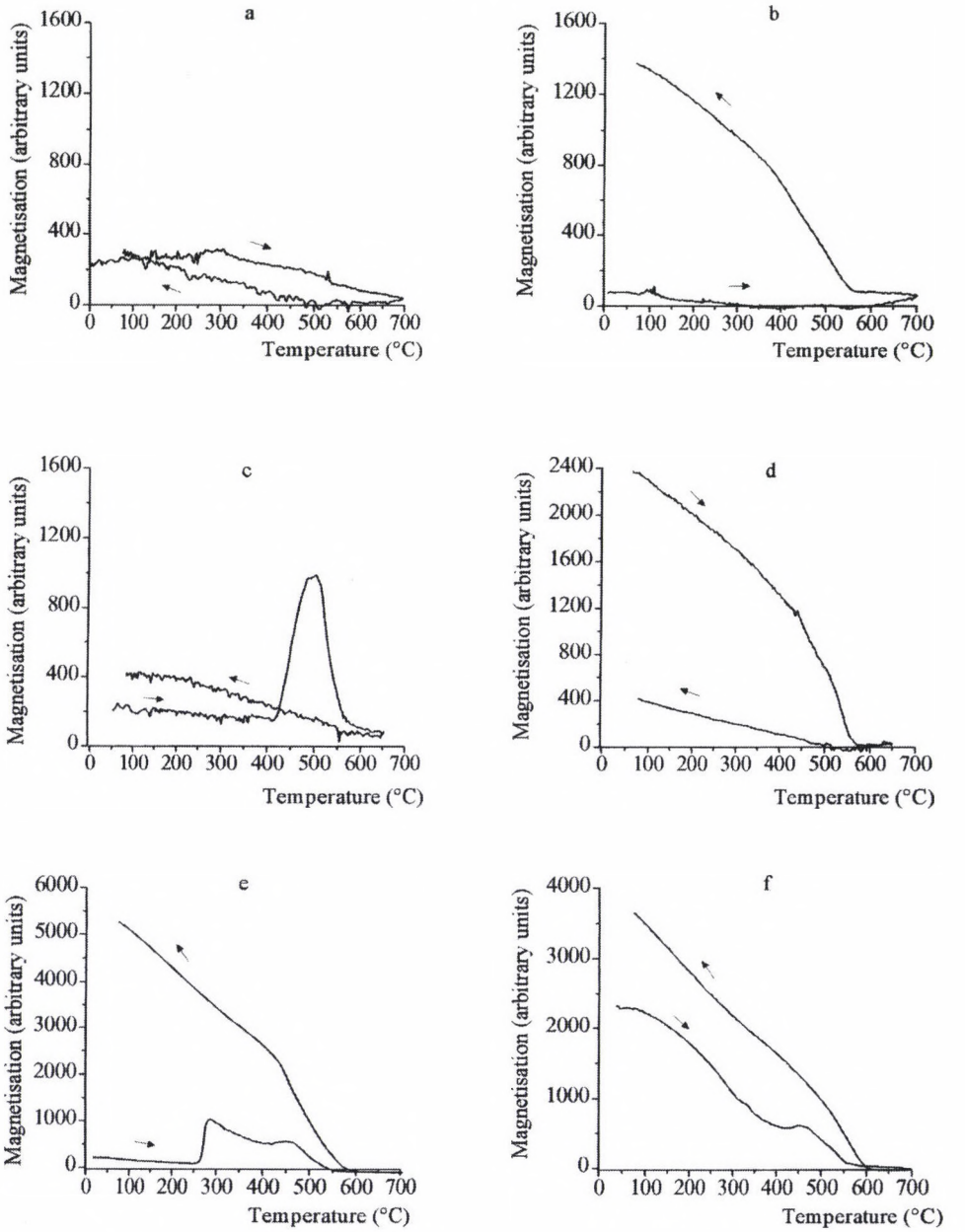


Fig. 3. Typical thermomagnetic behaviours of sediment from the Lake Fertő area when heated in air in a magnetic field of 0.4 T. a) Tadten-A 70 cm depth, b) Tadten-A 13 cm depth, c) Király 346 cm depth, d) Király 346 cm depth – sample heated to 500°C, cooled and heated the second time until 650°C, e) Tadten-A 36 cm depth, f) Tadten-A 36 cm depth – sample heated to 300°C, cooled and heated the second time until 650°C, d) and f) show thermomagnetic curves from the second heating until 650°C

and is not precisely identified, but it contains sulphur and iron ions. The observed behaviours, especially the two last ones with more ore less precisely identified iron minerals which transform with temperature (iron carbonates and iron sulphides), are independent of lithology. The presence of these minerals and their independence of lithology suggest they have a post-depositional (early diagenetic) origin.

During early diagenesis of sediments in anoxic conditions, an important pathway for the degradation of organic matter would be microbial sulphate reduction, detrital iron oxide reduction and methanogenesis. Bacterial reduction of detrital iron oxides (reduction of Fe^{+3} to Fe^{+2}) and sulphates (SO_4^{-2}), which are dissolved in water (saline and brackish water bodies) or originate from organic matter (fresh water bodies) increases the alkalinity of the sediment pore-waters and favours precipitation of iron sulphides with the most stable pyrite (FeS_2). The pathway by which it is formed from the initially amorphous precipitate of FeS involves the formation of intermediate iron monosulphides such as greigite and mackinawite. In environments with a low organic matter content and/or low sulphate conditions, iron sulphide formation will be limited to monosulphide, which persists upon burial. When all the sulphate present in the pore-waters is reduced and used to form iron sulphides, the organic matter degradation continues in the zone of methanogenesis. In this zone, if an Fe^{+2} reactive is present, iron carbonates (siderite) and/or iron phosphates (vivianite) can precipitate. Thus, the presence of iron carbonates, sulphides and phosphates in the sediment suggests two steps of organic matter degradation: in the zone of microbial sulphate reduction, and in the zone of methanogenesis (for more details see Jones and Bowser 1978, Berner 1971, 1980, 1981, Berner et al. 1979, Curtis 1987, Roberts and Turner 1993, Jelinowska et al. 1997, 1998).

1. The presence of iron sulphides and carbonates in Lake Fertő sediments suggests anaerobic early diagenetic conditions related to the bacterial degradation of the organic matter.

2. The presence of iron sulphides indicates the sulphate reducing zone in the sediment. Their transformation into magnetic monosulphides during thermomagnetic experiment suggests that it is monosulphide rather than pyrite; such transformations do not affect pyrite but is known for hexagonal pyrrhotite ($Fe_{1-x}S$; $0 < x > 0.13$), which alters to monoclinic pyrrhotite (Fe_7S_8) with a corresponding increase in magnetisation (Schwarz 1975, Dekkers 1988). If monosulphide persists in the sediment, it suggests that there is not enough sulphur ions to form pyrite, so possibly the basin was poor in salts or rich in iron ions. As this behaviour in the Lake Fertő is not related to only one lithological unit, it is most probable, that this is due to influence of the water salinity rather than to the iron abundance.

3. The presence of siderite suggests a zone of methanogenesis. In the samples with siderite we do not observe sulphides. This means that either they are not present, so the environment had no sulphate ions as in freshwater conditions or else that pyrite, which we did not observe with the methods we used, may be present in small quantities. In this case the salinity might have been higher.

4. It is interesting to note that the presence of monosulphides of iron and/or carbonates of iron indicate significant changes of the environment. Monosulphides suggest low salinity conditions while carbonates were formed either in freshwater or strongly salty water (if sulphur is absent or abundant respectively). Siderite in sandy sediments (related to fluvial transport) in Lake Fertő sequences would rather indicate a freshwater environment.

This study shows the potential of magnetic methods to trace the development of the basin in terms of the hydrological system (relation between evaporation and alimentation) of the lake. A more detailed study is, however, required to obtain a full record of its evolution.

References

Berner R A 1971: Principles of chemical sedimentology. McGraw Hill, New York

Berner R A 1980: Early diagenesis: a theoretical approach. Princeton Series in Geochemistry, Princeton University Press

Berner R A 1981: *J. Sed. Petrol.*, 51, 359–366.

Berner R A, Baldwin T, Holder G R 1979: *J. Sed. Petrol.*, 49, 1346–1350.

Curtis C 1987: In: Marine Clastic Sedimentology, J K Legett and G G Zuffa eds, Graham and Trotman, London, 108–123.

Dekkers M J 1988: *Geologica Ultraiectina*, 51.

Hoffman V 1992: *Phys. Earth Planet. Inter.*, 70, 288–301.

Hoffman V 1993: *Studia geoph. et geod.*, 37, 366–381.

Hus J J 1990: *Phys. Earth Planet. Inter.*, 63, 41–57.

Jelinowska A, Tucholka P, Gasse F, Fontes J-Ch 1995: *Geophys. Res. Let.*, 22, 953–956.

Jelinowska A, Tucholka P, Wieckowski K 1997: *Geophys. J. Int.*, 129, 727–736.

Jelinowska A, Tucholka P, Guichard F, Chalié F, Gasse F, Lefèvre I, Tribovillard N, Desprairies A, Badaut-Trauth D 1998: *Geophys. J. Int.*, 133, 499–509.

Jones B F, Bowser C J 1978: In: Chemistry, Geology, Physics. Lerman and Springer eds, New York, 179–227.

Kohlbeck F, Szarka L, Wesztergom V, Jelinowska A, Menvielle M, Tucholka P, Schott J J 1998: Limitations in identification of lake sediment layers by using precise geoelectric sounding. XXIII EGS, Nice

Kohlbeck F, Szarka L, Jelinowska A, Menvielle M, Schott J J, Tucholka P, Wesztergom V 2000: *Geophysical Transactions*, 43, 33–45.

Krs M, Novak F, Krsova M, Pruner P, Jansa J 1992: *Geologica Carpathica*, 43, 156–157.

Mann S, Sparks N H C, Frankel R B, Bazylinski D A, Jannasch H W 1990: *Nature*, 346, 258–261.

Reynolds R L, Tuttle F, Rice C A, Fishman N S, Karaszewski J A, Sherman D M 1994: *Am. J. Sci.*, 294, 485–528.

Roberts A P, Turner G M 1993: *Earth Planet. Sci. Lett.*, 115, 247–273.

Schwarz E J 1975: *Geol. Surv. Can.*, Paper 74–59, 1–24.

Snowball I F 1991: *Phys. Earth Planet. Inter.*, 68, 32–40.

Thompson R, Oldfield F 1986: Environmental magnetism. Allen and Unwin, London

Tric E, Laj C, Jehanno C, Valet J P, Kissel C, Mazaud A, Iaccarini S 1991: *Phys. Earth Planet. Inter.*, 65, 319–336.

Verosub K L, Roberts A P 1995: *J. Geophys. Res.*, 100, 2175–2192.

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 265–275 (2000)

MAGNETOTELLURIC PHASES: A CASE STUDY FROM THE SUKINDA THRUST AREA, EASTERN INDIA

K K Roy[1], S Srivastava[1], A K Singh[1]

[Manuscript received August 26, 1999, revised February 9, 2000]

Phases of the magnetotelluric impedance tensors are more important than the apparent resistivities. Because they suffer minimum static distortion and are useful for both qualitative and quantitative interpretation. In this paper, the nature and behaviours of the family of phases in magnetotellurics are discussed. Behaviours of the phases of rotated TE and TM mode impedances ϕ_{TE} or ϕ_{XY}, ϕ_{TM} or ϕ_{YX}; phases of the rotation invariant magnetotelluric tensors ϕ_B, ϕ_D and ϕ_C; phases of the telluric vectors ϕ_{ex} and ϕ_{ey} (discussed in the text) are presented using the field data collected across the Sukinda Collision Zone, District Dhenkanal, Orissa, India. All pseudosections show that phase is a more important tool than the apparent resistivity for qualitative interpretation. Phase pseudosections across the northern margin of the eastern ghats show the existence of a series of vertical faults just north of the collision zone and depth extent of these faults is around 12 to 15 km. Signature of the lithosphere – asthenosphere boundary is also reflected in the phase pseudosections. Electrical conductivity 2D MT models are obtained by joint inversion of apparent resistivity and phase. Phase is a must for magnetotelluric (MT) forward modelling and inversion.

Keywords: apparent resistivity; magnetotelluric phases; magnetotellurics; modelling

Introduction

In this paper the authors wanted to present the phases as a tool for qualitative interpretation of magnetotelluric data. The phases are ϕ_{XY}, ϕ_{YX}, ϕ_D, ϕ_C, ϕ_B, ϕ_{eX} and ϕ_{eY}. ϕ_{XY} and ϕ_{YX} are respectively the TE and TM mode phases of Cagniard impedances (Cagniard 1953); ϕ_B and ϕ_D are rotation invariant average and determinant phases (Berdichevskiy and Dmitriev 1976); ϕ_C is the central phase (Lilley 1993); ϕ_{eX} and ϕ_{eY} are the phases of the telluric vectors e_x and e_y (Bahr 1988). Twist and shear of Groom and Bailey (1989, 1991) are also measured as angles. These parameters suffer minimum static distortion. Therefore, they are better than apparent resistivities in the areas where static shift effects are prominent.

Both apparent resistivities and their phases are needed for 1D, 2D and 3D modelling. Only apparent resistivity can not and should not be used for inversion. Phase pseudosections are the important tools for qualitative and semiquantitative interpretation. Here also phases are much better parameters than apparent resistivities

[1]Department of Geology and Geophysics, Indian Institute of Technology, Kharagpur – 721 302, West Bengal, India

1217-8977/2000/$ 5.00

specially when apparent resistivity data are static shift contaminated. When apparent resistivities are more distorted than phases, one can try to retrieve apparent resistivity from phase data.

A series of examples are presented in this paper from the field data collected across the Sukinda Collision Zone, near Kamakhyanagar, district Dhenkanal, Orissa, Eastern India. Sukinda Collision Zone is the contact between the high-grade granulite terrain of the Proterozoic Eastern Ghats and the low-grade Archaean Singhbhum granites and the iron ore group of rocks.

Magnetotelluric phases

1. Phase of the rotated TE-mode Cagniard impedance (i.e., ϕ_{XY} or ϕ_{TE})

$$\phi_{XY} = \tan^{-1}\left[\frac{\mathrm{IM}\left|\frac{E_X}{H_Y}\right|}{\mathrm{Re}\left|\frac{E_X}{H_Y}\right|}\right]. \tag{1}$$

2. Phase of the rotated TM-mode Cagniard impedance (i.e., ϕ_{YX} or ϕ_{TM})

$$\phi_{YX} = \tan^{-1}\left[\frac{\mathrm{IM}\left|\frac{E_Y}{H_X}\right|}{\mathrm{Re}\left|\frac{E_Y}{H_X}\right|}\right]. \tag{2}$$

3. Phase of the rotation invariant tensor Z_{average}:

$$Z_{\text{average}}\ (Z_B \text{ or } Z_1) = \frac{1}{2}(Z_{XY} - Z_{YX}) \tag{3}$$

$$\phi_{\text{average}}\ (\phi_B \text{ or } \phi_{\text{ave}}) = \text{phase of } (Z_{XY} - Z_{YX})/2 \tag{4}$$

(Berdichevskiy and Dmitriev 1976).

Ingham and Hutton (1982) also modelled apparent resistivity and phase curves by using Z_B and ϕ_B which are invariant with the angle of the orientation of the measurement axes.

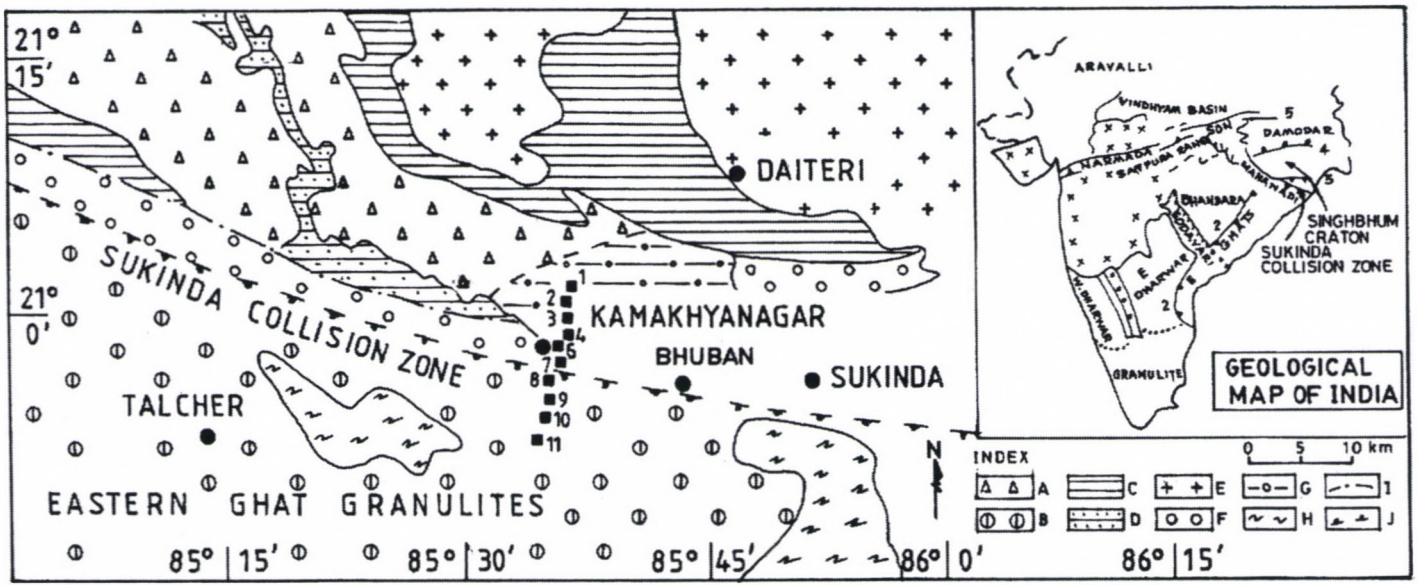
4. Phase of the rotation invariant tensor $Z_{\text{determinant}}$ or Z_{det} or Z_D.

$$Z_{\text{determinant}}(Z_D) = [Z_{XX}Z_{YY} - Z_{XY}Z_{YX}]. \tag{5}$$

Here Z_D has the dimension of Z^2.

$$\phi_{\text{determinant}}(\phi_D) = \text{phase of } [Z_{XX}Z_{YY} - Z_{XY}Z_{YX}]/2 \tag{6}$$

(Berdichevskiy and Dmitriev 1976).



1. Kankadahad ; 2. Urubangi ; 3. Batagaon-1 ; 4. Batagaon-2 ; 6. Kohnapura ; 7. Indipur ; 8. Mahulpal ; 9. Bijadihi ; 10. Kadalipal ; 11. Tumasingha ; A. Pala Lahara Gneiss ; B. High Grade Granulite Facies Eastern Ghat Rocks ; C. Iron Ore Group shales, tuffs, phyllites ; D. BHJ, BHQ and sandstone conglomerate of Iron Ore Group ; E. Singhbhum Granite Phase-III Bonai Granite, Chakradharpur Granite ; F. Kolhan Group and equivalents ; G. Mayurbhanj Granite H. Khondalite ; I. Fault ; J. Thrust fault.

Fig. 1. Location of the Singhbhum Craton in India; the location of the Sukinda Collision Zone and the geological map of the Singhbhum Craton; the location of the magnetotelluric sites across the Collision Zone

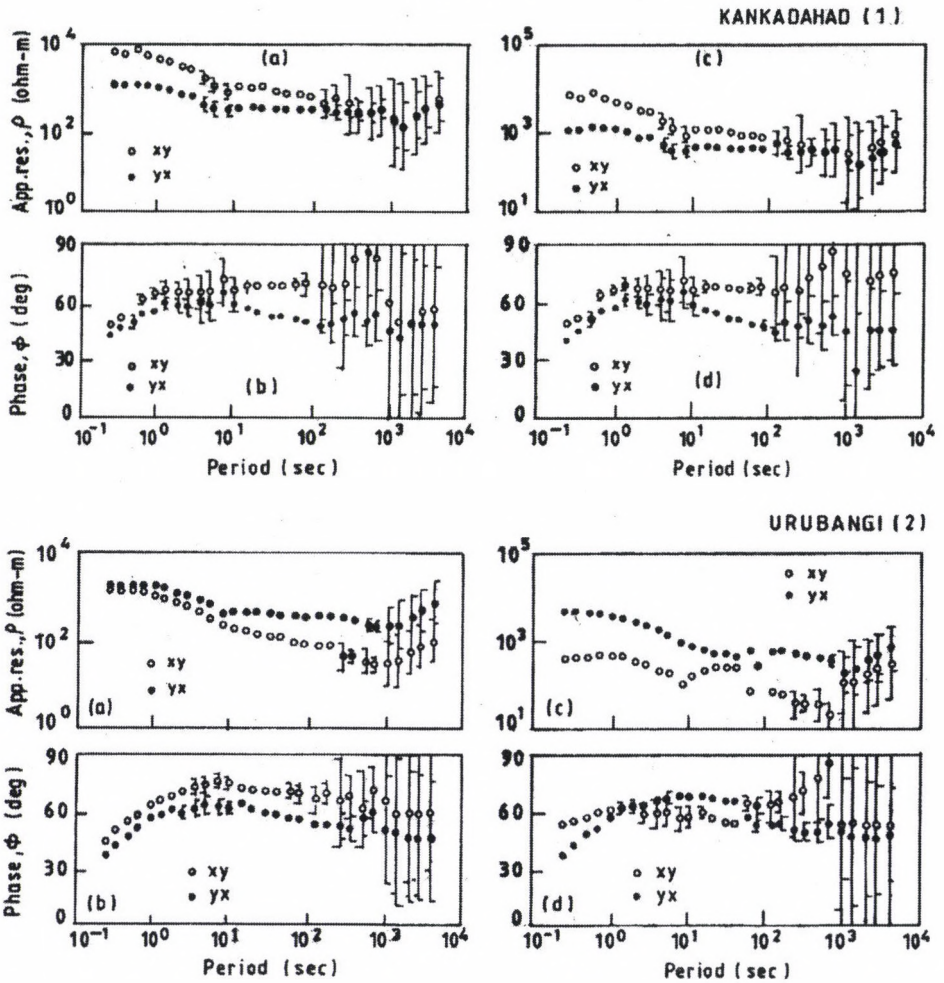


Fig. 2. Unrotated TE and TM mode a) apparent resistivities b) phases c) Swift rotated apparent resistivities and d) their phases for MT site Kankadahad (1) and Urubangi (2)

5. Phase of the rotation invariant tensor Z_{central} or Z_{cen} or Z_C.

$$Z_{\text{central}}(Z_C) = [X^2 + Y^2]^{1/2} \tag{7}$$

$$\phi_{\text{central}} = \tan^{-1}\left(\frac{Y}{X}\right), \tag{8}$$

where

$$X = \frac{1}{2}\left[(Z_{XXr} + Z_{YYr})^2 + (Z_{XYr} - Z_{YXr})^2\right]^{1/2} \tag{9}$$

$$Y = \frac{1}{2}\left[(Z_{XXi} + Z_{YYi})^2 + (Z_{XYi} - Z_{YXi})^2\right]^{1/2}, \tag{10}$$

where r and i stands for the real and imaginary components respectively (Lilley 1993).

6. Phases of the rotation invariant eigen values of the impedance tensor Z (Eggers 1982).

$$\lambda^+ = Z_1 + (Z_1^2 - \det Z)^{1/2} \tag{11}$$

$$\lambda^- = Z_1 - (Z_1^2 - \det Z)^{1/2}, \tag{12}$$

where det stands for determinant and λ^+, λ^- are the two eigen values of the impedance tensor Z (Eggers 1982). Here det Z is the determinant of the matrix Z and Z_1 is defined by Eq. (3).

Since Z_1 is rotationally invariant parameters, therefore it is implied that λ^+ and λ^- are also rotationally invariant parameters. The two apparent

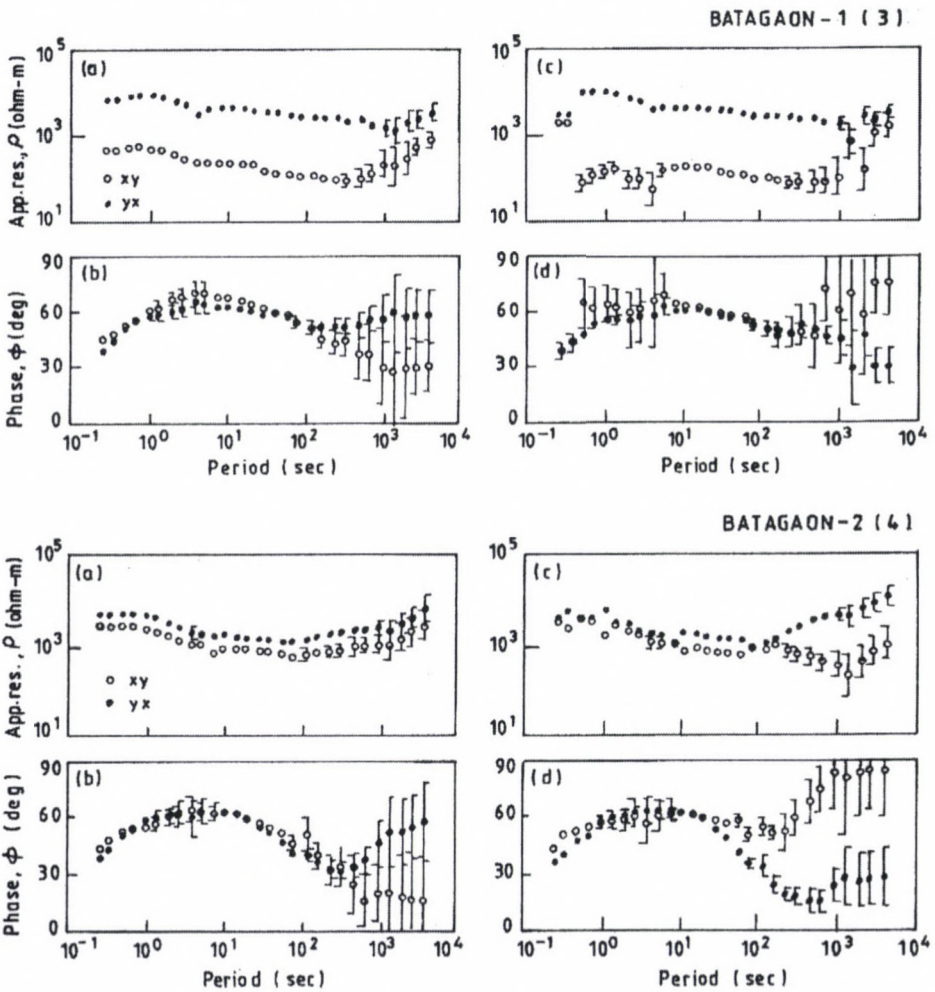


Fig. 3. Unrotated TE and TM mode a) apparent resistivities b) phases c) Swift rotated apparent resistivities and d) their phases for MT site Batagaon-1 (3) and Batagaon-2 (4)

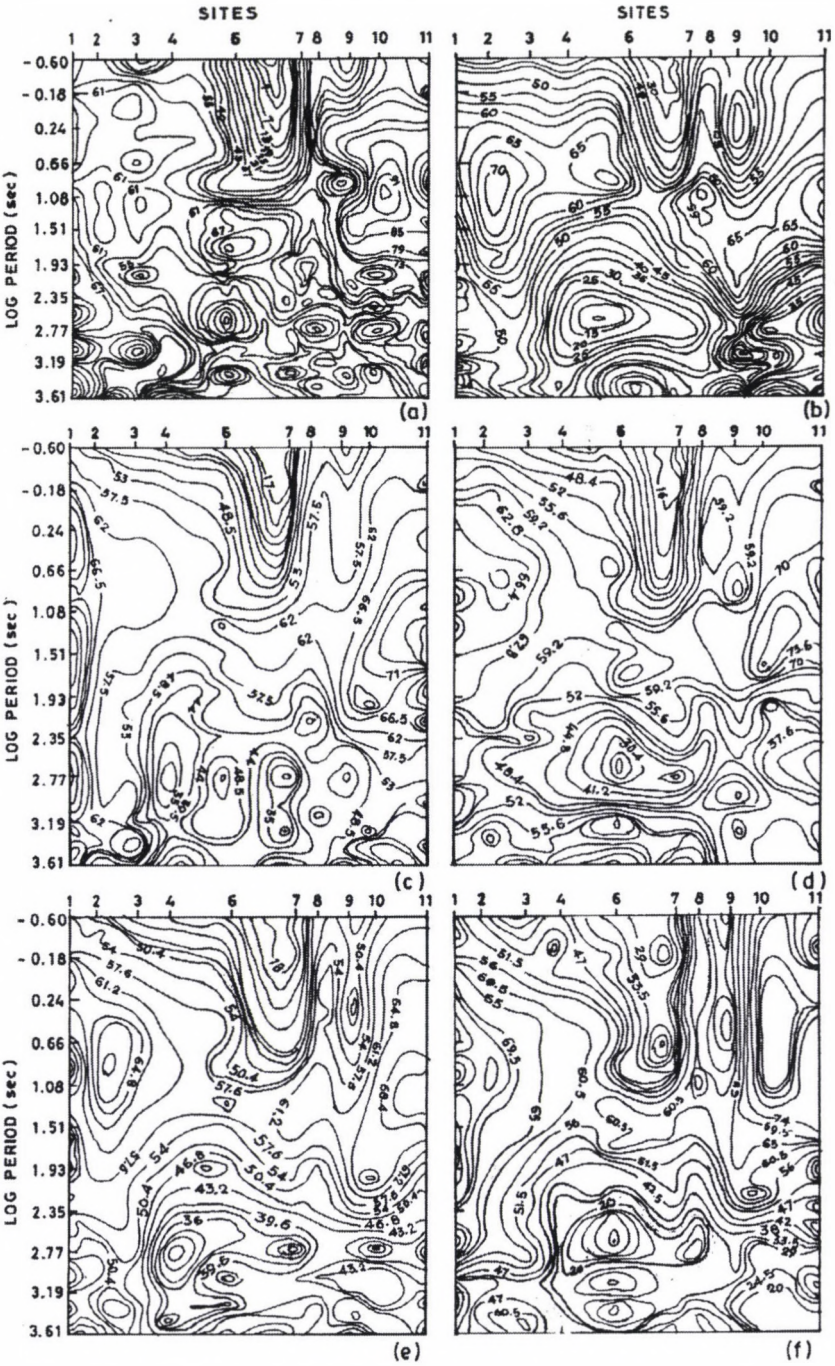


Fig. 4. Pseudosections of a) ϕ_{TE} or ϕ_{XY}, b) ϕ_{TM} or ϕ_{YX}, c) ϕ_{average} or ϕ_B, d) $\phi_{\text{determinant}}$ or ϕ_D, e) ϕ_{central} or ϕ_C, f) ϕ_{ex} along the Kankadahad-Tumasingha profile

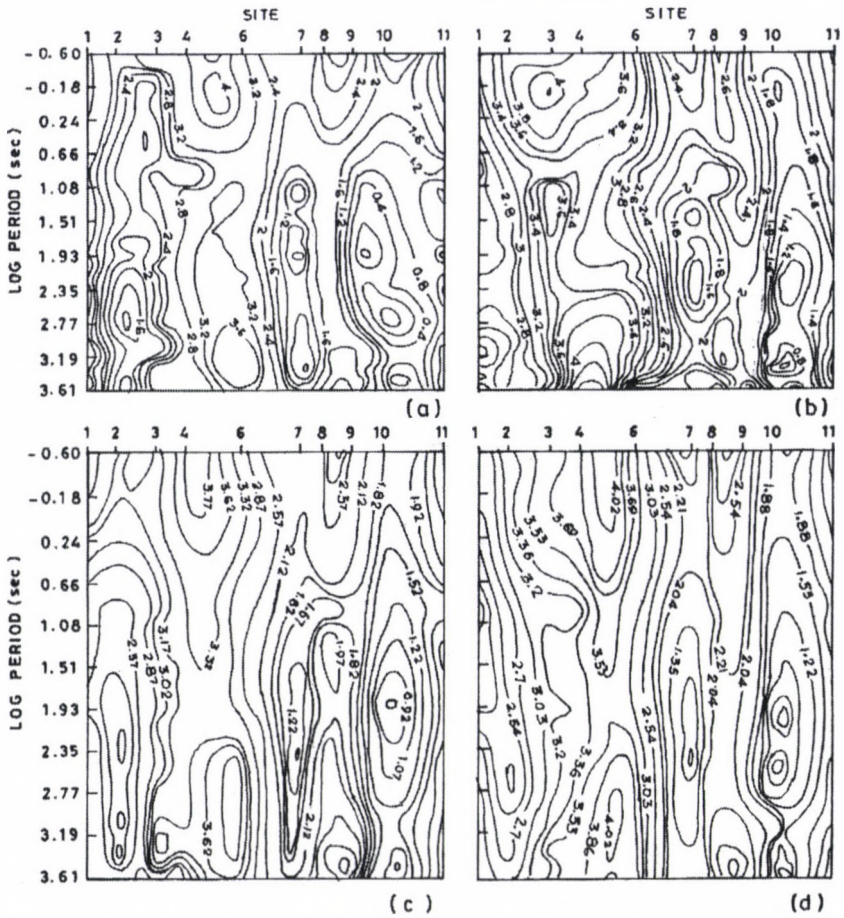


Fig. 5. Pseudosections of a) ϱ_{TE} or ϱ_{XY}, b) ϱ_{TM} or ϱ_{YX}, c) $\varrho_{\text{determinant}}$ or ϱ_D and d) ϱ_{central} or ϱ_C along the Kankadahad-Tumasingha profile

resistivities, using eigen values λ^+ and λ^- are

$$\varrho_{\lambda+} = \frac{1}{\omega\mu_0}|\lambda^+|^2 \tag{13}$$

$$\varrho_{\lambda-} = \frac{1}{\omega\mu_0}|\lambda^-|^2 . \tag{14}$$

$\phi_{\lambda+}$ and $\phi_{\lambda-}$ are respectively the phases of $\varrho_{\lambda+}$ and $\varrho_{\lambda-}$.

It is interesting to see that the field measured phases maintain the perfect relation

$$\phi_D = (\phi_{\lambda+} + \phi_{\lambda+})/2 . \tag{15}$$

Ranganayaki (1984) has shown that ϕ_D pseudosection almost depicts the subsurface with greater clarity. Ingham (1988) also suggested that the use of

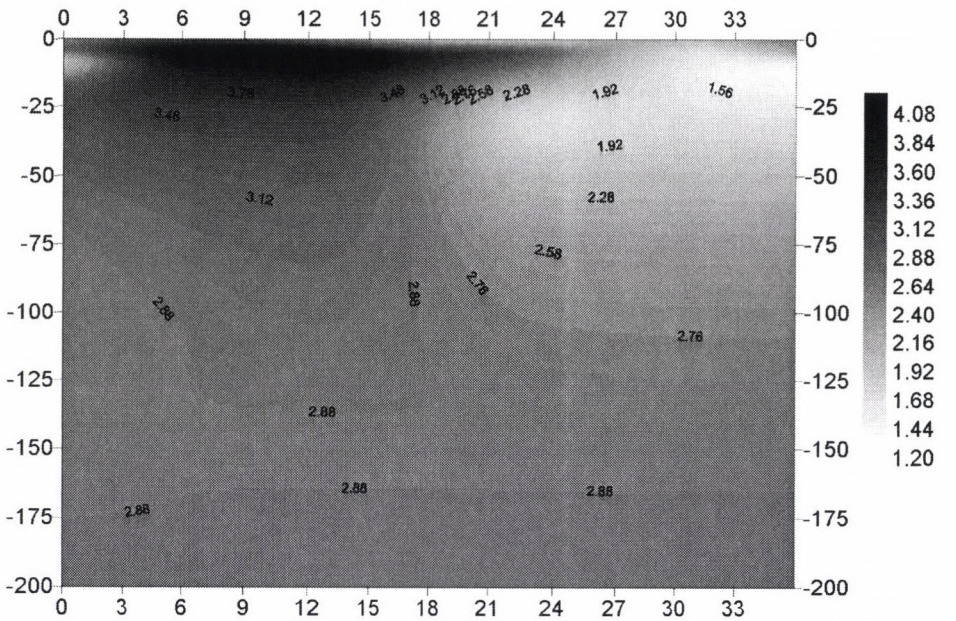


Fig. 6. 2D MT model of the Sukinda Collision Zone obtained by joint inversion of $\varrho_{\text{determinant}}$ (ϱ_D) and $\phi_{\text{determinant}}$ (ϕ_D) using Rapid Relaxation Inversion (RRI) algorithm of Smith and Booker (1991)

'invariant' impedances may provide a simple, relatively accurate method of magnetotelluric interpretation even in region of three-dimensionality.

7. Bahr (1988), from his theory of tensor decomposition has defined the concept of telluric vectors and their phases. The phases of the telluric vectors (ϕe_X and ϕe_Y) are given by the following relation

$$\phi e_X = \tan^{-1}\left[\frac{(\mathrm{Im}Z_{XX})^2 + (\mathrm{Im}Z_{YX})^2}{(\mathrm{Re}Z_{XX})^2 + (\mathrm{Re}Z_{YX})^2}\right]^{1/2} \tag{16}$$

$$\phi e_Y = \tan^{-1}\left[\frac{(\mathrm{Im}Z_{XY})^2 + (\mathrm{Im}Z_{YY})^2}{(\mathrm{Re}Z_{XY})^2 + (\mathrm{Re}Z_{YY})^2}\right]^{1/2}. \tag{17}$$

In this paper an attempt has been made to show that the phase plots are very important for qualitative interpretation of MT data.

Results and discussion

Figure 1 shows the location of the Sukinda Collision Zone and the magnetotelluric observation sites across the Collision Zone. Figures 2 and 3 show the rotated and unrotated Cagniard TE and TM mode apparent resistivities and their phases for the station Kankadahad (1), Urubangi (2), Batagaon-1 (3) and Batagaon-2 (4) respectively along the Kankadahad-Tumasingha profile. Figure 4 shows the phase

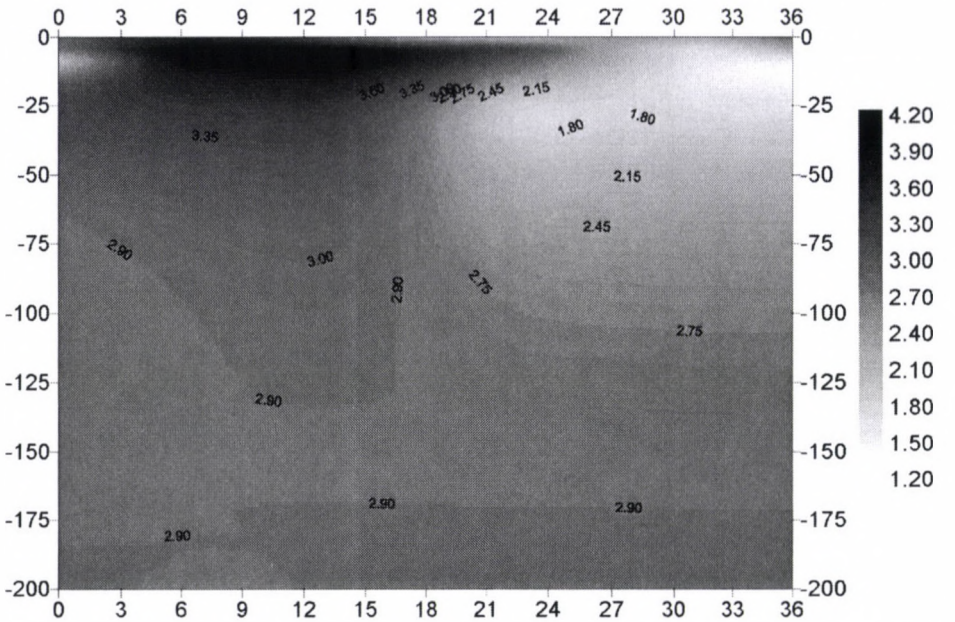


Fig. 7. 2D MT model of the Sukinda Collision Zone obtained by joint inversion of ϱ_{central} (ϱ_C) and ϕ_{central} (ϕ_C) using Rapid Relaxation Inversion (RRI) algorithm of Smith and Booker (1991)

pseudosections along the Kankadahad-Tumasingha (profile length: 36 km) for the parameters ϕ_{XY}, ϕ_{YX}, ϕ_B, ϕ_D, ϕ_C and ϕ_{ex} respectively. The pseudosections are plotted with the MT stations and their mutual distance as the abscissa and the log time period of the MT signals as the ordinate. The parameter values are presented in the form of the contours. The phase contours show that there are faults in this collision zone. Signatures of the faults are clearly seen in the pseudosections. The same type of signatures is repeated in all the phase pseudosections. Figure 5 shows the apparent resistivity pseudosections for the parameters ϱ_{XY}, ϱ_{YX}, ϱ_D and ϱ_C along the Kankadahad-Tumasingha profile. Since static shift problem exists in this area because of accumulation of charges at the faulted boundaries, apparent resistivity pseudosection can not be used. Static shift is a frequency independent vertical shift of the apparent resistivities (see Fig. 5). So the signatures of faults reaches upto the bottom of the pseudosections with MT signal period upto 4096 seconds.

These phases with minimum static distortion are more informative. Most of these pseudosections show that there are vertical faults. The signature of these faults went upto the log time period of 1.08 sec of the MT signal. The apparent skin depth of these signals is about 10 to 15 km.

It is interesting to note that all the phase pseudosections have given the signatures at the same location and the depth extent exhibited by most of the parameters are more or less the same. Phase pseudosections clearly show the brittle upper crust of thickness of 15 km near the Kamakhyanagar and are overlying the ductile lower crust. Phase pseudosections show three horizontal boundaries at the log time period

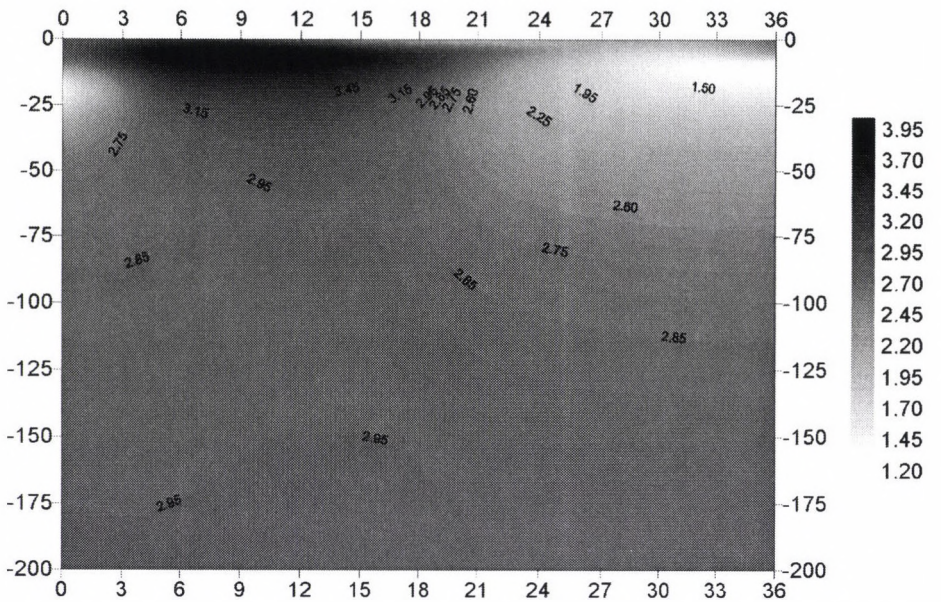


Fig. 8. 2D MT model of the Sukinda Collision Zone obtained by joint inversion of ϱ_{average} (ϱ_B) and ϕ_{average} (ϕ_B) using Rapid Relaxation Inversion (RRI) algorithm of Smith and Booker (1991)

of 3.19 sec., 2.77 sec. and 2.35 sec. The apparent skin depths for these time periods are respectively 200 to 300 km, 60 to 70 km and 30 to 40 km. The first horizontal boundary, which is at an apparent skin depth of 200 to 300 km, is the boundary of the lithosphere-asthenosphere. The other two horizontal boundaries are debatable but are obtained in many pseudosections.

Figures 6, 7 and 8 show the 2D MT model of the Sukinda Collision Zone obtained by three pairs of parameters viz., (ϱ_D, ϕ_D), (ϱ_C, ϕ_C) and (ϱ_B, ϕ_B) using Rapid Relaxation Inversion (RRI) algorithm of Smith and Booker (1991) respectively. The model shows a sharp change in the electrical conductivity at a depth of about 130 to 140 km. This contact is assumed to be the depth of the lithosphere-asthenosphere boundary. It may be a zone where there is unusual concentration of graphites or the hydrogen ions.

In this paper the authors tried to project (i) The importance of the qualitative interpretation of magnetotelluric data in the tectonically disturbed zone (ii) The role of MT phases of minimum static shift effect in qualitative interpretation. (iii) All quantitative interpretations should be the joint inversion of apparent resistivity and phase data. (iv) A series of vertical faults are indicated by the phases. (v) The nature of the contact of the Archaean Singhbhum Craton of low grade rocks and the Proterozoic high grade granulite suggests that it may be a collision zone.

Acknowledgement

Authors are grateful to the Department of Science and Technology, New Delhi for sanctioning of the project "Geophysical Investigation for Crustal Evolution and Geodynamic modelling of Lithosphere across the Sukinda Thrust, Orissa" (ESS/CA/A9- 22/92).

References

Bahr K 1988: *J. of Geophysics*, 62, 119–127.
Berdichevskiy M N, Dmitriev V I 1976: *Acta Geod. Geoph. Mont. Hung.*, 11, 447–484.
Cagniard L 1953: *Geophysics*, 18, 605–635.
Eggers D E 1982: *Geophysics*, 47, 1204–1214.
Groom R W, Bailey R C 1989: *J. Geoph. Res.*, 94(B2), 1913–1925.
Groom R W, Bailey R C 1991: *Geophysics*, 56, 496–518.
Ingham M R 1988: *Geoph. J. Int.*, 92, 165–169.
Ingham M R, Hutton V R S 1982: *Geoph., J.R. astr. Soc.*, 69, 579–594.
Lilley F E M 1993: *Geophysics*, 58, 1498–1506.
Ranganayaki R P 1984: *Geophysics*, 49, 1730–1748.
Smith J T, Booker J R 1991: *J. Geoph. Res.*, 96(B3), 3905–3922.

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 277–282 (2000)

INFLUENCE OF TEMPERATURE AND BAROMETRIC PRESSURE VARIATIONS ON EXTENSOMETRIC DEFORMATION MEASUREMENTS AT THE SOPRON STATION

G MENTES[1]

[Manuscript received September 22, 1999]

Results of the deformation measurements obtained by a quartz tube extensometer installed at the Geodynamical Observatory of the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences in Sopron (Hungary) are described. In the same time the dependence of the deformation data on the temperature and pressure is also investigated.

Keywords: deformation; deformation measurement; quartz tube extensometer

1. Introduction

In 1990 a quartz tube extensometer with electrical recording was installed at the Geodynamical Observatory of the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences in Sopron (Mentes 1991). The instrument is assembled from quartz tubes with length of 2.5 m. The total length of the instrument is 22 m, its azimuth is 116° and the co-ordinates of the observatory are: latitude: 47.7° N; longitude: 16.5° W. Figure 1 shows the scheme of the observatory which is an artificial tunnel driven in gneiss. The distance of the extensometer from the entrance is about 30 m, the overlay of the gallery is 60 m thick. Due to technical reasons the instrument is not in the tunnel axis. The extensometer is separated from the gallery by thermal insulation. The annual temperature variation in the gallery at the instrument is less then 0.5°C and the daily one is about 0.05°C.

The regular, daily calibration of the extensometer is made by a magnetostrictive coil. In addition the instrument was in-situ calibrated by the method developed by Mentes (1995, 1998) in the observatory. The calibration factor obtained by this measurements: 0.219 nm/mV. Figure 2 shows the principle of the extensometers. As it can be seen the bedrock is part of the instrument. By the usual calibration of the instrument only the rigidity of the rod and the sensitivity of the electronics can be controlled. If there are faults in the bedrock they can cause uncontrollable displacements and therefore the instrument does not properly measure the tectonic movements.

To avoid the errors caused by the local movements a tidal evaluation of the data is also carried out. The instrument and its place are supposed to be suitable

[1]Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H–4901 Sopron, POB 5, Hungary, e-mail: mentes@ggki.hu

1217-8977/2000/$ 5.00

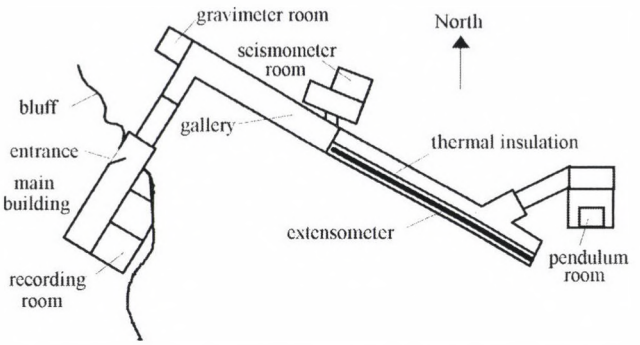


Fig. 1. Scheme of the Geodynamical Observatory in Sopron

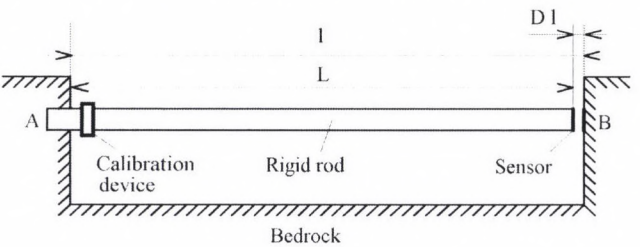


Fig. 2. The principle of extensometers

for recording tectonic movements if the tidal waves can be detected. The results of the tidal evaluation show that the instrument functions properly. In this paper only the results of the deformation measurements are given and the effects of the barometric pressure and temperature variations are investigated because they can cause large local movements and therefore suppress the tectonic movements in the recorded data.

2. Result of the deformation measurements

The extensometer has recorded since May, 1990. Because of the data gaps in the first year usable data are only available from 1991 till present. The slow tectonic movements can be very simply obtained from the recorded extensometric data by filtering the data by means of a low-pass filter. In the present case an average was calculated from the hourly sampled data for each day and the daily data were filtered by a 28 day moving average filter. Figure 3 shows the obtained deformation curve from 01.01.1991 till 31.12.1996. The same curve can also be obtained after the tidal analysis subtracting the tidal waves from the hourly data series but the averaging method is much more simple. This curve contains yet seasonal and yearly variations caused probably by weather fluctuations. These variations get eliminated if we calculate the trend of the deformation curve. The trend gives the rate of the displacement which is a compression of about 27 μm/year.

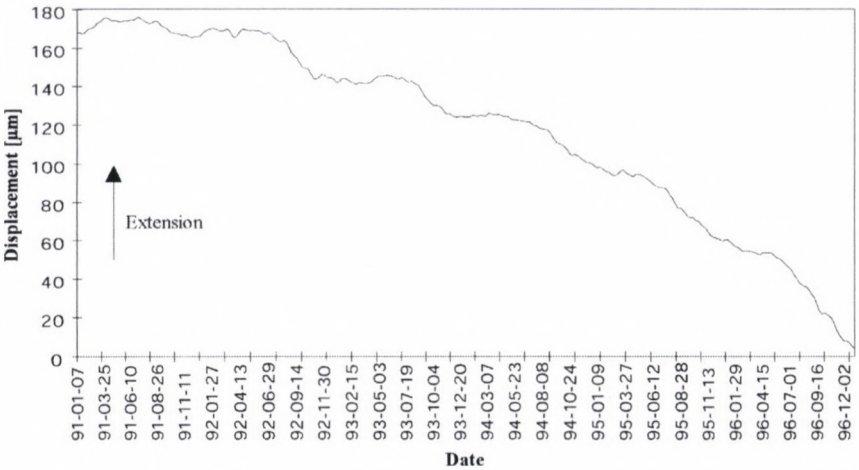


Fig. 3. Deformation curve measured by extensometer at the Sopron Geodynamical Observatory from 01.01.1991 till 31.12.1996

3. Effect of the temperature

In order to examine the relationship between the extensometric data and the temperature the record made in 1996 was used. Because the daily temperature variations in the observation room are less than 0.05°C, the extensometer is not disturbed by the inner temperature variations. Therefore in the first step the relationship between the daily averages of the outdoor temperature and strain data was investigated. To calculate the correlation between strain and temperature data the trend was subtracted from both data series. The variations of the strain and the ones of the outdoor temperature are shown in Fig. 4. The correlation coefficient between the two curves is $r = 0.852$. In the second step both data were filtered by a 14 day moving average filter (Fig. 5) then the correlation changed to $r = 0.919$. Applying a 28 day moving average filtering to both data (Fig. 6) we obtain a correlation coefficient $r = 0.933$. These results show that the deformation of the bedrock cannot follow the short periodic variations of the outdoor temperature. The correlation coefficient has a maximum value of about $r = 0.965$ if we shift the temperature curve by 18 days backward. That means that there is a phase shift of about 18 days between the outer temperature and the measured deformation caused by it. Calculating a regression coefficient between the displacement and the shifted temperature the displacement is obtained between the ends of the extensometer caused by a temperature change of 1°C. At the present extensometer with a length of 22 m this displacement is about 0.5 μm.

The obtained results prove very clearly that the extensometric record is disturbed by the variations of the outdoor temperature and it should be corrected before tidal evaluation. In the case of the deformation evaluation the effect of the temperature is eliminated because of the very strong low-pass filtering (calculation of long term averages) and the calculation of the trend. For this reason long (many years) continuous records are needed to determine the tectonic movements.

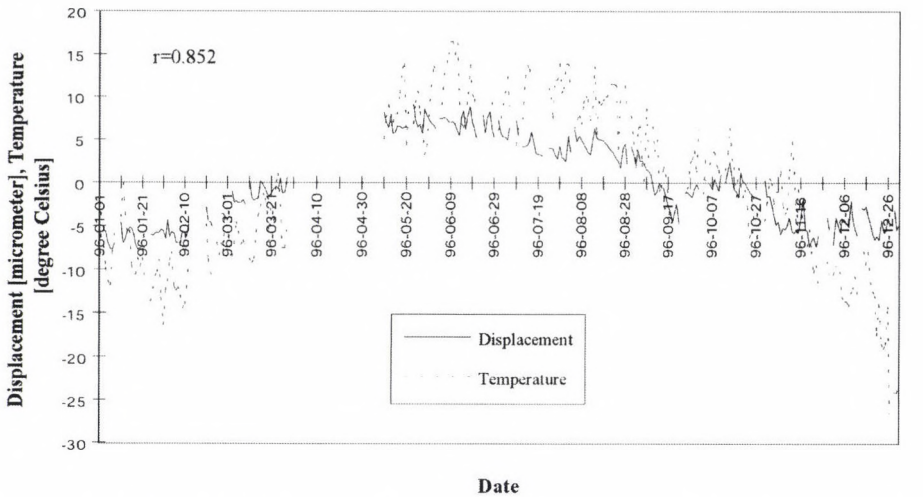


Fig. 4. Variations of the daily averages of the extensometric data and the ones of the outdoor temperature measured at the Sopron Geodynamical Observatory in 1996

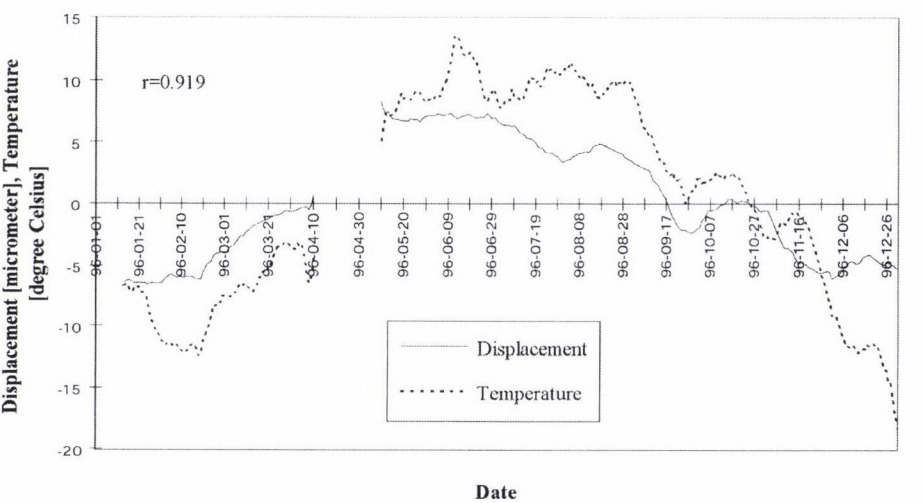


Fig. 5. The variations of the daily averages of the extensometric data and the ones of the outdoor temperature filtered by a 14 day moving average filter

4. Effect of the atmospheric pressure

The investigation of the influence of the air pressure variations on the tidal measurements similarly to the temperature effects is a very important task because the deformation of the rock, and the one of the recording room at the observatory due to the atmospheric loading can cause large errors at all kinds of instruments. Therefore the precise correction of this effect is very important (Kroner and Jentzsch 1998).

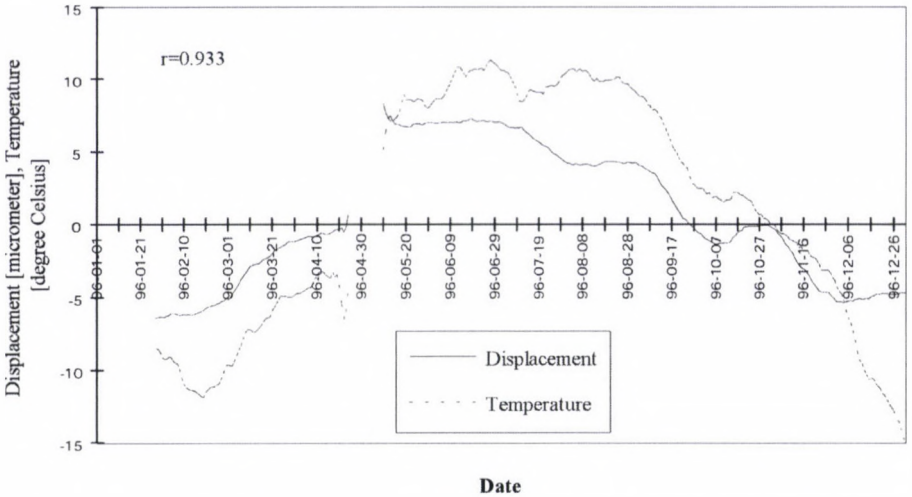


Fig. 6. The variations of the daily averages of the extensometric data and the ones of the outdoor temperature filtered by a 28 day moving average filter

Data preparation for investigation of the influence of the air pressure variations to the measured deformation was made identically to the one described in the previous chapter. Figure 7 shows the variations of the daily averages of the extensometric data and the ones of the air pressure measured in 1996. The correlation coefficient between the two curves is $r = 0.101$. The change of the correlation coefficient is not perceptible if the data are much stronger filtered or shifted relatively to each other. That means that the air pressure variations have negligible influence on the extensometer data. Onoue and Takemoto (1998) obtained similar result for extensometer at the Donzurubo Observatory. For two other extensometers they found a close correlation between the measured displacements and the air pressure variations. They state that this difference between the extensometers is due to the topography of the surrounding area of the observatory.

Figure 1 shows that the extensometer is not perpendicular to the rock-wall and therefore the instrument is not sensitive to the displacements caused by the air pressure variations. Another reason is that due to the construction and the topography of the area of the observatory the extensometer is deeply placed below the surface.

Earlier tidal measurements by means of horizontal pendulums prove that the Sopron Observatory is mechanically very stable.

5. Conclusions

The results of the investigations showed that the temperature variations can strongly influence the extensometric data. But this effect is eliminated if we determine only tectonic movements from very long (many years) data series by calculating the trend of the measured displacements. On the other hand the recorded data should be corrected before tidal evaluation.

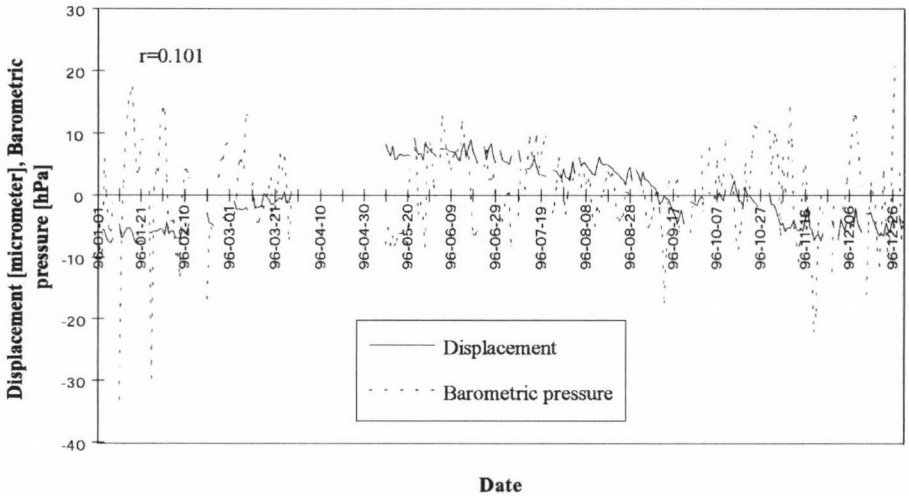


Fig. 7. Variations of the daily averages of the extensometric data and the ones of the air pressure measured at the Sopron Geodynamical Observatory in 1996

In contrast with the effect of the temperature variations the atmospheric loading has no remarkable influences on the extensometric data at the Sopron station.

The results of the deformation measurements show that the extensometer and the recording site are suitable for observation of tectonic movements. For elimination of the disturbing effects a complex analysis of the measured data is needed. The obtained results contribute to a better understanding of the dynamics of the recording site and the surrounding rock and can be used for the correction of the cavity effect (Brimich et al. 1998).

Acknowledgements

This research was supported by the Hungarian National Scientific Research Foundation (OTKA) in the frame of the research projects T 014896 and T 019030.

References

Brimich L, Kohút I, Kostecký P 1998: In: Proceedings of the 13th International Symposium on Earth Tides, Brussels, 397–412.

Kroner C, Jentzsch G 1998: In: Proceedings of the 13th International Symposium on Earth Tides, Brussels, 423–430.

Mentes Gy 1991: *Marees Terrestres Bulletin d'Informations, Bruxelles*, 110, 7936–7939.

Mentes Gy 1995: *Marees Terrestres Bulletin d'Informations, Bruxelles*, 121, 9070–9075.

Mentes Gy 1998: In: Proceedings of the 13th International Symposium on Earth Tides, Brussels, 43–50.

Onoue K, Takemoto S 1998: In: Proceedings of the 13th International Symposium on Earth Tides, Brussels, 157–164.

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 283–294 (2000)

ANALYTICAL DETERMINATION OF Q_{MED} (I.E., OF THE ERROR OF SAMPLE MEDIANS) FOR ARBITRARY $\underline{n}$ SAMPLE SIZE

L Csernyák[1], B Hajagos[1], F Steiner[2]

[Manuscript received August 17, 1999]

Analytical way is given to determine the error (Q or q) of the sample medians for arbitrary (i.e., also for very small) sample sizes. The also presented Monte Carlo solution is computer time consuming but valid for whatever estimates, too, e.q., for the most frequent values.

Keywords: asymptotic scatter; median; sample median; semi-interquartile range; semi-intersextile range

1. Introduction

$F(x)$ and $f(x)$ denote the distribution and density functions of a random variable. This parent distribution has the meaning of measuring errors. In the present investigations these parent distribution can be an arbitrary type of the complete $f_c(x)$ supermodel, defined by Eq. (1) by Steiner and Hajagos (1998) (and referred by Steiner and Hajagos (2000) also as Eq. 1).

For L_1-based procedures the central notion is the theoretical median (med_{th}) defined by

$$\int_{-\infty}^{\mathrm{med}_{th}} f(x)dx\,. \tag{1}$$

In asymptotical case, i.e., if the number $\underline{n}$ of the elements x_i in the sample coming from an arbitrary $f_c(x)$ standard error-distribution tends to the infinity, the scatter σ of the sample medians can be calculated simply as

$$\sigma = \frac{A}{\sqrt{n}}\,, \tag{2}$$

where the value A (called asymptotic scatter) can be also easily determined as

$$A = \frac{1}{2 \cdot f(\mathrm{med}_{th})}\,. \tag{3}$$

In the practice it is usually supposed that if n is large enough, the error of the sample medians can be characterized according to Eq. (2) by the scatter or

[1]Institute of Mathematics, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary
[2]Geophysical Department, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary

1217-8977/2000/$ 5.00

with 0.9674 times the scatter, because this product results in the semi-intersextile range Q if the estimates are Gauss-distributed. (The product $0.6745 \cdot \sigma$ equals for the Gaussian distribution the semi-interquartile range denoted by q and is called by Bessel as "probable error".) Huber (1981), however, verified that under rather general conditions the type of the estimates are of Gaussian type.

2. Analytical determination of Q_{MED} (and of q_{MED}) for an arbitrary even sample size $\underline{n}$

We restrict ourselves to even $\underline{n}$-values. The conventional definition of the sample median (denoted by "med") is well known in this case

$$\mathrm{med} = \frac{1}{2}\left[x_{n/2} + x_{n/2+1}\right] . \tag{4}$$

This definition is, however, not consequent in respect of the L_1-norm philosophy (which is accentuated by the basic Eq. (1)): the mean value supposes that we accept the L_2 philosophy, too. Instead to use a definition based upon a mixture of philosophies, Steiner and Hajagos (1999a) gave a more sophysticated definition (denoted by MED) as Eq. (3a) for the sample medians

$$\mathrm{MED} = \left\{ \begin{array}{l} \text{a single random number from the ones} \\ \text{uniformly distributed between } x_{n/2} \text{ and } x_{n/2+1} \end{array} \right. . \tag{5}$$

The definition given by Eq. (5) hardly needs any comment: all values which are greater than $x_{n/2}$ and less than $x_{n/2+1}$ are equally justified supposed to be a MED-value (the actual choose is made by a random number generator); really, the sample can be written (using Dirac-δ-s) as a density function of the form $\frac{1}{n}\sum_{i=1}^{n}\delta(x_i)$ and comparing this with the definition of med_{th} in Eq. (1), we find a concordance with the MED-definition given in Eq. (5). We use no more mixed philosophies, only L_1 is accepted.

$G(\mathrm{MED})$ and $g(\mathrm{MED})$ denote the distribution and density functions of the (n-dependent) set of MED estimates; the semi-intersextile range

$$Q_{\mathrm{MED}} = G(5/6) \tag{6}$$

is accepted to characterize the error of the MED-estimates. This value Q_{MED} depends naturally upon the parent distribution (this statement is valid also for the asymptotic case, see Eq. (3)) but Q_{MED} depends also upon the sample size $\underline{n}$: the degree of dependence is expected to be not negligible especially in case of small samples.

For determining Q_{MED} the conceptual simplicity of definition in Eq. (5) has evident advantages. Greater MED-values as $G(5/6)$ (see Eq. (6)) occur with probability $1/6$ but in two ways:

a) the relation

$$x_{n/2} \geq Q_{\mathrm{MED}} \tag{7}$$

is fulfilled and therefore the definition in Eq. (5) generates a MED-value (with the probability 1) for which MED $\geq Q_{\mathrm{MED}}$ is fulfilled;

b) the second variant: the relations

$$x_{n/2} < Q_{\mathrm{MED}} \tag{8a}$$

and

$$x_{n/2+1} \geq Q_{\mathrm{MED}} \tag{8b}$$

are fulfilled simultaneously and in this case the relation MED $\geq Q_{\mathrm{MED}}$ is valid with the probability

$$\frac{x_{n/2+1} - Q_{\mathrm{MED}}}{x_{n/2+1} - x_{n/2}} \,. \tag{9}$$

Denoting the whole probabily to get MED-values in way a) by P_1 and the whole probability of the b) variant by P_2, evidently

$$P_1 + P_2 = 1/6 \tag{10}$$

defines the Q_{MED}-value.

Cramér (1945) (and also Vincze (1968)) published (fortunately enough) not only the occurring probability density function of a sample element characterized by an arbitrary index i for ordered samples

$$f_{x_i}(x) = n \left(\begin{array}{c} n-1 \\ i-1 \end{array} \right) [F(x)]^{i-1} \cdot [1 - f(x)]^{n-i} \cdot f(x) \tag{11a}$$

(see p. 76 by Vincze (1968)) but also densities of the simultaneous occurrence of two elements characterized by the index-pairs i and k. We write this two variable probability immediately for the case $i = n/2$ and $k = n/2 + 1$:

$$\begin{aligned} P(x_{n/2} < x, x_{n/2+1} < y) \quad = \quad & \frac{2 \cdot (n/2)!}{[(n/2-1)!]^2} \cdot \int_{-\infty}^{x} \int_{-\infty}^{y} [F(u)]^{n/2-1} \cdot \\ & \cdot [1 - F(v)]^{n/2-1} f(a) f(v) du dv \,. \end{aligned} \tag{11b}$$

Having these two equations, it is a mathematical task to give the formulas for both P_1 and P_2:

$$P_1 = \frac{n}{2} \left(\begin{array}{c} n \\ n/2 \end{array} \right) \int_{Q_{\mathrm{MED}}}^{\infty} F(t)^{n/2-1} \cdot [1 - F(t)]^{n/2} \cdot f(t) dt \tag{12}$$

and

$$\begin{aligned} P_2 \quad = \quad & \tfrac{n^2}{4} \left(\begin{array}{c} n \\ n/2 \end{array} \right) \int_{-\infty}^{Q_{\mathrm{MED}}} \left[\int_{Q_{\mathrm{MED}}}^{\infty} \frac{z - Q_{\mathrm{MED}}}{z - y} [1 - F(z)]^{n/2-1} \cdot f(z) dz \right] \cdot \\ & \cdot F(y)^{n/2-1} f(y) dy \,. \end{aligned} \tag{13}$$

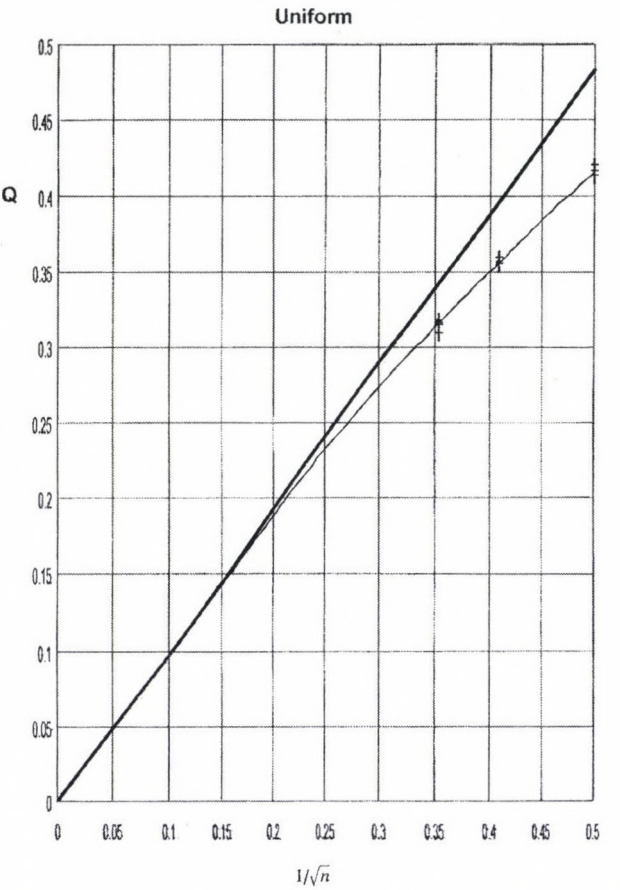


Fig. 1. Semi-intersextile ranges Q of sample medians vs. $x = 1/\sqrt{n}$ in case of the parent distribution of uniform type. The straight line represents the asymptotic rule. The "+" signs correspond to Monte Carlo results (see Section 4)

If the semi-interquartile range is preferred as error characteristics instead of Q_{MED}

$$q_{\mathrm{MED}} = g(3/4)\,, \tag{14}$$

the only difference is the substitution of q_{MED} in both Eqs (12) and (13), — in this case, however, instead of Eq. (10) the followings must be postulated for the P_1 and P_2 probabilities:

$$P_1 + P_2 = 1/4\,. \tag{15}$$

Table I

Number of the figure	Type	Q_∞	$Q_{MED}(1/\sqrt{n})$-values in case of			Coefficient of the polynom in Eq. (17)		
			$n=4$	$n=6$	$n=8$	a	b	c
1	uniform	0.96742157	0.41500900	0.35579034	0.13601485	0.02011102	–0.80342452	0.39015948
2	Gaussian	1.21248312	0.56806503	0.47119849	0.41202751	–0.00452043	–0.43093242	0.37616697
3	$a=9$	0.42784205	0.21116214	0.17384405	0.15150256	0.24206723	–1.00136637	0.93129027
4	$a=5$	0.64494771	0.31457800	0.25677805	0.22292926	0.00056976	–0.37650568	0.55485059
5	$a=3$	0.96742157	0.49602386	0.39672856	0.34137487	0.01877360	–0.38919558	0.90694117
6	Cauchy	1.51962224	0.88683084	0.67218012	0.56512951	0.12532887	–0.80305460	2.44217549

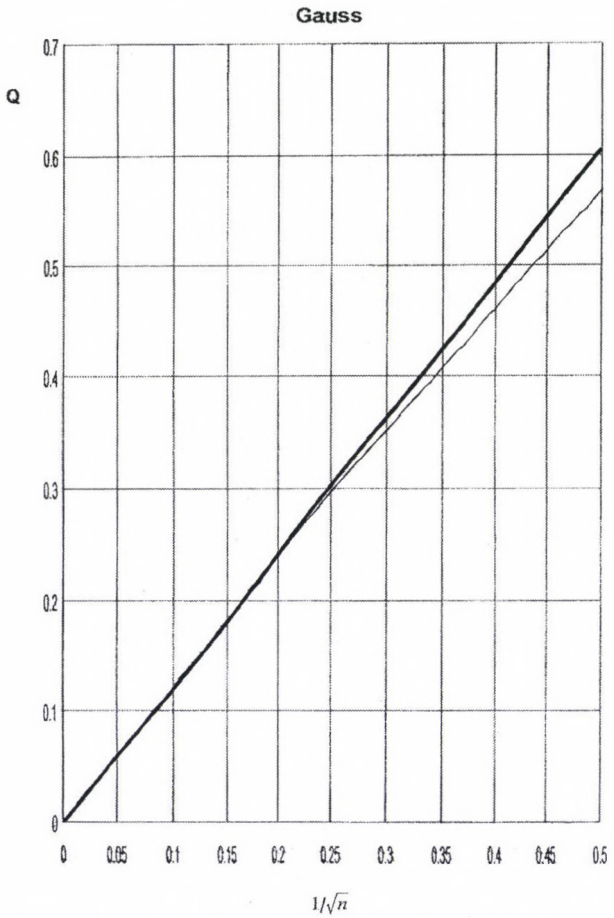


Fig. 2. Semi-intersextile ranges Q of sample medians vs. $x = 1/\sqrt{n}$ in case of the parent distribution of Gaussian type. The straight line represents the asymptotic rule

Table II

Type	Semi-intersextile ranges of the MED (sample median) values for six parent distribution types depending upon the sample size n
uniform	$Q(n) = 0.96742157x + 0.01945584x^2 - 0.77725021x^3 + 0.37744870x^4$
Gaussian	$Q(n) = 1.21248312x - 0.00548095x^2 - 0.52249829x^3 + 0.45609610x^4$
Jeffreys ($a = 9$)	$Q(n) = 0.42784250x + 0.10356654x^2 - 0.42842664x^3 + 0.39844514x^4$
statistical ($a = 5$)	$Q(n) = 0.64494771x + 0.00036746x^2 - 0.24282648x^3 + 0.35784961x^4$
midway ($a = 3$)	$Q(n) = 0.96742157x + 0.01816198x^2 - 0.37651648x^3 + 0.87739445x^4$
Cauchy	$Q(n) = 1.51962224x + 0.19045254x^2 - 1.22033963x^3 + 3.71118419x^4$

3. Results of the analytical calculations of Q_{MED}

The curve definited by the Q_{MED}-values vs. $1/\sqrt{n}$ starts as a straight line characterized by the first derivative

$$Q_\infty = 0.9674 \cdot A\,, \tag{16}$$

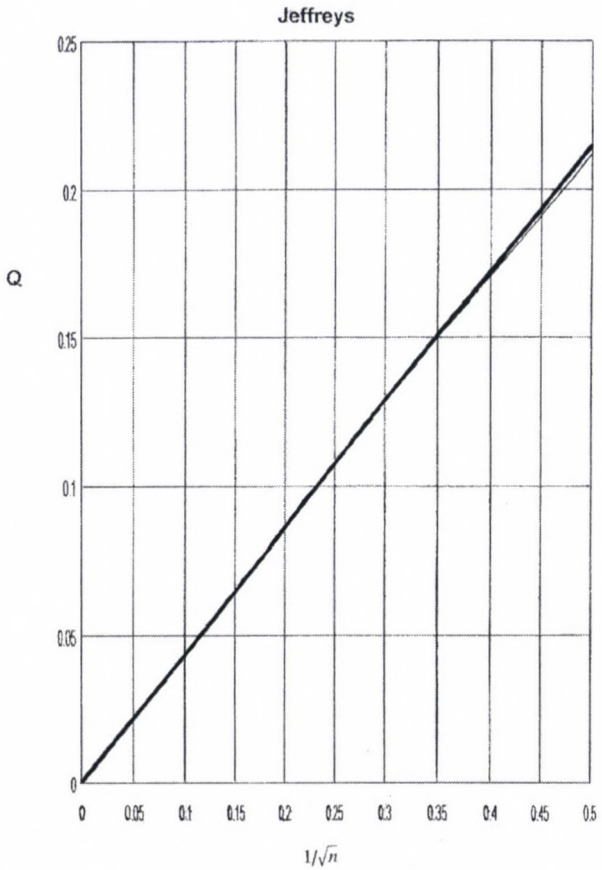


Fig. 3. Semi-intersextile ranges Q of sample medians vs. $x = 1/\sqrt{n}$ in case of the parent distribution of Jeffreys-type. The straight line represents the asymptotic rule. The seeming coincidence has its graphical reasons: at $x = 0.5$ the difference is 1.3%

where A is defined by Eq. (3). Fortunately enough, the Q_{MED} $(1/\sqrt{n})$ curves are in such a high degree smooth that a fourth order polynomial is enough to describe it, on one hand, and on the other hand, for the determination of the constants $\underline{a}$, $\underline{b}$ and $\underline{c}$ of this polynomial

$$Q_{\text{MED}}(1/\sqrt{n}) = Q_\infty \frac{1}{\sqrt{n}} \left[1 + \frac{a}{\sqrt{n}} + \frac{b}{n} + \frac{c}{n \cdot \sqrt{n}}\right] \tag{17}$$

is enough to determine (using Eqs (10), (12) and (13)) only three $_{\text{MED}}(1/\sqrt{n})$ values for $n = 4$, $n = 6$ and $n = 8$, and after this only interpolation-like calculation is necessary: the three accurately known $Q_M(1/\sqrt{n})$ values obviously give us the coefficients $\underline{a}$, $\underline{b}$ and $\underline{c}$ in Eq. (17). (It should be once more mentioned that the Q_∞ value is the product of A given by Eq. (3) and of the known conversion number 0.9674.) The four accurately known values Q_∞ and for $n = 4$, 6 and 8 calculated $Q_{\text{MED}}(1/\sqrt{n})$-values and the $\underline{a}$, $\underline{b}$ and $\underline{c}$ polynomial coefficients are given in Table I;

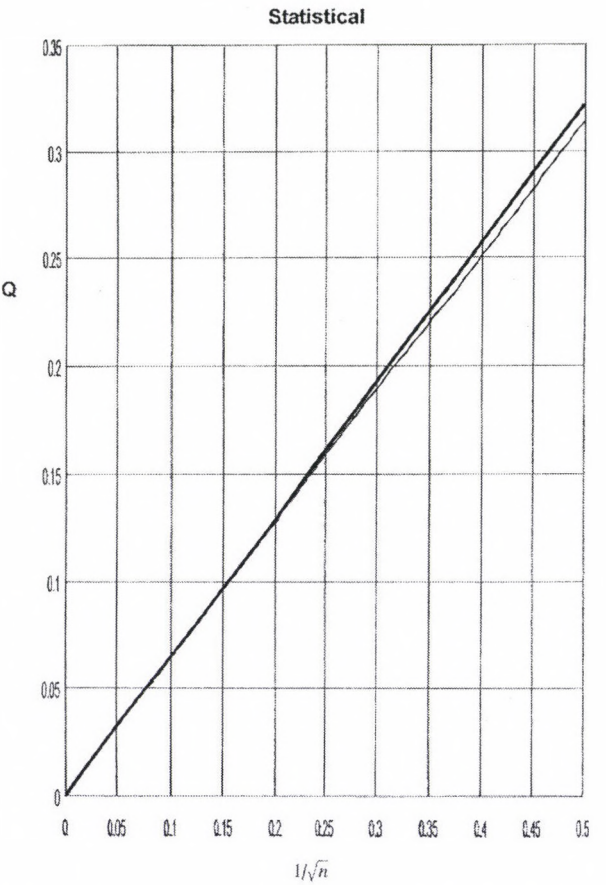


Fig. 4. Semi-intersextile ranges Q of sample medians vs. $x = 1/\sqrt{n}$ in case of the parent distribution of statistical type. The straight line represents the asymptotic rule

the corresponding $Q_{\mathrm{MED}}(1/\sqrt{n})$ curves are shown for different parent distributions in Figs 1 to 6. (A comment to Fig. 3: the Q-curve and the asymptotic straight line cannot be separated graphically for the Jeffreys-distribution: the greatest difference is only 1.3% for the smallest sample size $n = 4$.)

4. Discussion and conclusions

Figures 2–6 show that any correction is hardly needed for the often occurring parent distribution types: the asymptotic rule (see the straight lines in these figures) can be used in the overwhelming majority of cases of the statistical practice. Theoretical aspects, or other reasons, however, are eventually not satisfied with an "error of error" characterized by 5–10% percent; the first row in Table I (and Fig. 1) show that by the uniform type for $n = 4$ the difference is already about 17%, perhaps not acceptable even by the practicioners. (The really great differences occur if the tails are extremely heavy, i.e., the nearer is the $\underline{\underline{a}}$ type parameter to the unity the greater are the differences between the $Q_{\mathrm{MED}}(1/\sqrt{n})$ values and according

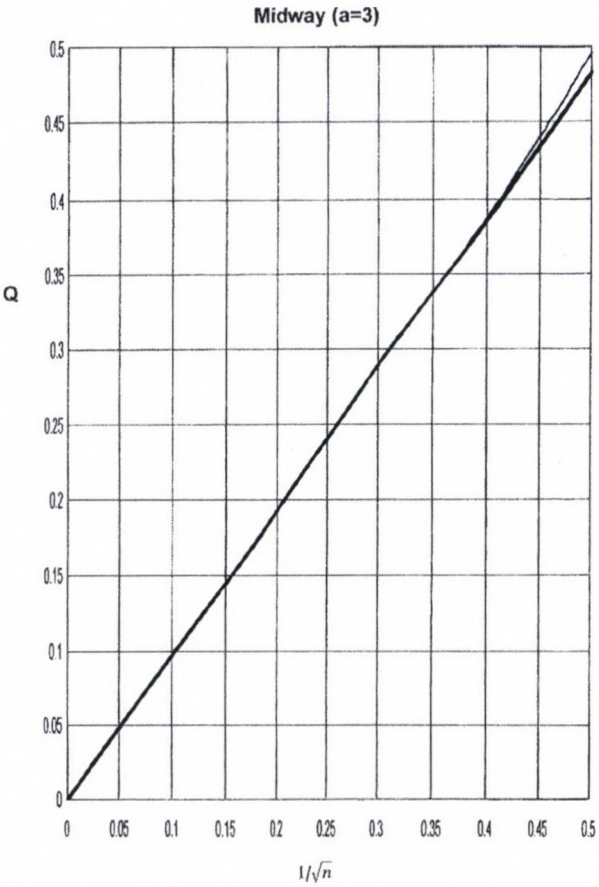


Fig. 5. Semi-intersextile ranges Q of sample medians vs. $x = 1/\sqrt{n}$ in case of the parent distribution of midway type. The straight line represents the asymptotic rule

to the "asymptotic rule" calculated expression $0.9674 \cdot A/\sqrt{n}$, see Eq. (16), and, in addition, this is true not only in case of extremely small samples. The authors will discuss such cases in a separate paper.)

To make for the reader incidentally necessary exact calculations easier, Eq. (17)-type polynoms are given for all six types of Table I in Table II, with simplifications in the notations: instead of $1/\sqrt{n}$ simply x, instead of $Q_{\text{MED}}(1/\sqrt{n})$ only $Q(n)$ is written.

Mathematicians perhaps are not sceptical concerning the results in Eqs (10), (13) and (14), practicioners are, however, inclined to make the "test of pudding", too, i.e., to carry out some Monte Carlo determination of the Q_{MED}-values. For this reason the uniform and the Cauchy types were chosen, see Table III, in the latter case to the sample sizes $n = 16$; 14; 12; 10; 8; 6 and 4, for the uniform type to sizes $n = 8$; 6 and 4. The Monte Carlo calculations — as it can be seen in Table III — were repeated five times (see the "+" signs in Figs 1 and 6) and the basis of every Monte Carlo calculation was a set containing 12000 samples.

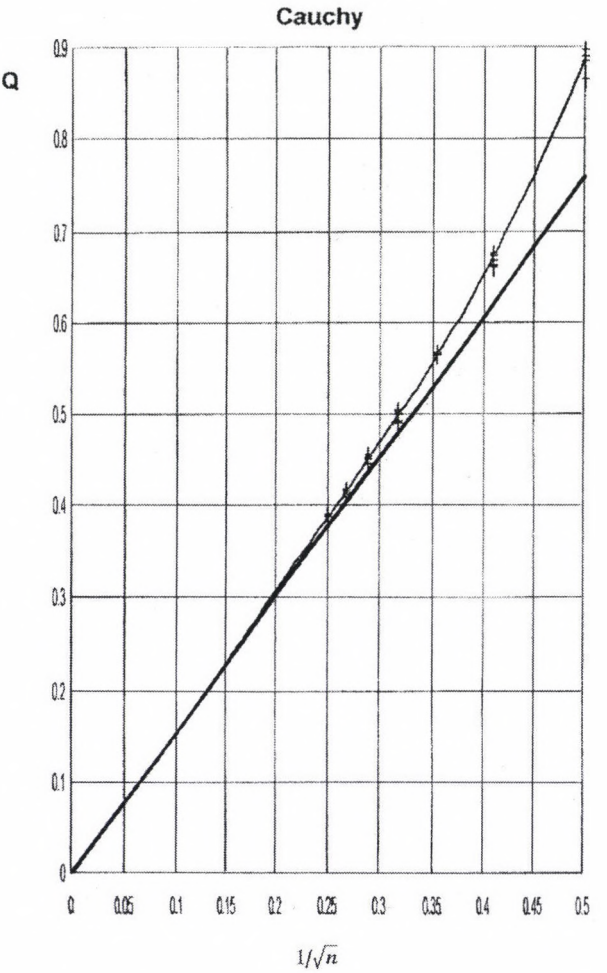


Fig. 6. Semi-intersextile ranges Q of sample medians vs. $x = 1/\sqrt{n}$ in case of the parent distribution of Cauchy type. The straight line represents the asymptotic rule. The "+" signs correspond to Monte Carlo results (see Section 4). The Monte Carlo results for $n = 10$, 12, 14 and 16 confirm that a fourth order polynomial is completely sufficient to describe the Q-curve vs. $x = 1/\sqrt{n}$

The middle value of the five-element $Q(n)$ groups can be compared also numerically with the analytical counterpart even if the reader finds Figs 1 and 6 convincing enough: in some respects this comparison is very instructive. The analytical values are given for six types to the sample sizes $n = 4$, 6 and 8 in Table I but Table III needs four surplus value, too: in case of the Cauchy type $Q(10) = 0.49791784$, $Q(12) = 0.45055228$, $Q(14) = 0.41478189$ and $Q(16) = 0.38649705$ holds.

It is perhaps not superfluous to give also numerically some middle values according to Table III in Table IV.

If one compares these values with the corresponding analytically calculated $Q_{\mathrm{MED}}(1/\sqrt{n})$-values given in Table I, the difference amounts only to some tenths of

Table III

Uniform		Cauchy	
$1/\sqrt{n}$	$Q(n)$	$1/\sqrt{n}$	$Q(n)$
0.3536	0.315264	0.25	0.387236
0.3536	0.318155	0.25	0.388651
0.3536	0.315858	0.25	0.388439
0.3536	0.315725	0.25	0.387619
0.3536	0.31002	0.25	0.390963
0.4082	0.359706	0.2673	0.417065
0.4082	0.358829	0.2673	0.41553
0.4082	0.356909	0.2673	0.41816
0.4082	0.355932	0.2673	0.412615
0.4082	0.354761	0.2673	0.409328
0.5	0.415749	0.2887	0.452006
0.5	0.413703	0.2887	0.450229
0.5	0.416052	0.2887	0.455326
0.5	0.42079	0.2887	0.449841
0.5	0.415842	0.2887	0.44616
		0.3162	0.504372
		0.3162	0.502156
		0.3162	0.488541
		0.3162	0.492833
		0.3162	0.500738
		0.3536	0.564529
		0.3536	0.56446
		0.3526	0.566981
		0.3536	0.564502
		0.3536	0.566294
		0.4082	0.675252
		0.4082	0.665149
		0.4082	0.668809
		0.4082	0.672996
		0.4082	0.661864
		0.5	0.886252
		0.5	0.884151
		0.5	0.898323
		0.5	0.88991
		0.5	0.865277

a percent. This circumstance, as a "test" of the analytical calculation, is evidently not astonishing, — the other aspect is, however, much more significant: it is proven that the Monte Carlo technics is accurate enough to determine $Q(n)$-curves for whatever estimates whitout analytical alternatives. The most frequent values are such estimates, consequently the authors shall use the above defined Monte Carlo technics in the near future to calculate $Q_{MFV}(1/\sqrt{n})$ values.

Table IV. Some Monte Carlo results for the semi-intersextile ranges of the MED values

Type	Sample size		
	$n = 4$	$n = 6$	$n = 8$
Cauchy	0.886252	0.668809	0.564529
uniform	0.415842	0.356909	0.315725

Acknowledgement

The present research work was supported by the Hungarian Ministry of Education (project number 0914/1997).

References

CramérH 1945: Mathematical Methods of Statistics. Almqvist and Wiksells, Uppsala

Huber P J 1981: Robust Statistics. Wiley, New York

Steiner F ed. 1997: Optimum Methods in Statistics. Akadémiai Kiadó, Budapest

Steiner F, Hajagos B 1998: *Acta Geod. Geoph. Hung.*, 33, 259–277.

Steiner F, Hajagos B. 1999a: *Acta Geod. Geoph. Hung.*, 34, 59–64.

Steiner F, Hajagos B 1999b: *Acta Geod. Geoph. Hung.*, 34, 65–69.

Steiner F, Hajagos B 2000: P-norm based statistical procedures are more efficient than L_1-based ones for all error-types of the complete supermodel $f_c(x)$. *Acta Geod. Geoph. Hung.* (present issue)

Vincze J 1968: Mathematical Statistics (in Hungarian). Műszaki Könyvkiadó, Budapest

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 295–304 (2000)

P-NORM BASED STATISTICAL PROCEDURES ARE MORE EFFICIENT THEN THE L_1-BASED ONES FOR ALL ERROR-TYPES OF THE COMPLETE SUPERMODEL $f_c(x)$

F STEINER[1] and B HAJAGOS[2]

[Manuscript received August 17, 1999]

The paper shows that P-norm based procedures are significantly more robust than the L_1-based ones. For these reason it was introduced the notion and the defining formulae for the indices of relative robustness. (These definitions are usable also generally, i.e., not only for P and L_1.)

Keywords: absolute efficiency; asymptotic case; asymptotic variances; norms of deviations; relative efficiency; relative robustness; small samples

1. The behaviour of the relative asymptotic efficiencies e_p/e_{L_1} for the whole supermodel $f_c(x)$

The supermodel $f_c(x)$ of all bell-shaped errors-distributions is defined by Steiner and Hajagos (1998) by the following formulae:

$$f_c(x) = \begin{cases} \dfrac{p^{1-1/p}}{2 \cdot \Gamma(1/p)} \cdot e^{-|x|^p/p} & (\infty > p \geq 2) \\ \dfrac{\Gamma(a/2)}{\sqrt{\pi} \cdot \Gamma[(a-1)/2]} \cdot (1+x^2)^{-a/2} & (\infty > a > 1) \end{cases} . \tag{1}$$

Figure 2c and the last columns of Table I p. 276 by Steiner and Hajagos (1998) show that the heaviness of the flanks of the error distributions continuously increases from vanishing probability to the maximum one.

The absolute asymptotic efficiency e_N of a given N norm is defined for an arbitrary $f(x)$ error distribution as

$$e_N = \frac{A^2_{\min}}{A^2_N} , \tag{2}$$

where $A^2_{\min}$ is the reciprocal value of the so-called Fisher-information depending only upon the $f(x)$. The efficiencies are given usually in percent.

We restrict ourselves for sake of simplicity to cases if the x_i data ($1 \leq i \leq n$) are direct measurement for a single quantity to be determined (denoted in general

[1]Geophysical Department, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary
[2]Institute of Mathematics, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary

1217-8977/2000/$ 5.00

by T). If the statistical algorithm is based upon the L_1-norm (in our case $L_1 = \frac{1}{n}\sum_{i=1}^{n}|x_i - T|$), the T-estimate is the sample median (a more sophisticated version of the latter is given by Steiner and Hajagos (1999b)). The asymptotic variance of this estimate is to be calculated in unimodal and to the origin symmetrical case according to the following simple formula:

$$A_{L_1}^2 = \frac{1}{4 \cdot f^2(0)} \,. \tag{3}$$

If the minimum place of the P-norm, i.e., of the expression

$$P = \left\{ \prod_{i=1}^{n} [(2\varepsilon)^2 + (x_i - Z)^2] \right\}^{1/2n} \tag{4}$$

is accepted as the best estimation of T — the latter is usually denoted as M in case of the P-norm (whilst the equation for the ε dihesion also

$$\varepsilon^2 = \frac{3 \cdot \sum_{i=1}^{n} \frac{(x_i - M)^2}{[\varepsilon^2 + (x_i - M)^2]^2}}{\sum_{i=1}^{n} \frac{1}{[\varepsilon^2 + (x_i - M)^2]^2}} \tag{5}$$

must be simultaneously fulfilled), the formula for the estimate (called most frequent value) is the following:

$$M = \frac{\sum_{i=1}^{n} \frac{x_i}{4\varepsilon^2 + (x_i - M)^2}}{\sum_{i=1}^{n} \frac{1}{4\varepsilon^2 + (x_i - M)^2}} \,. \tag{6}$$

The expression of the asymptotic variance of this estimate (it is justified to denote it by A_P^2) is evidently more complicated than Eq. (3):

$$A_P^2 = \frac{\int_{-\infty}^{\infty} \frac{(x - M)^2}{[4\varepsilon^2 + (x - M)^2]^2} f(x)dx}{\left[\int_{-\infty}^{\infty} \frac{4\varepsilon^2 - (x - M)^2}{[4\varepsilon^2 + (x - M)^2]^2} f(x)dx \right]^2} \tag{7}$$

(see Eq. (5-90) by Steiner (1990)). In Eq. (7) the M-value fulfils the integral counterpart of Eq. (6):

$$M = \frac{\int_{-\infty}^{\infty} \frac{x}{4\varepsilon^2 + (x - M)^2} f(x)dx}{\int_{-\infty}^{\infty} \frac{1}{4\varepsilon^2 + (x - M)^2} f(x)dx} \,. \tag{8}$$

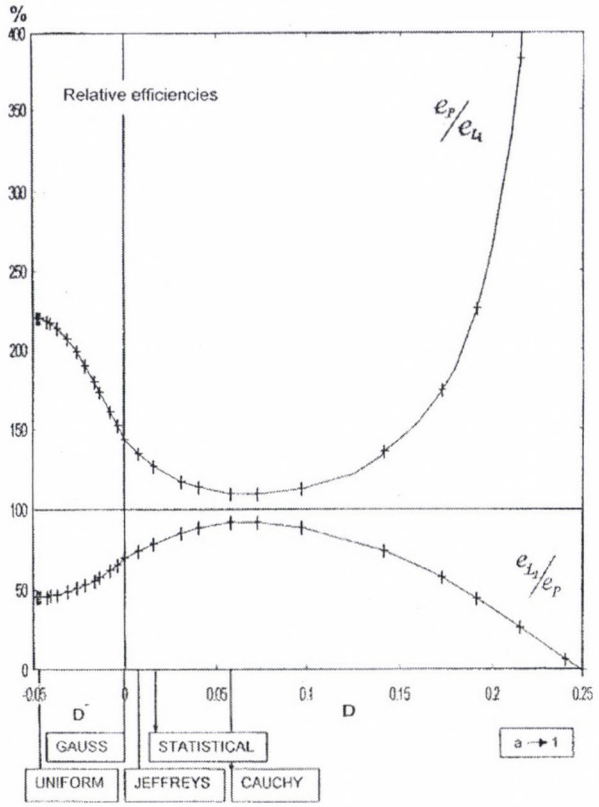


Fig. 1. Relative efficiency curves e_{L_1}/e_P vs. the type distance from the Gaussian. In case of heavier tails than the Gaussian, this distance is denoted by D, in opposite case by D^-

Concerning the $A_{\min}$-values, Eq. (1) clearly shows that there are two different formulae needed, if we would calculate absolute asymptotic efficiencies according to Eq. (2). In this paper, however, the goal is only the comparison of two absolute asymptotic efficiencies using the notion of the relative efficiency, i.e., the quotient of them:

$$e_P/e_{L_1} = \frac{A^2_{\min}}{A^2_P} \cdot \frac{A^2_{L_1}}{A^2_{\min}} = \frac{A^2_{L_1}}{A^2_P} \tag{9}$$

and therefore $A_{\min}$-values are not needed.

In Fig. 1 both the e_P/e_{L_1} and e_{L_1}/e_P relative efficiency curves are shown as $e_P/e_{L_1} \to \infty$ if $D \to 0.25$. For all types of the complete $f_c(x)$ supermodel the relation $e_P > e_{L_1}$ is fulfilled. For some numerical values see Table I where besides the type parameters and relative asymptotic efficiencies also the values of the ε dihesions are given (as surplus information). One column of Table I contains the actual type-distances D^- and D from the Gaussian type (D^- in direction of the diminishing, D in direction of the increasing heavinesses of the tails; concerning the correct and general definition of type-distances see e.g. Eq. (1–2) by Steiner (1997).

Table I

Type parameters of the supermodel $f_c(x)$	Type distances from the Gaussian type (D^- and D)	e_P/e_{L_1} (%)	e_{L_1}/e_P (%)	The value of the dihesion ε in standard case
	D^-			
$p = \infty$	0.04804	219.92	45.47	0.7185223
$p = 200$	0.04745	219.90	45.48	0.7356970
$p = 100$	0.04686	219.85	45.49	0.7481034
$p = 50$	0.04569	219.64	45.53	0.7682812
$p = 20$	0.04222	218.28	45.82	0.8122672
$p = 15$	0.04031	217.02	46.08	0.8307641
$p = 10$	0.03656	213.70	46.79	0.8600029
$p = 20/3$	0.03119	207.00	48.31	0.8907265
$p = 5$	0.02584	198.85	50.29	0.9108802
$p = 4$	0.02088	189.82	52.68	0.9236502
$p = 10/3$	0.01619	180.37	55.44	0.9309372
$p = 3$	0.01322	173.99	57.48	0.9333825
$p = 5/2$	0.00762	161.37	61.97	0.9336774
$p = 20/9$	0.00370	152.20	65.70	0.9306420
$p = 2$	0.00000	143.41	69.73	0.9253681
	D			
$a = 9$	0.00804	134.26	74.48	0.3346663
$a = 5$	0.01601	126.96	78.77	0.4817797
$a = 3$	0.03130	117.24	85.30	0.6972796
$a = 2.5$	0.04082	113.41	88.17	0.8119323
$a = 2$	0.05809	109.66	91.19	1.0000000
$a = 1.75$	0.07309	109.04	91.71	1.1502030
$a = 1.5$	0.09746	112.46	88.92	1.3776428
$a = 1.25$	0.14249	135.60	73.74	1.7665380
$a = 1.15$	0.17225	174.20	57.41	2.0133731
$a = 1.1$	0.19150	225.38	44.37	2.1708114
$a = 1.05$	0.21536	383.16	26.10	2.3601228
$a = 1.01$	0.24077	1661.6	6.2	2.5415764
$a \to 1$	0.25000	∞	0.0	2.592

2. Index of the asymptotic relative robustness

The superiority of the P-norm procedures over the L_1-norm ones is characteristically visualized in Fig. 1 (and given numerically in Table I): not only the qualitative relationship $e_P > e_{L_1}$ is fulfilled for all types of the complete f_c(x) supermodel but also $e_P > e_{L_1}$-values amounting to several hundred percents are occurring in both ends of the abscissa. The left end corresponds to the uniform distribution type characterised by the relative efficiency value 220%, i.e., in this case 120% surplus data are needed using the L_1 norm instead of the P-norm if we would achieve same accuracy; this surplus value is even in the Gaussian case far not negligible: it amounts to 43%. This rate of surplus data needed can be, however, several hundred percents, too, see e.g. the type $a = 1.05$ in Table I; here 283% surplus data were needed using L_1 instead of P. These values (even the 43% characterizing the Gaussian case) enhance the economic meaning of the relative efficiencies.

In the last example ($a = 1.05$) the tails are too heavy: such types occur very rarely. It would be therefore misleading to define an index of relative robustness analogously to the general robustness (see e.g. Section 7.2 by Steiner (1997)); we would get perhaps the not informative infinite value "characterizing" mostly the rarely occurring error types of very heavy tails. We must take therefore into consideration defining robustness the occurrence probability densities of often occurring types in the practice of statistics and geostatistics. According to modern statisticians practically only tails greater than the Gaussian tail can be expected: e.g., according to Dutter (1986/1987) the type belonging to the type parameter $a = 5$ can be most commonly expected in geosciences. Introducing the notation $t = 1/(a-1)$ and taking into consideration conclusions and comments of other authors, too, the probability densities of the occurrence of error types in the practice can be given by the following formula:

$$f(t) = 16 \cdot t \cdot e^{-4t} \tag{10}$$

(see the $f_D(t)$-curve in Fig. 7.4 by Steiner (1997)). The maximum corresponds to $a = 5$, the Cauchy-type shows not vanishing probability density — about 20% of the maximum value — but $f(2)$ belonging to $a = 1.5$ is practically zero).

As the relative efficiencies e_P/e_{L_1} can be evidently also expressed in function of t in the D-domain denoted e.g. by $e_{\mathrm{rel}}(P/L_1; t)$, the index of the asymptotic relative robustness (denoted by $e_{\mathrm{rel}}(P/L_1)$) can be defined (and in the same time calculated) according the following equation:

$$r_{\mathrm{rel}}(P/L_1) = \int_0^\infty e_{\mathrm{rel}}(P/L_1, t) \cdot f(t)dt \tag{11}$$

and this results in the value 121%, i.e., in a significantly higher value than 100%.

In special cases (e.g. making a series of measurements in a laboratory), the error types can occur in the vicinity of the Jeffreys distribution characterized by $a = 9$. In this case instead of $f(t)$ in Eq. (10)

$$\overline{f}(t) = 64 \cdot t \cdot e^{-8t} \tag{12}$$

is to be used and using this weight function the $\overline{r}_{\mathrm{rel}}(P/L_1)$ index of the asymptotic relative robustness will be the following expression:

$$\overline{r}_{\mathrm{rel}}(P/L_1) = \int_0^\infty e_{\mathrm{rel}}(P/L_1; t) \cdot \overline{f}(t)dt \tag{13}$$

resulting in the value 128.5%.

It is perhaps not superfluous to remark that substituting an arbitrary N_1 norm instead of P and a similarly arbitrary norm N_2 instead of L_1, Eqs (10) and (13) are general definitions of the asymptotic relative robustnesses, — in this paper, however, only the norm P is compared with the norm L_1.

3. Index of the relative robustness in case of finite samples

Huber (1981) proved that under rather general conditions the type of the statistical estimates converges to the Gaussian type if $\underline{n}$ (the number of the elements x_i contained in the sample) tends to infinity. As a consequence, the scatter of the estimates can be calculated as $A/\sqrt{n}$ (where the A asymptotic scatter depends both upon the used statistical algorithm and the parent distribution of the x_i data,) if n is large enough.

Steiner and Hajagos (1999a) have shown that for many parent distributions the values of efficiencies can significantly differ from the asymptotic one in case of $n = 5$. In the present paper in Sections 1 and 2 the asymptotic case was only treated; it arises the question if the constated and by numerical values (e.g. by indices of the relative robustnesses) characterised superiority of the P norm over the L_1 one is only valid for the case $n \to \infty$ — and perhaps the opposite situation characterizes the small n-values? The answer to this question is fortunately a definite not —, but we feel ourselves to be obliged to deal with this question a somewhat more deeply.

All characteristics of the dispersion of a given error distribution are proportional to the scale parameter S. Consequently, the quotient of two arbitrary dispersion characteristics of each parent distribution type is a well defined number. For example, the semi-intersextile range Q (see e.g. Fig. P.2 by Steiner (1997)) can be expressed as $0.9674 \cdot \sigma$ (σ means the scatter; see the last column of Table 1.1 by Steiner (1990)). It should be mentioned that σ is at the same time the parameter of scale (S) of the Gaussian distribution.

Using Monte Carlo method, and generating N times samples of $\underline{n}$ elements on ground of a chosen parent distribution, our first guess would perhaps be to calculate the variance of N estimates got by minimizing the L_1 norm, on one hand, and on the other hand, the variance of N estimates got by minimizing the P norm: according to Eq. (9) the quotient of these two quantities would give the relative efficiency for the given finite $\underline{n}$ (and for the chosen parent distribution).

The calculation of variances, however, have its drawbacks, see e.g. the introduction of Appendix IX by Steiner (1997). It is more convenient to determine the empirical semi-intersextile ranges of the N estimates both in case of the L_1 norm (where the estimates are sample medians) and in case of the P norm (where the estimates of the same N samples are most frequent values denoted by M). According to the already mentioned proportionality, for arbitrary value of $\underline{n}$ the relative efficiency can be calculated according to the followings:

$$e_P/e_{L_1} = Q^2_{\mathrm{MED}}/Q^2_M \,. \tag{14}$$

The repetition number was chosen as $N = 12000$ and the sample medians were calculated according to the formula given by Steiner and Hajagos (1999b). (The whole Monte Carlo calculation was carried out 5 times.) We have decided to calculate for the smallest possible sample characterized by $n = 4$; if the quotients calculated according to Eq. (14) for the most frequently occurring parent distributions are greater than 100%, the qualitative conclusions got for the asymptotic case in Sections 1 and 2 concerning to the superiority of the P norm over L_1 remain valid for finite samples, too.

Table II. Values of the relative efficiencies e_P/e_{L_1} for some parent distributions for the asymptotic case ($n = \infty$) and for two variants of the case $n = 4$. (The second variant is treated below in Section 4)

D (type distance from the Gaussian type)	a (type-parameter, see Eq. (1))	t $\left(t = \frac{1}{a-1}\right)$	$n = \infty$	$n = 4$ (standard procedure for M with double iteration)	$n = 4$ (M is calculated according to Eq. (6) but with *a priori* known ε-value)
0	∞	0	143%	121%	129%
0.00804	9	0.125	134%	118%	124%
0.01601	5	0.25	127%	115%	119%
0.03130	3	0.5	117%	109%	116%
0.05809	2	1	110%	107%	130%
0.09746	1.5	2	112%	102%	228%

Table III. Indices of relative robustness

	$n = \infty$	$n = 4$ (standard case)	$n = 4$ (a priori known ε)
$r_{\text{rel}}(P/L_1)$	120.9%	115%	121%
$\bar{r}_{\text{rel}}(P/L_1)$	128.5%	111%	122%

Table II shows both for the asymptotic case and for $n = 4$ the e_P/e_{L_1} values being all greater than 100% for the parent distributions characterized by the t-values 0; 1/8; 1/4; 1/2; 1 (and also 2) which define the curve of the relative efficiencies figuring in the formulae for calculating the indices of robustness (see Eqs (11) and (13)). It seems to be informative to show all indices of robustness in a single Table III:

The superiority of P over L_1 is not only for $n = \infty$ but also generally valid.

4. The special case of "*a priori* known ε"

The last columns of both Table II and III belong to a special and up to the present not treated case: if ε is *a priori* known. This situation is perhaps rare but one must not exclude: e.g., if the cause of error lyes in our well known measuring instrument, ε can be a well defined value.

The "*a priori* known ε" however, needs not be by all means a "well defined value of ε". As a first step to verify this statement, we modify Eq. (7) for arbitrary S scale parameter instead of 2ε:

$$A^2 = \frac{\int_{-\infty}^{\infty} \frac{(x-MFV)^2}{[S^2+(x-MFV)^2]^2} f(x)dx}{\left[\int_{-\infty}^{\infty} \frac{S^2-(x-MFV)^2}{[S^2+(x-MFV)^2]^2} f(x)dx\right]^2}, \tag{15}$$

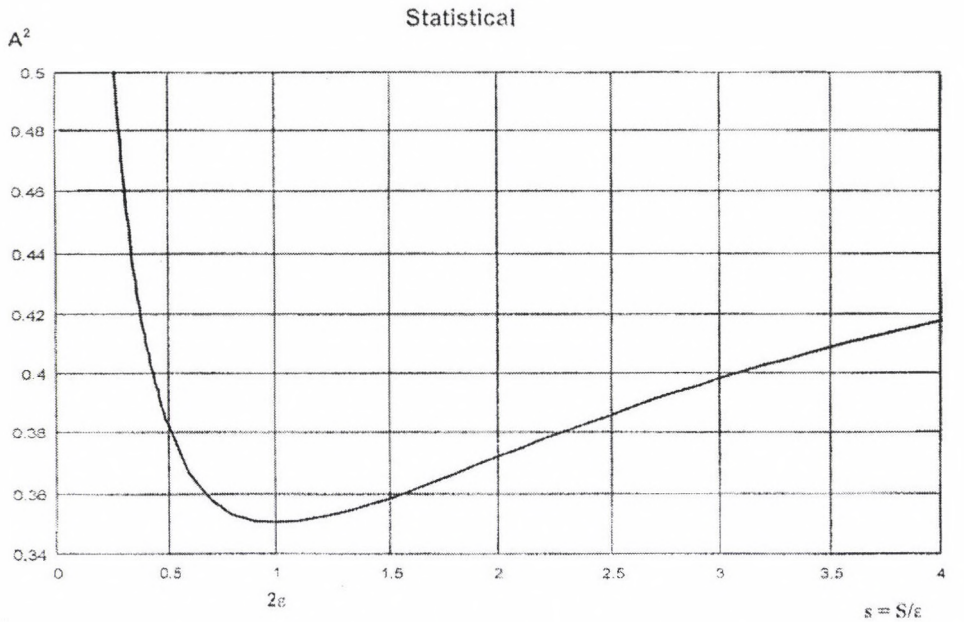


Fig. 2. The curve of the asymptotic variance A^2 according to Eq. (15) for the statistical distribution-type $a = 5$ vs. $s = S/2\varepsilon$, where S means the actually applied scale parameter. The minimum place is at $s = 1$ (i.e., at $S = 2\varepsilon$) but because of the flat minimum domain even at the twice value of the correct S (or s) the increase of A^2 (figuring in the efficiency formula) is only 5.7% — and 2% if the "*a piori* known ε" is by 50% greater as the theoretical value of the dihesion (see the row "$a = 5$" in Table IV of the asymptotic variances). Taking only the half value, the increase in A^2 is 10%

MFV means the most frequent value, i.e., M in case of $k = 2$ and M_C if $k = 1$. Figures 2 and 3, and their captions verifies that a rough estimation is enough.

The caption of Fig. 2 refers to Table IV, where not only the characteristics of the geostatistical ($a = 5$) type are given.

For other types the minimum place of the A^2-values are naturally not any more at $S = 2\varepsilon$, see e.g. the rows for $a = 3$ (midway-type) and for $a = 9$ (Jeffreys-type), — the Table IV, however, belongs to the standard P-norm using always $S = 2\varepsilon$ in the formula Eq. (15): this is in every row "the 100%".

If we have some information concerning the actual error distribution type, we can choose instead of $k = 2$ the best k using Table V. (As a curiosity, it should be remarked that the long ago proposed $k = 1/2$ in case of very heavy tails is discovered to belong actually to $a = 1.15$.)

Because of the great robustness of the P_k-norms, it is enough to use the following k-values: 3; 2; 1 and 1/2. At the end of this paper in the authors show Fig. 3 for the P_C-norm ($k = 1$) an analogous curve showed in Fig. 2 for $k = 2$ (P norm) as in the practice often the minimization of the P_C-norm is the statistical background of an inversion or regression algorithm.

It is perhaps not superfluous to make two concluding remarks:

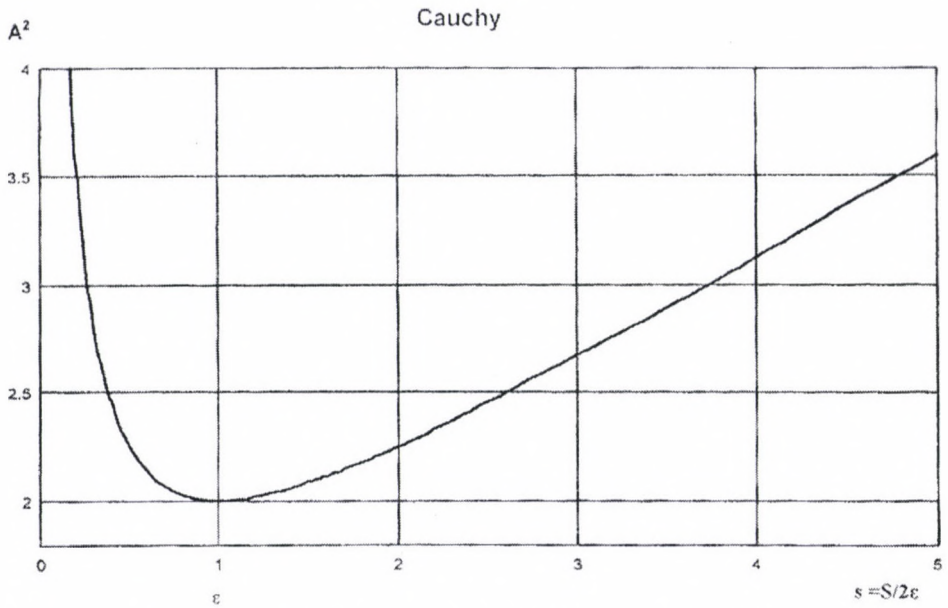


Fig. 3. The curve of the asymptotic variance A^2 for the Cauchy distribution. The conclusions are the same as in case of Fig. 2: an "*a priori* known ε" is practically not strictly equal with the theoretical value of the dihesion (a rough estimate is satisfying)

A) The case of an "*a priori* known ε" can be neglected if $\underline{n}$ is large enough to get a reliable ε-value according to Eq. (5).

B) The superiority of the most frequent procedures over the L_1 ones was demonstrated earlier in respect of the resistance (i.e., in respect of the outlier-insensitivity). Two examples from Steiner (1997):

   I. Figure 8.15 on page 176 where T means the distortion caused by asymmetric lying outliers for every demonstrated occurrence probability p of outliers the median-distortions are significantly greater than the most frequent value (MFV) distortions.

   II. Figure 10.13 on page 246 shows that a great amount of asymmetrically lying outliers can fully breakdown L_1-results, while for the same case the error using the P norm amounts less than 10%.

Acknowledgement

The present research work was supported by the Hungarian Ministry of Education (project number 0914/1997).

Table IV

Type	ε	$A^2(\varepsilon)$	$A^2(4/3\varepsilon)$	$A^2(2\varepsilon)$	$A^2(3\varepsilon)$	$A^2(4\varepsilon)$
$a = 1.15$	2.01337	20.8911 67.7%	23.8529 77.3%	30.8686 100%	43.0461 139.4%	56.7026 183.7%
$a = 1.5$	1.37760	4.79579 78.4%	5.14827 84.2%	6.113355 100%	7.84202 128.3%	9.75350 159.5%
$a = 2$ (Cauchy)	1.0000	2.00000 88.9%	2.04167 90.7%	2.25000 100%	2.66667 118.5%	3.12500 138.9%
$a = 3$ (midway)	0.69728	0.857748 100.6%	0.834303 97.8%	0.852980 100%	0.928614 108.9%	1.01619 119.1%
$a = 5$ (statist)	0.48178	0.386126 110.3%	0.362364 103.5%	0.350078 100%	0.356945 102.0%	0.370169 105.7%
$a = 9$ (Jeffreys)	0.33467	0.181501 116.6%	0.166668 107.1%	0.155652 100%	0.152780 98.2%	0.153808 98.8%
$p = 1$ (Laplace)	0.80648	1.160338 92.0%	1.188078 94.2%	1.261174 100%	1.371311 108.7%	1.466957 116.3%
$p = 2$ (Gauss)	0.92537	1.346384 122.9%	1.216102 111.0%	1.095330 100%	1.035652 94.6%	1.015862 92.7%
$p = \infty$ (uniform)	0.71852	0.758139 166.7%	0.589869 129.7%	0.454721 100%	0.389075 85.6%	0.365079 80.3%

Table V

a	ε	k	$A_{\min}(k\varepsilon)$
1.15	2.01337	0.4972	4.27327
1.25	1.76654	0.5661	3.22490
1.50	1.37762	0.7269	2.16025
2.00	1.00000	1.0000	1.41421
3.00	0.69728	1.4342	0.91287
5.00	0.48178	2.0756	0.59161
9.00	0.33467	2.9879	0.39087

References

Dutter R 1986/1987: Mathematische Methoden in der Montangeologie. Vorlesungsnotizen, Manuscript, Leoben

Huber P J 1981: Robust Statistics. Wiley, New York

Steiner F 1990: Introduction to Geostatistics (in Hungarian). Tankönyvkiadó, Budapest

Steiner F ed. 1997: Optimum Methods in Statistics. Akadémiai Kiadó, Budapest

Steiner F, Hajagos B 1998: *Acta Geod. Geoph. Hung.*, 33, 259–277.

Steiner F, Hajagos B 1999a: *Acta Geod. Geoph. Hung.*, 34, 65–69.

Steiner F, Hajagos B 1999b: *Acta Geod. Geoph. Hung.*, 34, 59–64.

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 305–312 (2000)

RELATIVE EFFICIENCY OF THE CONVENTIONAL STATISTICS COMPARED TO THE L_1-NORM BASED STATISTICAL PROCEDURES, TESTED ON DISTRIBUTION TYPES OF THE GENERALIZED GAUSSIAN SUPERMODEL

F STEINER[1] and B HAJAGOS[2]

[Manuscript received January 5, 2000]

In the paper there are given the error type-domains where the conventional statistics works with greater efficiency than the L_1y-based statistics does, and vice versa. It turned out that the frontier of the two domains corresponds to the $f_a(x)$-type characterized by $a = 5.56$, i.e., near lying to the statistical (geostatistical) type and in the same time to the upper limit of the Jeffreys interval of types. — In the paper not only the asymptotical case is treated but also the case of small samples.

Keywords: generalized Gaussian supermodel; L_1-norm; L_2-norm; relative efficiency; relative efficiencies for small samples

1. Introduction

The drawbacks of the conventional, i.e., of the L_2-norm based procedures force more and more the practicioners in many disciplines (e.g. in geophysics) to work statistically on the L_1-norm instead of the conventional statistics. The use of the L_1-norm is namely the simplest alternative in the modern statistics but sufficient for solving many tasks. (The P-norm has efficiency advantages even compared to the L_1-norm, see Steiner and Hajagos (2000), the P-algorithms, however, are more difficult than the L_1 ones.)

In guaranteed outlier-free cases and by distributions characterized by short tail-domains, e.g., in the very neighbourhood of the Gaussian type, the conventional statistics works surely better (with greater statistical efficiency) than the L_1-norm does; the latter works best if the errors are Laplace distributed. For testing purposes therefore it seems to be the best to accept the "generalized Gaussian" supermodel (see e.g. Tarantola 1987, p. 26) defined by the following equation for density functions in the standard case ($T = 0$, $S = 1$):

$$f_p(x) = \frac{p^{1-1/p}}{1 \cdot \Gamma(1/p)} \cdot e^{-|x|^p/p} \qquad (p > 0)\,. \tag{1}$$

For $p = 2$ Eq. (1) gives evidently the standard Gaussian density $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, and similarly for $p = 1$ the density formula for the Laplace type $\frac{1}{2} e^{-|x|}$, i.e., the generalized Gaussian super-model really contains the eigendistributions both for the L_2 and the L_1 norms.

[1]Geophysical Department, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary
[2]Institute of Mathematics, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary

1217-8977/2000/$ 5.00

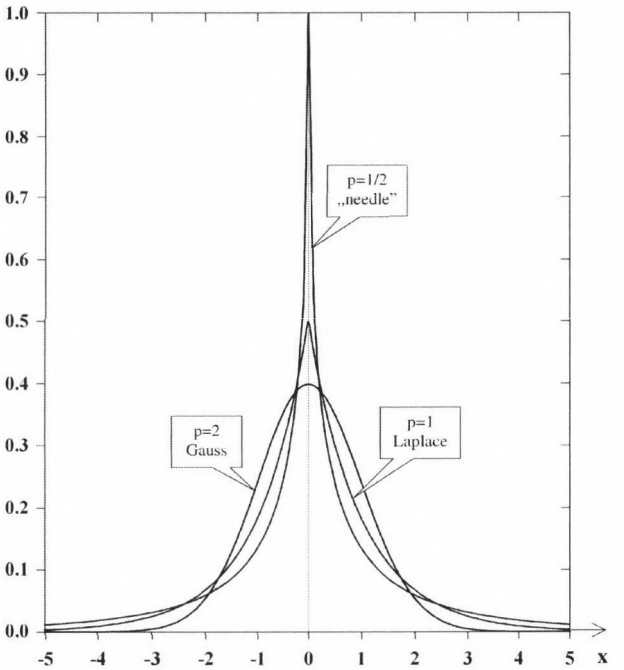


Fig. 1. The probability density curves for the Gaussian, Laplace and "neddle" distributions, i.e., $f_p(x)$-curves (see Eq. 1) for the type parameters $p = 2$, $p = 1$ and $p = 1/2$

In case of $p \to \infty$ $f_p(x)$ tends to $f_u(x)$, i.e., to the standard uniform distribution (having a value differing from zero only in the interval $-1 \leq x \leq +1$, namely the constant $f_u(x)$-value $1/2$), on the one hand, and the relative behaviour of the investigated norms can be of interest also for this distribution type. On the other hand, however, if $p \to 0$, the density curves in the symmetry point $x = 0$ became more and more (in limit infinitely) peaky, and these cases can be regarded from a practical point of view as not interesting. Figure 1 shows that even at $p = 1/2$ the $f_p(x)$ type can be justifiedly named "needle-distribution" (this name was used already by Steiner 1997 in Fig. 9.14); for comparison the Laplace and Gaussian density functions are also demonstrated in the same Fig. 1. The conclusion is the following: the treatments of the present paper are enough to carry out for the error type-parameter interval $0.5 \leq p \leq \infty$.

2. The asymptotic relative statistical efficiencies $e(L_2/L_1)$ and $e(L_1/L_2)$

The absolute statistical efficiency $\underline{e}$ of a statistical algorithm to determine (to estimate) the location parameter of a probability distribution is defined by the following quotient:

$$e = \frac{A^2_{\min}}{A^2}\,, \tag{2}$$

where A^2 is the asymptotic variance of the just used algorithm and $A^2_{\min}$ is — according to the Cramér-Rao inequality — the asymptotic variance of that algorithm, which results in the most accurate estimate of the location parameter in case of the actually occurring $f(x)$ density function. If the statistical procedure is based upon the minimization of a well defined norm N of the deviations X_i (i.e., of the differences of the i-th measured and calculated values), the efficiency can be written as $e(N)$ or as e_N.

Two statistical procedures based on the norms N_1 and N_2 can be compared for a given $f(x)$ parent distribution calculating the quotient $e(N_1)/e(N_2)$ resulting in the so-called relative asymptotic efficiency; for the sake of simplicity let's be its notation $e(N_1/N_2)$. (In the relative efficiency the variance $A^2_{\min}$ obviously plays not any more a role being

$$e(N_1/N_2) = \frac{A_2^2}{A_1^2}\,, \tag{3}$$

where the asymptotic variances minimize the N_1 and the N_2 norm of the deviations.) If $N_1 = P$ and $N_2 = L_1$ (see Table P.2 on page 20 by Steiner 1997 where the definitions of ten norms are given), Steiner and Hajagos (2000) have proven that the relation $e(P/L_1) > 100\%$ is valid for all types of the complete supermodel $f_c(x)$ defined by Steiner and Hajagos (1998). As the application of P-based algorithms are a little bit more complicated then that of the L_1-based procedures, at the present time the L_1-norm is usually accepted if a practitioner will not any more use the L_2-based classical statistics with its drawbacks: these algorithms are namely heavily influenced by outliers, and the statistical efficiency $e(L_2)$ drops quickly from 100% to lower values if the type-distance from the Gaussian type increases. In outlier-free and by short tails characterized cases, however, the classical statistics can be used more advantageously than the L_1-based procedures. Supposing that outliers are not present, the question can be posed as follows: for which p-interval is for the relative efficiency the inequality $e(L_2/L_1) > 100\%$ valid?

In the overwhelming majority of cases the location parameter is estimated (see Andrews et al. 1972). The arithmetic mean of the sample elements (denoted by $\overline{x}$) minimizes the L_2-norm of the deviations (i.e., the expression $\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$) and the sample median (*med*) minimizes the L_1-norm of them (i.e., med is the minimum place of the expression $\frac{1}{n}\sum_{i=1}^{n} |X_i|$). The asymptotic variances $A^2_{\overline{x}}$ and A^2_{med} can be calculated for $f_p(x)$-distributions as follows:

$$A^2_{\overline{x}} = P^{\frac{2}{p}} \frac{\Gamma(3/p)}{\Gamma(1/p)} \tag{4}$$

and

$$A^2_{\text{med}} = \frac{1}{4 \cdot f_p^2(0)} = \frac{p^{\frac{2}{p}} \cdot \Gamma^2(1/p)}{p^2} \tag{5}$$

(see e.g. by Steiner (1990) the second and first rows of the Table 5.6, taking naturally also Eq. (1) into consideration). According to Eqs (3), (4) and (5) the relative

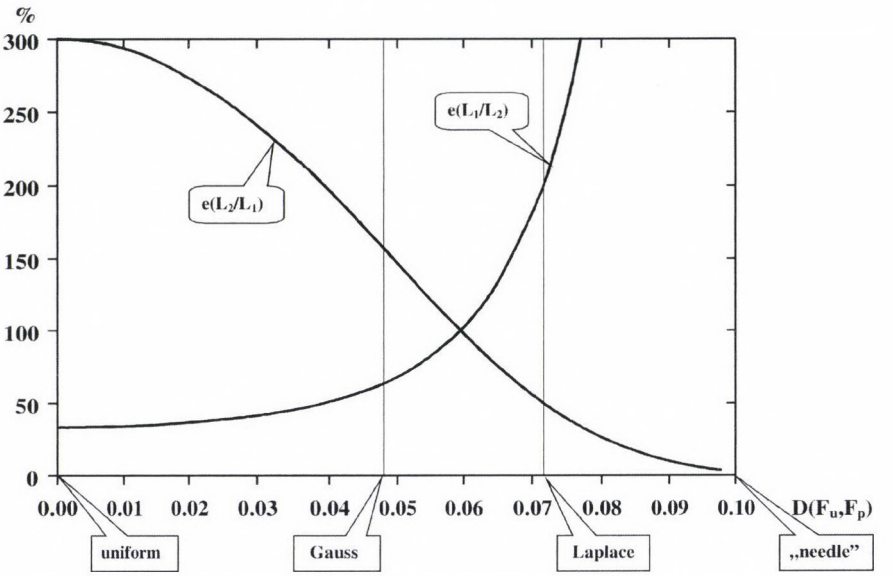


Fig. 2. Relative efficiency curves $e(L_2/L_1)$ and $e(L_1/L_2)$ in asymptotic case (i.e., if $n \to \infty$) for the types of the generalized Gaussian supermodel vs. the type-distance $D(F_u, F_p)$. The intersection point is at $p = 1.407$ (see the third and fourth columns of Table I) and this corresponds best to the upper end of the Jeffreys-interval

efficiency $e(L_2/L_1)$ in asymptotic sense is to be calculated simply as

$$e(L_2/L_1) = \frac{\Gamma^3(1/p)}{p^2 \cdot \Gamma(3/p)}\,, \tag{6}$$

the reciprocal of this expression gives evidently the relative efficiency $e(L_2/L_1)$. The curves of both quantities vs. the type-distance from the uniform type are presented in Fig. 2. The corresponding numerical values are given in the columns 3 and 4 in Table I, where the values of the type-parameter p and the type-distances are given in the second one in the first column $D(F_u, F_p)$ calculated according to

$$D(F_u, F_p) = \min_S \left\{ \max_x |F_p(0, S; x) - F_u(0, 1; x)| \right\} \tag{7}$$

(see e.g. Eq. (1-4) on page 44 by Steiner 1997).

The intersection point of both curves in Fig. 2 gives the upper limit as far as the conventional statistics works asymptotically better than the L_1-based procedures. What the accurate value belongs, Table I shows that the limit-type is characterized by the parameter value $p = 1.407$. In the geophysical practice (and even in its theory), however, the use of the $f_a(x)$ supermodel defined by the following standard densities

$$f_a(x) = \frac{\Gamma[a/2]}{\sqrt{\pi} \cdot \Gamma[(a-1)/2]} (1 + x^2)^{-a/2} \qquad (a > 1) \tag{8}$$

Table I

		$e(L_2/L_1,n)$ and $e(L_1/L_2,n)$ in percent							
p	$D(F_u,F_p)$	$n=\infty$		$n=7$		$n=5$		$n=3$	
1	2	3	4	5	6	7	8	9	10
100	0.00159	299.86	33.35	254.54	39.29	238.92	41.86	205.04	48.77
50	0.00311	299.43	33.40	254.20	39.34	238.61	41.91	204.59	48.88
20	0.00742	296.64	33.71	251.96	39.69	236.58	42.27	204.11	48.99
10	0.01388	287.82	34.74	244.83	40.84	230.02	43.47	199.45	50.14
5	0.02479	259.89	38.74	222.34	44.98	209.51	47.73	183.52	54.49
3	0.03647	213.62	46.81	186.88	53.51	177.97	56.19	159.79	62.58
2	0.04804	157.08	63.66	146.61	68.21	142.61	70.12	133.76	74.76
1.75	0.05226	135.64	73.72	131.83	75.86	129.59	77.16	124.12	80.57
1.5	0.05735	110.35	90.62	114.41	87.86	114.16	87.60	112.55	88.85
1.407	0.05960	99.95	100.05	107.18	93.30	107.71	92.84	107.66	92.89
1.25	0.06364	81.30	122.99	93.99	106.40	96.10	104.06	101.47	98.55
1.0	0.07168	50.00	200.00	70.45	141.94	74.38	134.45	81.47	122.75
0.9	0.07558	37.74	265.00	60.28	165.90	64.88	154.12	73.62	135.84
0.8	0.07999	26.31	380.13	49.82	200.74	54.95	181.99	65.12	153.56
0.75	0.08243	21.10	473.97	44.54	224.54	49.85	200.61	60.62	164.96
0.667	0.08691	13.47	742.57	35.78	279.50	41.23	242.56	51.61	193.76
0.55	0.09428	5.58	1792.97	23.98	417.06	29.18	342.76	41.02	243.80
0.5	0.09794	3.33	3000.00	19.26	519.19	24.16	413.94	35.75	279.76

ensures significant advantages, consequently the following question must be answered: for which $\underline{a}$-value reaches the type-distance $D(F_a, F_{p=1.407})$ its minimum value? The answer is $a = 5.56$ (being the minimum value only 0.003666). It can be also of interest to compare the type-distances of both each other corresponding error-distribution types from the uniform distribution. The following equations hold: $D(F_{a=5.56}, F_u) = 0.0595504$ and $D(F_{p=1.407}, F_u) = 0.0595264$. Consequently, the $f_a(x)$-type in case of $a = 5.56$ can be really accepted as limit type as far as the conventional statistics work asymptotically better than the L_1-based procedures. This type-domain contains the whole Jeffreys-interval defined by $10 \geq a \geq 6$ and the limit type ($a = 5.56$) is already near the (geo)statistical distribution characterized by $a = 5$; this type often occurs according to Dutter 1986/1987 in geostatistics.

3. Relative statistical efficiencies $e(L_1/L_1,n)$ and $e(L_1/L_2,n)$ for finite samples

In case of finite samples we cannot any more operate with asymptotic error-characteristics as it was done in Eq. (3) using A_1 and A_2. Instead of them the error of the estimates for the location parameter, i.e., $\overline{x}$ (arithmetic mean minimizing the norm L_2) and *med* (sample median in case of L_1) should be characterized by some of their semi-interquantile ranges, depending naturally upon the $\underline{n}$ sample size. From among the possibilities the choice of the semi-intersextile range Q is most advantageous (see e.g. Fig. 5.27 on page 203 by Steiner 1990: the relative

asymptotic scatters of the semi-interquantile ranges have its flat minimum zones for the mostly occurring error-types around $p = 1/6$). Remember that the definition of Q is simply $F^{-1}(1 - 1/6)$ if $f(x)$ is unimodal and symmetrical to the origin, both estimates $\overline{x}$ and *med*, however, have also this property if the parent distribution is symmetrical to the origin, too.

It is easy to verify (using the definition of characteristic function and the convolution theorem) that in the above mentioned circumstance for $Q_{\overline{x}}(n)$

$$\int_0^{nQ_{\overline{x}}(n)} \left\{ \int_0^{\infty} \left[\int_0^{\infty} \cos(xt) \cdot f(x) \cdot dx \right]^n \cos(yt) dt \right\} dy = 1/3 \tag{9}$$

must be hold. The solution of Eq. (9) needs not neglectable computing time but $Q_{\overline{x}}(n)$ is analytically determined and therefore highly accurate. (Monte Carlo determination would need irrealistic computer time to reach the same accuracy.)

Restricting ourselves to odd $n = 2k + 1$ sample sizes, it is enough to use the distribution function of the middle element of the ordered sample, i.e., $F_{n,k+1}(x)$:

$$F_{n,k+1}(x) = n \binom{2k}{k} \int_{-\infty}^{x} F(t)^k \cdot [1 - F(t)]^k f(t) dt \tag{10}$$

(see e.g. page 76 by Vincze 1968); we have now then to solve the following equation:

$$F_{n,k+1}[Q_{\text{med}}(n)] = 5/6\,. \tag{11}$$

Having the values $Q_{\overline{x}}(n)$ and $Q_{\text{med}}(n)$, analogously to Eq. (3) the relative efficiency $e(L_2/L_1, n)$ can be expressed as

$$e(L_2/L_1, n) = \frac{Q^2_{\text{med}}(n)}{Q^2_{\overline{x}}(n)}\,, \tag{12}$$

the reciprocal of this quotient gives evidently $e(L_1/L_2, n)$.

These pairs of relative efficiency curves are shown on Figs 3–5 for $n = 7.5$ and 3; the corresponding numerical values are summarized in Table I (see the columns 5–10). It is suprising that the place of the intersection point of both curves is near to the asymptotic one even at $n = 7$ (compare the Figs 2 and 3), — and even by the smallest sample size $n = 3$ is this place at $p = 1.27$ (being $e(L_2/L_1, 3) = 99.77\%$ and $e(L_1/L_2, 3) = 100.23\%$), which corresponds to $a = 4.48$ in the same sense which was discuted for the asymptotic case at the end of Section 2 (resulting there in the value $a = 5.56$). Our conclusion is that in outlier-free cases the conventional statistics can be used more advantageously than the L_1-norm based procedures as far as the often occurring neighbourhood of the (geo)statistical error-distribution type.

Let's make a last comment: is the sample size $n = 3$ not too small to investigate it? Not at all. Remember the old rule of thumb: one measurement is no measurement, and therefore e.g. in geodetic (or gravimetric) field measurements each

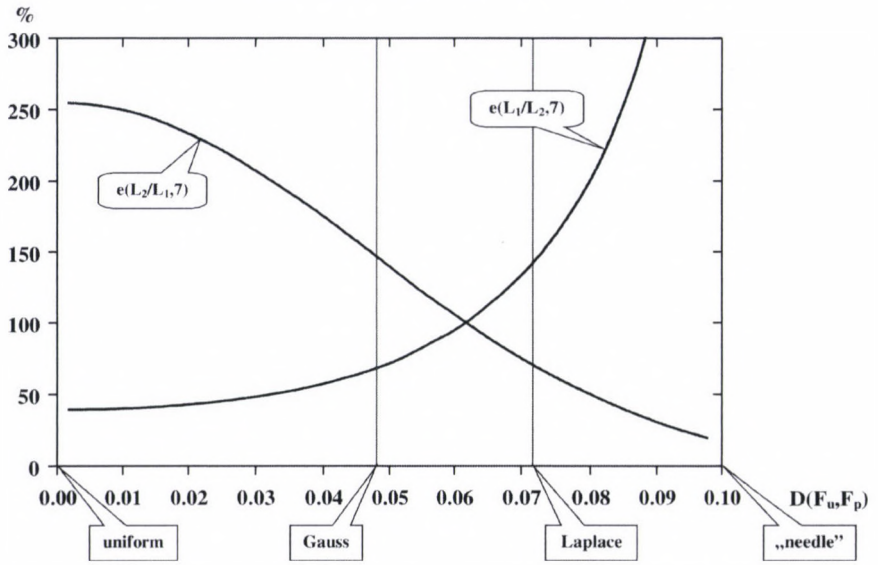


Fig. 3. Relative efficiency curves $e(L_2/L_1, n)$ and $e(L_1/L_2, n)$ for the types of the generalized Gaussian supermodel is the sample size is $n = 7$ vs. the typedistance $D(F_u, F_p)$

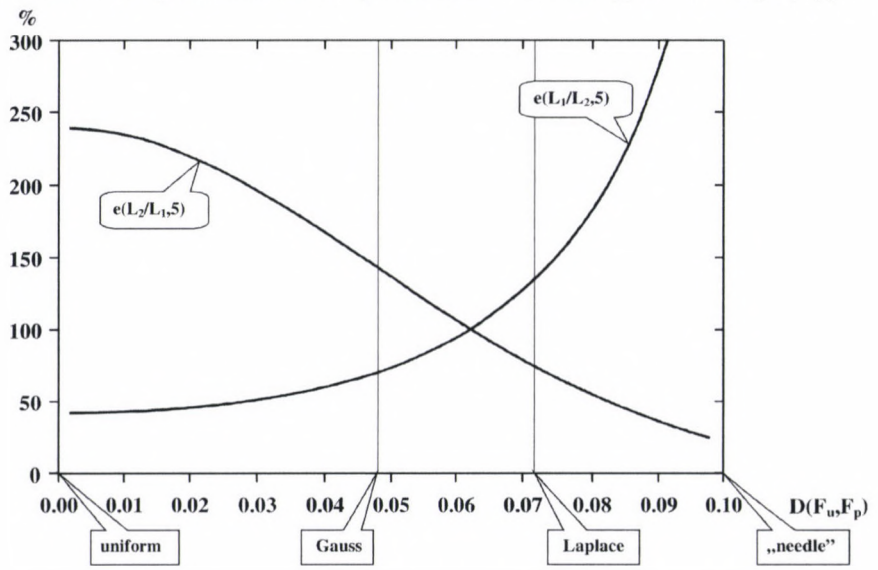


Fig. 4. Relative efficiency curves $e(L_2/L_1, n)$ and $e(L_1/L_2, n)$ for the types of the generalized Gaussian supermodel is the sample size is $n = 5$ vs. the typedistance $D(F_u, F_p)$

measurement is repeated most commonly three times and the arithmetic mean of these three values is accepted. (In such circumstances an eventually occurring outlier strikes in such degree the eye that it is immediately omitted by the observer.) The present paper justifies (as a by-product) that in such cases is not any modernization needed.

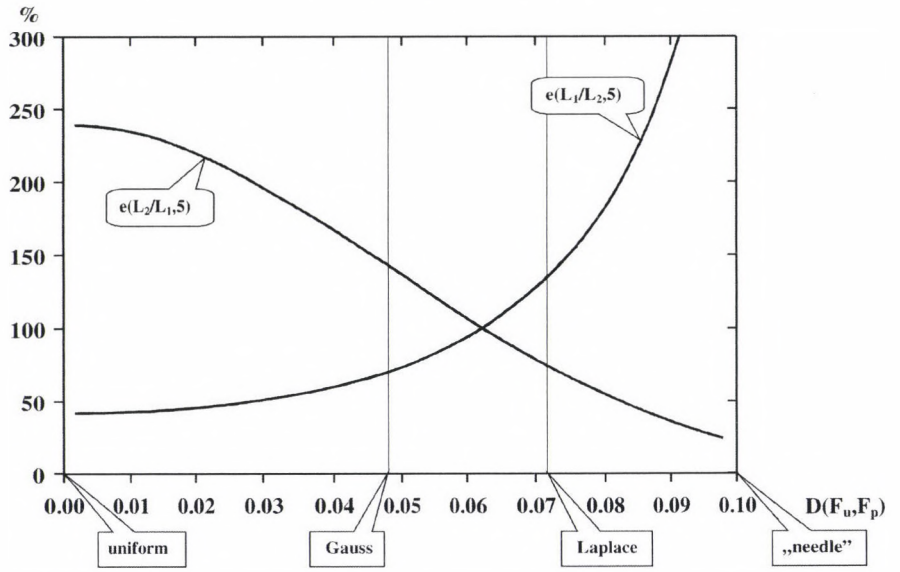


Fig. 5. Relative efficiency curves $e(L_2/L_1, n)$ and $e(L_1/L_2, n)$ for the types of the generalized Gaussian supermodel is the sample size is $n = 3$ vs. the typedistance $D(F_u, F_p)$. The abscissa of the intersection point of both curves are hardly shifted in the right direction even compared it to the asymptotic case (see Fig. 2) notwithstanding in the present case the sample size is extremely small

Acknowledgement

The present research work was supported by the Hungarian Ministry of Education (project number 0914/1997).

References

Andrews D F, Bickel P J, Hampel F R, Huber P J, Rogers Wbt, Tukey J W 1972: Robust Estimates of Location. Princeton University Press, Princeton, New Jersey

Dutter R 1986/1987: Mathematische Methoden in der Montangeologie. Vorlesungsnotizen, Manuscript, Leoben

Steiner F 1990: Introduction to Geostatistics (in Hungarian). Tankönyvkiadó, Budapest

Steiner F ed. 1997: Optimum Methods in Statistics. Akadémiai Kiadó, Budapest

Steiner F, Hajagos B 1998: *Acta Geod. Geoph. Hung.*, 33, 259–277.

Steiner F, Hajagos B 2000: P-norm based statistical procedures are more efficient then the L_1-based ones for all error-types of the complete supermodel $f_c(x)$. *Acta Geod. Geoph. Hung.* (present issue)

Tarantola A 1987: Inverse Problem Theory. Elsevier, Amsterdam

Vincze I 1968: Mathematical Statistics (in Hungarian). Műszaki Könyvkiadó, Budapest

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 313–317 (2000)

SEISMICITY OF THE SUB-CARPATHIAN REGION OF THE NORTH-EASTERN SECTION OF THE PANNONIAN BASIN

T Zsíros[1]

[Manuscript received April 9, 2000]

The Sub-Carpathian (Kárpátalja) region of the north-eastern section of the Pannonian Basin shows considerable seismicity by more than 600 events between 1662 and 1992 with magnitude M≤4.6. The distribution of epicenters indicates that the seismic activity is concentrating around the range of Avas, Kőhát and Gutin mountains. Based on the average recurrence we may expect
an earthquake of M3.0 or larger every 1 year
an earthquake of M4.1 or larger every 10 years
an earthquake of M4.7 or larger every 100 years
in the studied region.

Keywords: Carpathian; earthquake; magnitude; Pannonian Basin; seismicity

The seismic activity in the Pannonian Basin is highly nonhomogeneous, with some areas showing significant seismicity, as the number of events and/or the strength of earthquakes or both of them. One of these important seismic areas is the Sub-Carpathian (Kárpátalja) region in the north-eastern part of the Pannonian Basin defined as in Fig. 1. The first known earthquake record is from Nagybánya (Baia Mare) occurred in 1662. Its epicentral intensity was V-VI EMS which corresponds to 3.9 average magnitude (see below the conversion formula).

For this seismicity study the Hungarian earthquake catalog file (Zsíros 1997) is used, which includes 601 events located in the Sub-Carpathian source area between 1662–1992. The instrumental magnitudes — if they exist — are used as averaged values, while the macroseismic magnitudes are estimated by equation (1). This relation was determined for the whole Carpathian Basin using 303 events (Zsíros 1995).

$$M = 0.63 I_o + 1.06 \log h - 0.61 \,, \tag{1}$$

where
M – average instrumental magnitude
I_o – epicenter intensity
h – focal depth (km).

If the focal depth is unknown, $h = 10$ km is used as average depth value.

Though the Sub-Carpathian source zone is in the vicinity of Érmellék region, where the strong 1834 event with M6.2 occurred, it is characterized only by weaker shocks with a maximum magnitude of 4.6. The most important earthquakes of this zone with their basic parameters are summarized in Table I.

[1]Geodetic and Geophysical Research Institute of the Hung. Acad. Sci., H–1112 Budapest, Meredek u. 18, Hungary, e-mail: tibor@seismology.hu

1217-8977/2000/$ 5.00

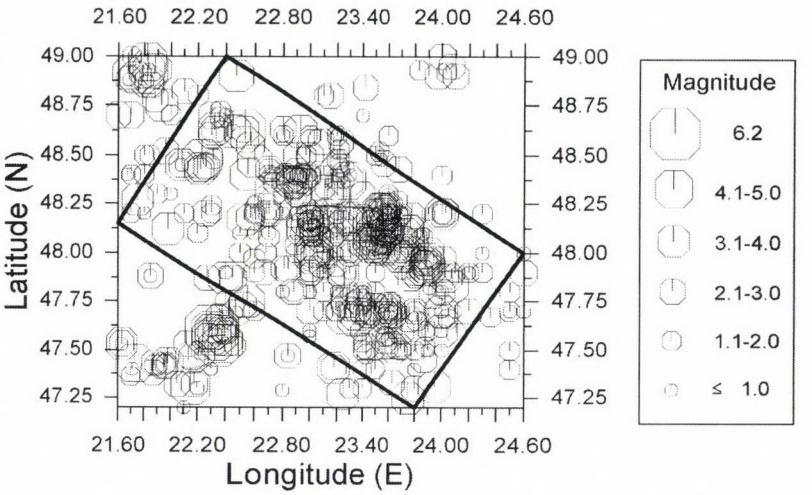


Fig. 1. Earthquake epicenters in the Sub-Carpathian region (magnitudes are averaged values)

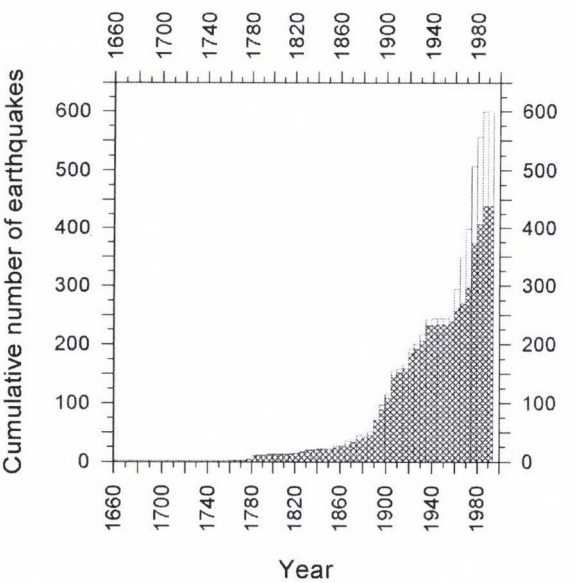


Fig. 2. Cumulative number of earthquakes

The cumulative number of earthquakes in the region is shown in Fig. 2 indicating sharper increase between 1890–1935 and 1960–1980. (The hatched area refers to earthquakes which have magnitude and/or epicenter intensity.)

The calculation of the energy released by earthquakes is based on the following relationship (Gutenberg and Richter 1942):

$$\log E = 1.5M + 4.8\,, \tag{2}$$

Table I. The most important earthquakes M≥4.0 in the Sub-Carpathian region

Date	Time	Epicentre		h	M	I	Location/Region
17840123	22	47.93N	23.88E		4.2	6.0	Máramarossziget
18721226	1440	48.37N	23.28E		4.2	6.0	Dolha
18930311	1025	47.99N	23.03E	9	4.2	6.0	Halmi area
18941007	2130	48.05N	23.47E		4.2	6.0	Bustyaháza
19260310	0210	48.00N	23.70E		4.2	6.0	Máramaros region
19260810	0210	48.00N	23.70E	9	4.0+	6.5	Taraczköz
19310407	0235	48.17N	22.53E	6	4.0	6.0	Beregdaróc
19360802	2127	48.63N	22.53E		4.2	6.0	Antalócz
19370914	0958	48.21N	23.54E		4.2	6.0	Kövesliget
19780630	021529.3	47.68N	23.27E		4.1+	6.0	Balotafalu
19790328	140243.	47.67N	23.35E		4.1+	6.0	Borhid
19790330	165615.7	47.70N	23.36E	13+	4.6+	6.5	Láposbánya area
19790922	190612.3	48.14N	23.58E		4.2+	6.5	Uglya
19820701	065000.	48.48N	22.23E	14+	4.3+	6.0	Nagydobrony

Time – local time (GMT+1 hour)
h – focal depth (km)
M – average magnitude
I – epicentral intensity
+ indicates instrumental data

where
M – magnitude
E – energy in joule.

The increase in the energy release curve (Fig. 3) is more or less monotonous between 1890–1940; the biggest jump is due to the M4.6 Láposbánya event (March 30, 1979); the 1940–1978 time interval is characterized by a small increase in energy.

As it can be seen in Fig. 1, a high concentration of earthquakes surrounds a relatively small area with lack of seismicity. This peculiarity is more evident in Fig. 4, where the smallest M≤3.0 events are excluded. The earthquakes occur along the dotted lines which connect Huszt, Nagyszőllős (Vinogradov), Nagybánya (Baia Mare), Kapnikbánya (Cavnic) and Rónaszék (Costiui) locations. That is the seismic activity is concentrating around the range of Avas, Kőhát and Gutin mountains.

The relative occurrence of small and large earthquakes over a limited magnitude range is expressed by relating the number of events N equal to or greater than magnitude M by the formula (Gutenberg and Richter 1949):

$$\log N = a - bM \,. \tag{3}$$

Studying the completeness of the data it is concluded that events with M≥3.0

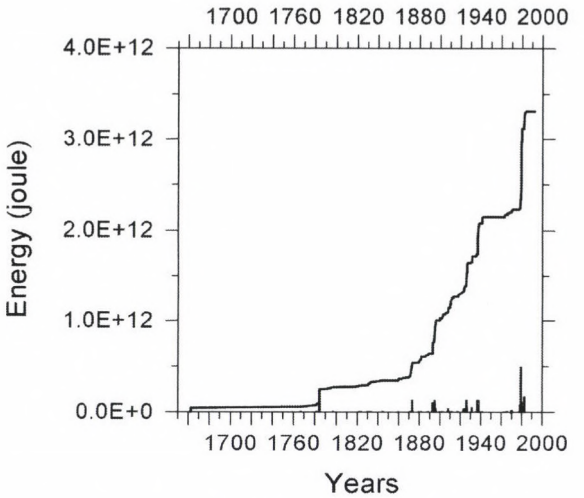


Fig. 3. Energy release curve of earthquakes

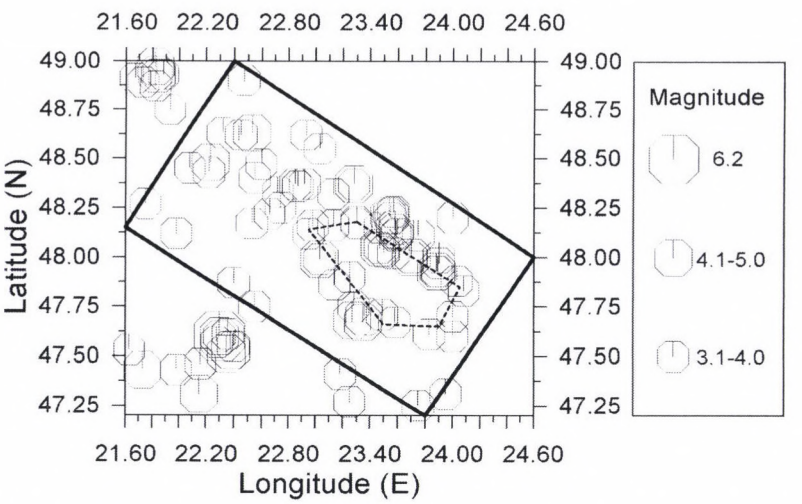


Fig. 4. Epicenter distribution of earthquakes above M3.0 in the north-eastern section of the Pannonian Basin. The sources are along the dotted lines delineate the range of Avas, Kőhát and Gutin mountains

can be regarded as complete dataset from 1880. The plot of the logarithm of the cumulative number of earthquakes per year above M3.0, against magnitude, (Fig. 5) gives the relation $\log(N) = 3.77 - 1.22M$ for the studied area. This linear fit, however, tends to overpredict the occurrence of the larger earthquakes, therefore the use of a truncated exponential relationship (Cornell and Vanmarcke 1969)

$$N = \alpha \frac{\exp(\beta M) - \exp(\beta M_u)}{1 - \exp(\beta M_u)}, \tag{4}$$

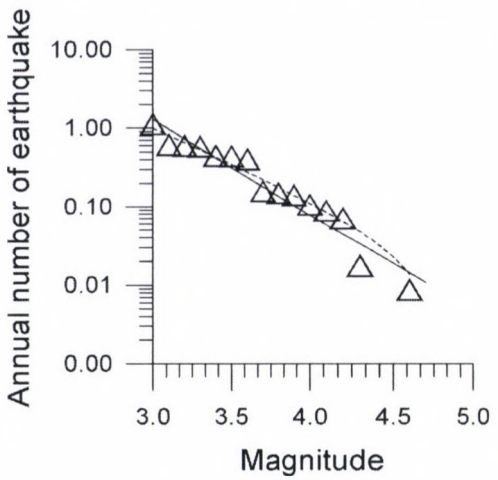


Fig. 5. Annual frequency of earthquakes. Triangles are observed data; the solid line is a linear fit and the dotted line is the preferred truncated fit

where α and β are related to a and b in Eq. (3), while M_u is maximum magnitude – seems to be a better solution. A least-square fit to the data gives values of 409.98 and 1.99 for α and β respectively and 4.81 is calculated for M_u. In this way it can be concluded about the average recurrence that we may expect

an earthquake of M3.0 or larger every 1 year
an earthquake of M4.1 or larger every 10 years
an earthquake of M4.7 or larger every 100 years

in the Sub-Carpathian region.

Acknowledgement

The preparation of this paper was supported by the Copernicus CT94-0238 project.

References

Cornell C A, Vanmarcke E H 1969: In: Proceedings of the Fourth World Conference on Earthquake Engineering, Santiago, A-1, 69–93.

Gutenberg B, Richter C F 1942: *Bull. Seism. Soc. Am.*, 32, 163–191.

Gutenberg B, Richter C F 1949: Seismicity of the Earth. Princeton Univ. Press

Zsíros T 1995: Earthquake magnitude and epicenter intensity relationship for the Carpathian region. Seismological Observatory, Budapest (manuscript)

Zsíros T 1997: Hungarian earthquake catalog. Seismological Observatory, Budapest (computer file)

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 319–332 (2000)

ON THE SEISMIC HAZARD ESTIMATION IN THE HIMALAYAN SEISMIC ZONES

D SHANKER[1]

[Manuscript received April 13, 2000]

The seismic hazard parameter estimation using Maximum Likelihood Method (MLM) has been made in the Himalayas and adjoing regions on the basis of the procedure which utilizes mixed data containing incomplete files of large historical events (extreme part of the catalogue) and complete files of most recent earthquakes (complete part of the catalogue) with same threshold magnitude. The entire earthquake catalogues used covers the period from 1720–1990. The maximum regional magnitude, $M_{\max}$, the activity rate of seismic events, λ, the mean return period, R, of earthquakes with a certain lower magnitude, $M \geq m$, as well as the parameter b of magnitude-frequency relationship have been determined for six seismic zones having different seismotectonic environments. High differences of the b parameters and hazard level from seismic zone I to VI reflect the high seismotectonic complexity and crustal heterogeneity, suggest that important project like hydroelectric project and town planning should preferably be located in the zones of lower hazard level.

Keywords: Himalayas; macroseismic; Maximum Likelihood Method (MLM); seismic hazard; techno-economic

Introduction

The natural hazards are inevitable ... Natural disaster are not. Earthquakes are among the most unavoidable natural hazards. One of the most frightening and destructive phenomena of nature is a severe earthquake and its terrible after effects. Numerous attempts to understand and predict this natural phenomenon have yielded partial successes, but most of the time nature maintains its superiority over present day science by striking in unexpected areas at the most unexpected times of day. The key point of seismic harad assessment is the determination of seismogenesis zones and structures withint he territory studied. It is carried out by the complex analysis of the seismological, geological, geophysical and geodetic information available. But the seismological aspect being the most important one is required to be considered in more detail.

Most probabilistic seismic hazard assessment procedure require the determination of seismic source uones, and a knowledge of their hazard parameters such as activity rate of level of catalog completeness, Guttenberg-Richter parameter b, and maximum possible magnitude $M_{\max}$. These parameters are then used to estimate the value of seismic hazard. As such information is not readily available for large part of the Indian subcontinent and most Indian seismic catalogues are highly uncertain and incomplete. Most of the available earthquake catalogues usually contain

[1]Department of Earthquake Engineering, University of Roorkee, Roorkee–247667, India, e-mail: dayasfeq@rurkiu.ernet.in

1217-8977/2000/$ 5.00

two types of information: macroseismic observations of major seismic events that occurred over a period of few hundred years, and complete instrumental data for relatively short periods of time, the last fifty at the most. The method which are generally used for the estimation of seismic hazard parameters (Weichert 1980, Dong et al. 1984) are not suitable for this type of data, because of the incompleteness of the macroseismic part of a catalogue or because of difficulties in estimating its growing incompleteness in earlier times.

Last two decades much effort has been made to obtaining realistic assessment of seismic hazard (Kiremidjian and Shah 1975, Mortgat and Shah 1979, McCann 1981, Wesnousky 1986). Seismic hazard studies of different tectonic regions have been carried out by various workers (Bath 1983, Markopoulos and Burton 1985, Papazachos 1988, Papadopoulos and Voidomatis 1987, Papadopoulos and Kijko 1991, for the Aegean region, Kijko and Sollevoll 1989, 1992 for western Norway costal region, Kaila and Narain 1976, Khattri et al. 1984, Rao and Rao 1979, Gupta and Srivastava 1990, Shanker and Singh 1997, Sing and Shanker 1993, Shanker and Sharma 1997, 1998, for territory of Indian subcontinent).

Shanker and Sharma (1998) have estimated seismic hazard parameters for the same region using only the complete parts of the database (1900–1990). In the present article we use the method suggested by Kijko and Sellevoll (1989) to estimate seismic hazard parameters in the six identified seismotectonic zones of the Himalayan belt utilizing mixed data containing incomplete files of large historical earthquakes (extreme par of catalogue) and complete files of most recent earthquakes (complete part of the catalogue).

Seismic zonation

In the seismic zonation procedure we considered a relatively small part of the lithosphere which includes the rupture zone (fault, deformation volume of the largest mainshock of this part of the lithosphere as well as the rupture zones of smaller mainshocks). The characteristic property of a seismogenic region is the interaction among its faults during the important seismic excitation (redistribution of stress etc.). Therefore, zonation in the present case is the procedure of defining the boundaries of a seismogenic region.

It was observed that although the accurate definition of seismogenic regions is important. It probably indicates that seismogenic regions, that is regions where interaction between faults occur, have not always very sharp boundaries, but there is a probability even for a distant fault to take part in the interaction but the probability decreases with increasing distance. The division of the studied area into seismotectonic segments, that is into part of the seismic zones having individual characteristics, is one of the basic requirements for the application of the estimation procedure for seismic hazard parameters. To check the effect of zonation on the main feature of the hazard determination procedure different divisions were made in areas with complex seismicity patterns and tectonics.

Based on the seismotectonics,s eismicity distribution and topographic variations the whole area is divided into four distinct sections namely, Hindukush Pamirs

(HKP), the Himalayas, the North East India (NEI) and the Burma-Andaman-Nicobar (BAN) section. Further the Himalayas have been subdivided into Kashmir-Himachal Pradesh Himalayas (KHP) and the central Himalayas. The central Himalayas is further subdivided into India-Western Nepal Border (IWNB). The following points form the basis of a short description of seismic zonation of the considered region.

(i) KHP region has V-Shaped lithosphere while Andaman Nicobar region shows Island Arc type of plate boundary.

(ii) NEI is characterized by several strong ($M > 7.0$) earthquakes with mulitple fault system. The pattern of energy release is distinctly different for HKP, NEI and BAN (Chauhan 1975).

(iii) NISB region of central Himalaya has been divided into two zones instead of one as taken by Chauhan (1975). This is because the strain release for IWNB zone is quite different from that of whole region (Srivastava 1973). Similar strain release for Hindukush region is also different (Drakopoulos and Srivastava 1972).

(iv) In IWNB region the largest magnitude of earthquake reported is 7.5. Great earthquake of magnitude greater than 8.0 have occurred in all the zones. The epicentral map of Himalaya (Fig. 1) shows that there are two distinct clusters of seismic activity located in IWNB and NISB region justifying the division of Central Himalaya into two separate zones.

(v) A rectangular shape has been taken for the seismic zones for the better presentation purpose only.

On the above criterion the whole Himalayan belt (20–36° N and 69–100° E) which is seismically very active and highly complicated from seismotectonic point of view, and for this reason it has been divided into six active zones based on seismotectonics, seismicity distribution and topographic variations. These six seismic zones (Shanker and Sharma 1998) along the major tectonic features of the area (rectangles I, II, III, IV, V and VI in Fig. 1) are:

Seismic zone I	Hindukush-Pamir (HKP) (25–36° N and 69–75° E)
Seismic zone II	Kashmir-Himachal Pradesh (KHP) (25–30° N and 75–80° E)
Seismic zone III	India-Western Nepal Border (IWNB) (25–32° N and 80–85° E)
Seismic zone IV	Nepal-India-Sikkim Border (NISB) (20–30° N and 85–90° E)
Seismic zone V	North-East India (NEI) (20–30° N and 90–95° E)
Seismic zone VI	Burma-Andaman-Nicobar (BAN) (20–30° N and 95–100° E)

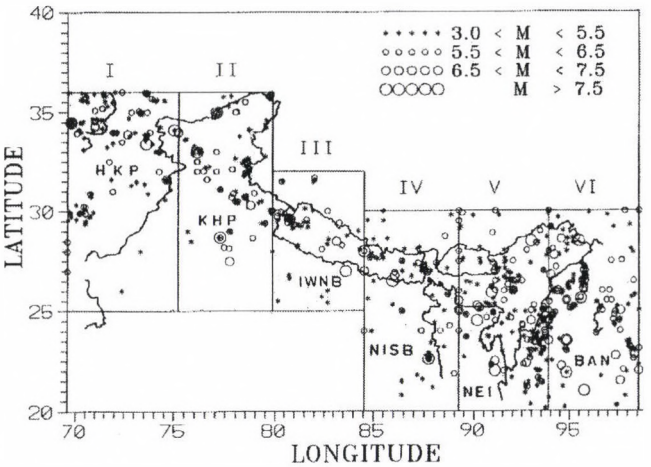


Fig. 1. Seismicity map of the Himalayas and its vicinity for the period (1720–1990). The identified seismic zones for the analysis are shown by rectangles denoted by I, II, III–VI (after Shanker and Sharma 1998)

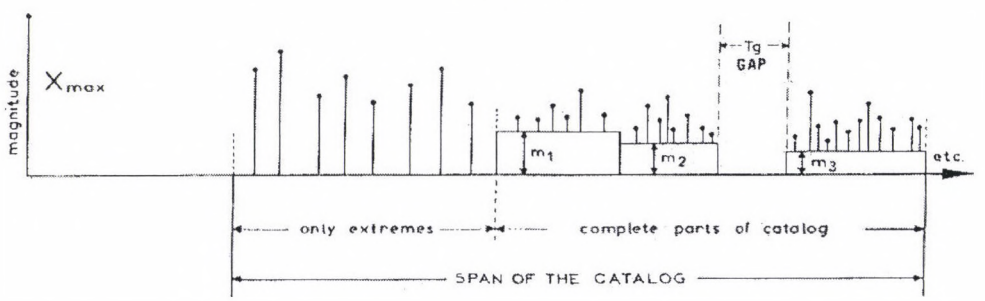


Fig. 2. Illustration of data used for seismic hazard parameter estimation (after Kijko and Sellevoll 1987) taking into account extreme as well as complete data files. This approach also accepth gaps (T_g: period for missing earthquake data). Notations appear in the figure have been given in the Appendix

Methodology

Generally, the available earthquake catalogues contain two types of information: one is macroseismic observations of major seismic events that occurred over a period of a few hundred years and other complete instrumental data for relatively short periods of time (Fig. 2). The methods which are generally used for the estimation of seismic activity parameters (parameter b in G-R equation, earthquake activity rate λ, and $M_{\max}$) are not suitable for this type of data due to incompleteness of the macroseismic (extreme) part of the catalogue.

The suitable methods for analyzing the macroseismic part of the catalogue are extreme distribution, that allow for varying time intervals from which maximum magnitudes are selected. This method incorporating the incomplete part of the catalogue into the analysis is very far from being optimum, as a great deal of information contained in small shocks is wasted (Kijko and Sellevoll 1987). Another

method for estimating the seismic activity parameter is to reject the macroseismic observations that are incomplete and to use any standard method for the data from the other complete part of the catalogue. It is obvious that this procedure is also highly ineffective, as the quantitative assessment of recurrence of strong seismic events based on observations over a short period of time is burdened with large error (Knopoff and Kagan 1977, Dong et al. 1984). Kijko and Sellevoll (1987, 1989) have developed an approach utilizing both the information of strong events contained in the macroseismic part of the catalogue as well as that contained in the complete catalogue which contains complete data above a certain magnitude threshold. This method assumes the Poissonian model of earthquake occurrence and the doubly truncated Gutenberg-Richter distribution. In this paper we use this approach to estimate seismic hazard parameters in six major seismotectonic zones (Fig. 1) of Himalayas and its vicinity.

Seismological database

In the present analysis the earthquake catalogue (1720–1990) has been used. The earthquake catalog from 1720–1986 has been taken from Chandra (1992) and data after that has been taken from the National Geophysical Data Center, Boulder, Colorado, USA. The plot of epicentres with magnitude $M \geq 3.0$ are shown in Fig. 1. But in actual calculations wer have used earthquake data $M \geq 4.5$, due to the discontinuity in the catalogue. The whole span of the database has been divided into two parts one is macroseismic part of the catalogue (1720–1900) and other is complete part of the catalogue (1900–1990).

Estimation of seismic hazard parameters

A well known technique (Kijko and Sellevoll 1989) for the estimation of seismic hazard parameters for the complete data files have been used. The short frame work of the mathematical forumulations and necessary notations are given in the Appendix for the readers. The method accepts mixed data, one containing only the largest earthquakes (extreme part) and other containing data sets complete to different threshold of magnitude (complete part). A computer program based on this method and written by the above author is also used for the analysis. According to the requirement of this computer program, the whole span of database has been broken in to two parts one extreme part (1720–1900) and other complete part (1900–1990) as illustrated in Fig. 2. It was assumed that in order to estimate the seismic hazard in the present study, for each zone, the maximum possible magnitude $M_{\max}$ was determined with standard error as 0.10, threshold magnitude 5.0 and $M_{\min} = 4.0$. Table I summarizes the input data used for the seismic hazard evaluation int he six seismic zones of the Himalayan belt. The selection of extreme part of the catalogue is demonstrate in Fig. 3. The inset in the histogram of the extreme and complete part shows the extreme magnitude selected from a 10 year time period, constituting the extreme data file for seismic hazard estimation.

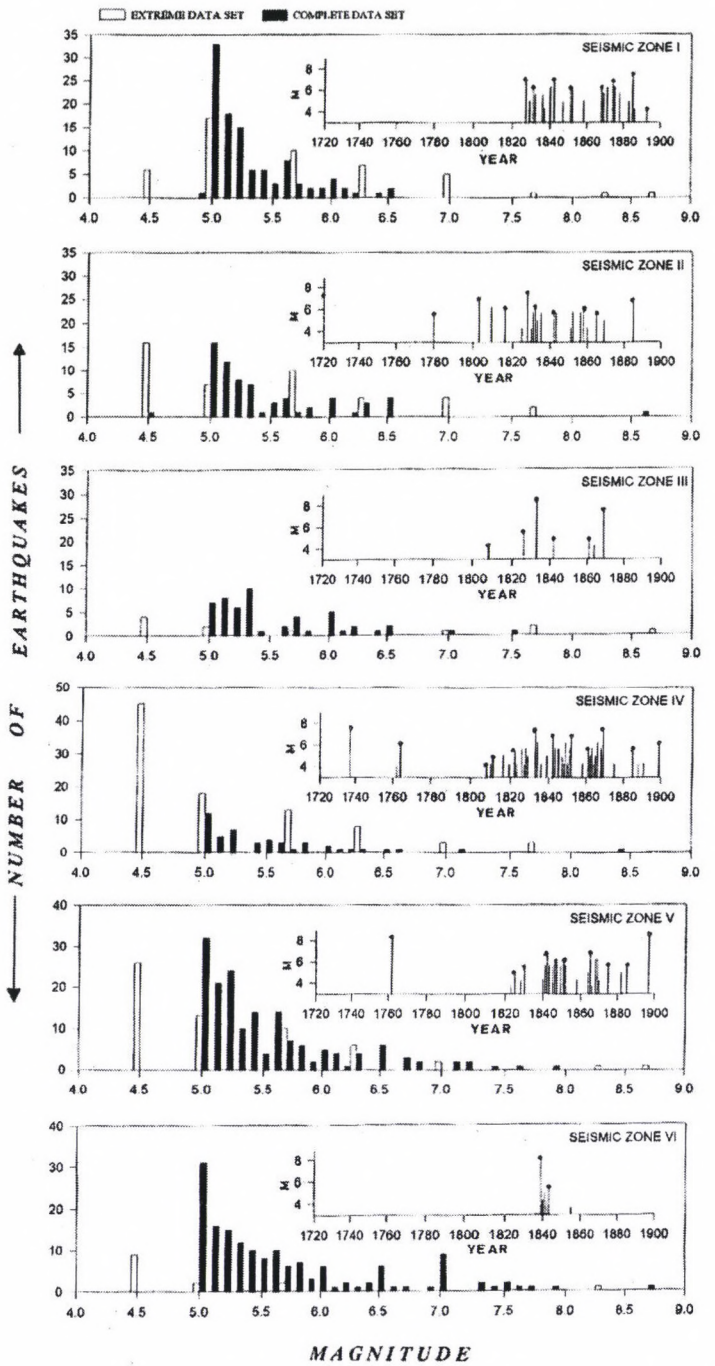


Fig. 3. Frequency distribution histograms for the extreme (1720–1990) and complete (1900–1990) part of the whole data set for the six seismic zones. The extreme are selected (marked as) from the period 1720–1900 with 10 year time interval as shown in the inset with magnitude vs. time plots

Table I. Input data used for seismic hazard analysis

Seismic zone	Number of events extreme (N_e)	complete (N_c)	Standard deviation	Max. obs. magnitude
I	7	107	0.37	7.7
II	9	68	0.66	8.6
III	5	52	0.55	8.7
IV	12	46	0.66	8.4
V	8	166	0.59	8.7
VI	3	157	0.75	8.7

Result and discussion

The zone I, Hindukush-Pamir (HKP) is limited by 25–36° N and 69–75° E, and has a V-shaped lithosphere as inferred by seismicity. The depth range of earthquakes in this region is 5 to 250 km. The largest earthquakes of Richter magnitude 8.0 occurred on 21 Oct. 1907 and 7th July 1909. The seismicity distribution is along Herat fault north of Kabul, Chaman fault and mountain ranges in the Pamir Knot. The focal mechanism indicates thrust faulting. The extreme part of the catalogues starts at 1720/01/01 and ends at 1889/12/03. The complete parts of the catalogues starts at 1900/01/01 and ends at 1990/12/03 with threshold magnitude 5.0. Analysis gives $\beta = 1.98 \pm 0.14$, $\lambda = 16.25 \pm 3.33$, $M_{\max} = 8.2 \pm 0.86$ and $b = 0.84 \pm 0.6$, where $M_{\min} = 3.3$. For the above computation the data contribution to the parameters are 36.2% and 6.2% for extreme part as well as 63.8% and 93.8% for the complete part of the catalogue for β and λ, respectively.

Seismic zone II, Kashmir-Himachal Pradesh (KHP) extended from 25–30° N and 75–80° E is seismically very active. The earthquake are shallow in nature. The largest earthquake of magnitude 8.0 was Kangra earthquake of 1905, which killed over 20000 people. Srinager 1885 major earthquake took a toll of 6000 lives. The prominent fault in this region are the Main Boundary Thrust (MBT), Indus sature zone (ISZ) and the Kaurik fault. In this zone the number of events in the extreme part is 9 and in the complete part is 68. The value of hazard parameter for this zone are $\beta = 1.83 \pm 0.16$, $\lambda = 2.35 \pm 0.29$, $M_{\max} = 9.1 \pm 2.10$, $b = 0.77 \pm 0.07$, where $M_{\min} = 4.0$. The data contribution to the parameters are $\beta = 49.6\%$ and $\lambda = 11.8\%$ and 88.2% for extreme and complete part, respectively.

Seismic zone III, India-Western-Nepal Border (IWNB) these earthquake are shallow to a depth of 23 km. The largest earthquake in this region occurred on 28 Aug. 1916 in Dharchula. Utterkashi earthquake (1991) of magnitude 7 killed 768 people. The seismic activity is different along the Main Central Thrust (MCT), as well as other thrust like Almora and Vaikrita thrust and tear fault like Yamuna Tear. The value of hazard parameters are $\beta = 1.15 \pm 0.16$, $\lambda = 1.18 \pm 0.26$, $M_{\max} = 9.2 \pm 0.74$ and $b = 0.49 \pm 0.07$, where $M_{\min} = 3.30$. Here the data contribution by the complete part are 54.2% and 45.8% for β and 91.1% and 8.9% for λ, respectively.

Nepal-India-Sikkim Border (NISB) i.e. zone IV is characterized by scattered

seismic activity associated with MCT and MBT and some tear fault. The largest earthquake on 15 Jan. 1934 had a magnitude 8.3. The recent earthquake of August 22, 1988 ($M \geq 6.8$) though smaller in magnitude than the great Bihar-Nepal earthquake of 1934 shows almost smaller miezoseismal pattern. For this seismic zone the hazard parameters has been calculated as $\beta = 1.29 \pm 0.17$, $\lambda = 1.77 \pm 0.41$, $M_{\max} = 8.9 \pm 0.76$, while $M_{\min} = 3.30$ and $b = 0.55 \pm 0.07$. The data contribution by the complete part and extreme part are 44.5% and 55.5% for β and 79.6% and 20.4% for λ, respectively.

The seismic zone V, North-East India (NEI) is seismically very active. Earthquakes are shallow to intermediate depths (up to 200 km). The focald epths icnreases towards Burmese arc, and eastern margin of the Indian plate. The largest earthquake of magnitude 8.7 which occurred on 12 June 1897 was most disastrous in the history of earthquake. Another earthquake of magnitude 8.5 occurred on 15 August 1950 in the eastern syntaxial bend close to India-Tibet border. The hazard parameters for this region are $\beta = 2.16 \pm 0.12$, $\lambda = 6.70 \pm 0.61$, $M_{\max} = 9.2 \pm 3.8$, where $M_{\min} = 4.0$ and $b = 0.92 \pm 0.5$. These values agree well with the values obtained only by the complete part of the data files (Shanker and Sharma 1997). The data contribution by the complete part and extreme part are 46.0% and 54.0% for β and 96.4% and 4.6% for λ, respectively.

In The Burma-Andman-Nicobart (BAN) i.e. seismic zone VI, the earthquakes have occurred from a shallow depth of 5 to 230 km. The largest earthquake in this region has been reported to be of magnitude 8.1 which occurred on 26 June 1941. The focal mechanism of many of earthquakes in this region show predominance of strike-slip movement. The calculated hazard values are $\beta = 1.78 \pm 0.11$, $\lambda = 3.91 \pm 0.31$, $M_{\max} = 9.2 \pm 3.8$, where $M_{\min} = 4.0$ and $b = 0.75 \pm 0.05$. The data contribution by the complete part and extreme part are 53.4% and 46.6% for β and 98.2% and 1.8% for λ, respectively.

Seismic data interpretation of Table II indicate that β and b-values from Zone I to VI but Zone V (NEI) has highest value of β and consequently high b-value showing the activity of the zone. The λ is highest in Zone I and lowest in Zone III and hence lower return period for Zone I and greatest for Zone III. From Fig. 4 it is clear that Zone II, III and IV show greater return period for $M \geq 6.0$ than the Zone I, V and VI, implying strong events more frequently occur in the Zone I, V and VI than the rest of the zones. Frequency distribution curve (Fig. 3) alsso support our view points. Considering the return period for $M = 6.0, 6.5, 7.0$, one may conclude that Zone VI is most active in the whole region. Probability of occurrence in 50, 100 and 1000 years return period (Fig. 5) implies that probability of occurrence of earthquakes decreases with magnitude in Zone I to VI.

In general, earthquake hazard is sharply higher in Zone I, II and V. Zone III and IV show relatively low hazard level than other zones and Zone VI is most hazardous zone thant he other zones int he whole Himalayan belt. The distribution of hazard potential for zone to zone is informative and useful for engineering point of view. So, the hazardous zones may be avoided for techno-economic and civil engineering constructions or proper risk evaluation be done before going for such planning.

Table II. Seismic hazard parameters

Seismic zone	±	λ	β	b-value	M_{max}	$R_{6.0}$
I	±16.25	3.33±1.98	0.14±0.84	0.06	8.2	13.1
II	±2.35	0.29±1.83	0.16±0.77	0.07	9.1	16.4
III	±1.18	0.26±1.15	0.16±0.49	0.07	9.2	19.2
IV	±1.77	0.41±1.29	0.17±0.55	0.07	8.9	18.9
V	±6.70	0.61±2.16	0.12±0.92	0.05	9.2	11.2
VI	±3.91	0.31±1.78	0.11±0.75	0.05	9.2	9.0

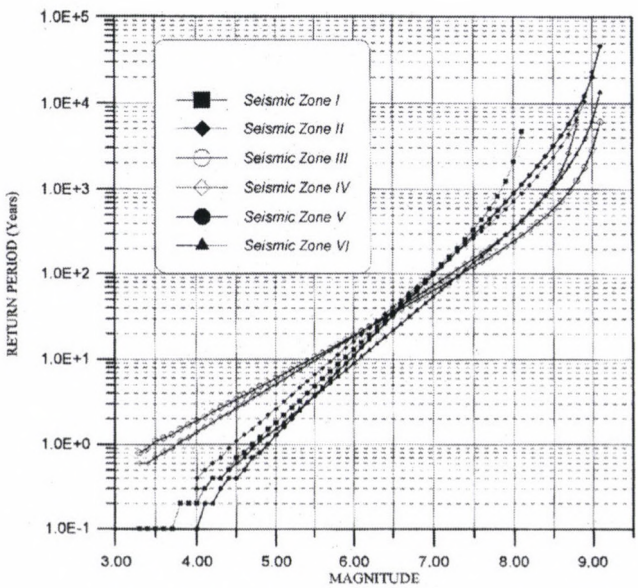


Fig. 4. Return period estimated using mixed (1720–1990) for the six seismic zones

Summary and concluding remarks

The reliability of the results for seismic hazard estimation in a given region depends on the methodology and information input used. For the sake of simplicity we have applied a MLM for mixed catalogue covering about 270-years data. As a consequence, the hazard of each seismic zone (I–VI) is independently determined. In conclusion, the hazard level is sharply high in zone NEI and BAN. Zone KHP, IWNB and NISB show intermediate hazard. The same can also be concluded from b-value distribution. The variation of seismic hazard from zone to zone may be used as a characteristic parameters in the application of microseismic zonation in young mountainous areas. The major earthquakes ($M \geq 7.5$) are likely to occur at those points along thrust and plate boundaries where excessive accumulation of strains/stresses are occurred. However, the stress accumulation will be restricted to local weak zones only along the thrust and junction of three lineaments. Thus,

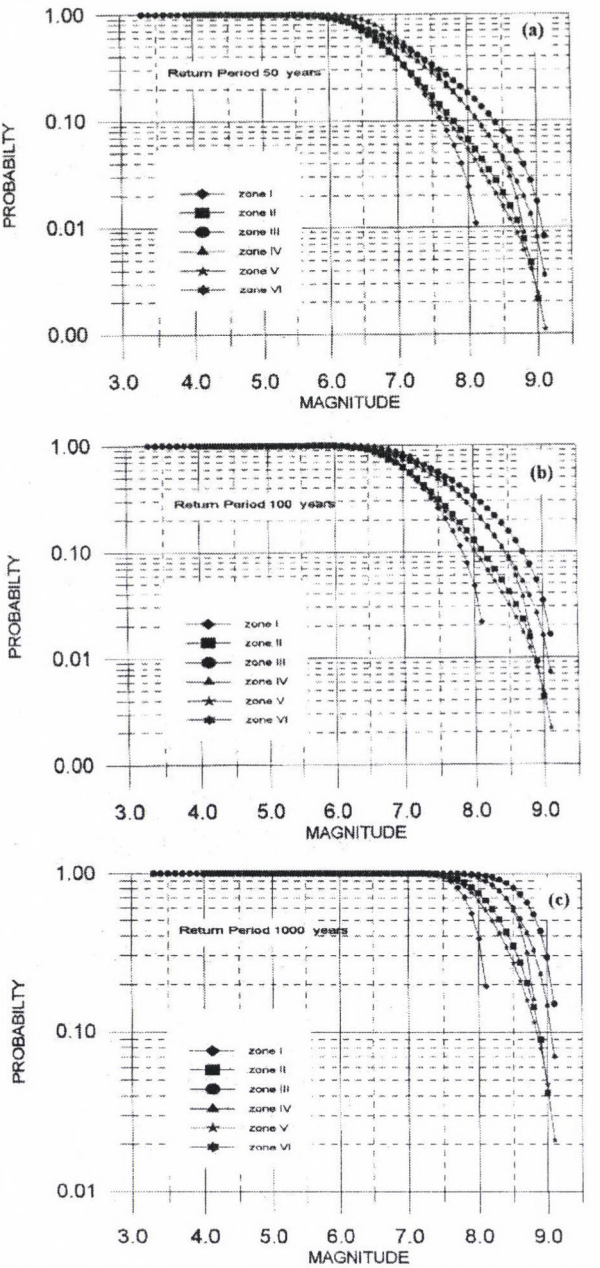


Fig. 5. Probability-magnetitude diagram for six seismic zones for the return period a) 50 years, b) 100 years ©1000 years

there is a possibility of adjoining areas of much lower seismic risk. Such areas of low seismic risk within the zone of high tectonic activity may be demarcated as those areas where major earthquakes have not occurred in the past two hundred years. It

appears that major earthquake will continue to occur repeatedly only the zone of strain accumulation and stress concentration. Thus, the zones of lower earthquake risk may be demarcated reliably on the basis of seismic history.

Seismological investigation suggest that sress concentration and major earthquakes are likely to occur where three plate boundaries are meeting such as NE-Himalaya and Kashmir. It is true that major earthquakes have been occurring in these zones. As such many areas of lesser earthquake risk may exit in Himalaya between Assam and Jammu as only two plates are interacting in the normal direction. Thus attention should be focussed on microseismic zonation.

The important projects like hydroelectric project and town planning should preferably be located in the areas of lower seismic risk. It is time to generate confidence and interest among geophysicists and geoscientists to develop computer modelling of 3-dimensional non-linear geodynamic analysis and seismic migration technique. Apart from its importance in seismicity and earthquake prediction studies, b is an indicator of the mechanical properties in the seismogenic materials, such as stress concentration, crack density and degree of heterogeneity. The hazard parameters λ, β and b-value in each seismic zone represent the seismic potential of each zone. Thus attention should be focused on these zones for future engineering or other practical purposes.

Acknowledgements

First author is indebted to Dr. S Basu, Professor and Head, Department of Earthquake Engineering, University or Roorkee, Roorkee, India for his helpful suggestion, nice advice and providing excellent computer facilities. Financial support provided by Department of Science and Technology (DST), New Delhi is gratefully acknowledge by the first author.

Appendix

Mathematical formulation and notations

For the sake of simplicity following notations have been used to estimate seismic hazard parameters in six major seismotectonic zones (seismic zones) of Himalayas using mixed catalogue.

b	=	slope of Gutenberg-Richter (G-R) magnitude-frequency relation
β	=	parameter interrelated to b ($b = \beta \log e$)
λ	=	activity rate of the seismic events
λ_o	=	value of λ corresponding to m_o
m_o	=	magnitude m_i; $i = 1, 2, 3, \ldots, s$
m_i	=	magnitude threshold int he i-th complete sub catalogue
s	=	number of complete sub catalogues
$M_{\max}$	=	maximum possible magnitude in the specified region
m	=	a certain earthquake magnitude
R	=	return period of earthquake with magnitude $M > m$
$R_{6.0}$	=	return period of earthquake with magnitude M = 6.0

T = time period covered by the whole catalogue; $(T_c + T_e)$
T_c = time interval covered by the extreme part of the catalogue
t_c = time interval covered by the complete part of the catalogue
T_i = time length of the i-th complete sub catalogue; $i = 1, 2, 3, \ldots, s$
n = total number of earthquakes
n_e = number of earthquakes in extreme part of the catalogue with magnitude
X_{ej} = selected from the time intervals t_j; $j = 1, 2, 3, \ldots, n_e$ (by definition $\sum t_j = T_e$)
n_i = number of events in the i-th complete sub catalogue with magnitude $X_{ij}^3 m_i$; $i = 1, 2, 3, \ldots, s$, $j = 1, 2, 3, \ldots, n_i$
n_c = number of events in the complete part of catalogue (by definition $\sum n_j = n_c$)
$X_{\max}$ = largest observed magnitude in the whole catalogue
$< X >$ = mean earthquake magnitude determined from the extreme and complete part of the catalogue.

Assuming the Poisson occurrence of earthquakes with activity rate λ and validity of the doubly truncated G-R relation, we get the following equations through the maximum likelihood method (Kijko and Sellevoll 1987, 1989):

$$\frac{1}{\lambda_o} = \phi_1^E + \phi_1^C \tag{1}$$

$$\frac{1}{\beta} =< X > -\phi_2^E + \phi_2^C + \lambda_o(\phi_3^E + \phi_3^C)\,, \tag{2}$$

where

$\phi_1^E = r_o B_1$
$\phi_2^E = r_o E(m_o, m_{\max})$
$\phi_3^E = r_o B_2 + \phi_2^E B_1$
$\phi_1^C = \sum_{i=1}^{S} T_i C_i / n$
$\phi_2^C = \sum_{i=1}^{S} r_i \left[E(m_i, m_{\max}) + (D_i/C_i)\right]$
$\phi_3^C = \sum_{i=1}^{S} D_i T_i / n.$

Moreover, we have

$n = \sum_{i=1}^{S} n_i$
$r_i = n_i / n$
$A_1 = e^{-\beta m_o}$
$A_2 = e^{-\beta m_{\max}}$
$A(m) = e^{-\beta m}$
$B_1 = (< t > A_2 - < tA >)/(A_2 - A_1)$

$$B_2 = (< tX_eA > - < t > m_{\max}A_2)/(A_2 - A_1)$$
$$C_i = 1 - F(m_i)$$
$$D_i = E(m_o, m_i) - E(m_o, m_{\max})F(m_i)$$

where $i = 1, 2, 3, \ldots, s$.

In addition

$$E(x, y) = [xA(x) - yA(y)]/[A(x) - A_2]$$
$$F(x) = [A_1 - A(x)]/[A_1 - A_2]$$

and $< t >= \sum t_i/n_e$

$$< tA >= \sum t_i A(X_{ei})/n_e$$
$$< tX_eA >= \sum t_i X_{ei} A(X_{ei})/n_e$$

where the summation is from $i = 1, 2, 3, \ldots, n_e$.

The indices E and c are introduced in order to distinguish different sources of function ϕ. If they are derived from the extreme part of catalogue, they are marked as E. Otherwise they follow from the complete parts (subcatalogues) and are marked as c.

Now putting the condition that the largest observed magnitude $X_{\max}$, is equal to the largest expected magnitude in the span of the whole catalogue, T, we get (Kijko 1988):

$$\begin{aligned} X_{\max} &= \text{EXPECTED}(x_{\max}|T) \\ \text{EXPECTED}(x_{\max}|T) &= m_{\max} - \frac{E_1(TZ_2) - E_1(TZ_1)}{\beta_e^{-TZ_2}} - m_o e^{-\lambda_o T}. \end{aligned} \tag{3}$$

In Eq. (3) $Z_1 = \overline{\lambda}A_1$, $Z_2 = \overline{\lambda}A_2$, $\overline{\lambda} = \lambda_o/(A_2 - A_1)$ and $E_1(.)$ denote an exponential integral function. By introducing the condition (3) into (1) and (2), we get a set of equations determining the maximum likelihood solution which can be solved by iteration. Applying the above theory for the present study a special cases $s = 1$, $r_o = 0$ ($n_o = 0$) and $m_{\min} = m$ implies that extreme magnitude are not taken into consideration, and the catalog is composed of only one complete part.

References

Bath M 1983: *Tectonophysics*, 98, 165–208.

Chandra U 1992: *Current Science*, 62, 40–71.

Chauhan R K S 1975: Some quantitative seismicity studies in north east India including Kuch. Mausam, 30, 23.

Dong W M, Bao A B, Shah H C 1984: *Bull. Seis. Soc. Am.*, 72, 725–737.

Drakopoulos J C, Srivastava H N 1972: *Annali di Geofisica*, 23, 593–606.

Gupta G D, Srivastava H N 1990: *Memoir Geological Society of India*, 23, 173–199.

Kaila K L, Narain H 1976: Evolution of the Himalaya based on the seismotectonics and deep seismic sounding. Int. Seminar, Himalayan Geology, Spl. 13–17.

Khattri K N, Rogers A M, Perkins D M, Algermissen S T 1984: *Tectonophysics*, 108, 93–134.

Kijko A, Sellevoll M A 1987: Estimated earthquake hazard parameters from incomplete data files. Seis. Ser., 11, Seismological Obs. University Bergen, Norway

Kijko A, Sellevoll M A 1989: *Bull. Seis. Soc. Am.*, 79, 645–654.

Kijko A, Sellevoll M A 1992: *Bull. Seis. Soc. Am.*, 82, 120–134.

Kiremidjian A S, Shah H C 1975: Seismic hazard mapping of California. Technical Report 29. The John A Blume Earthquake Engineering Center, Department of Civil Eng., Stanford University

Knopoff L, Kagan Y 1977: *J. Geoph. Res.*, 82, 5647–5657.

Markopoulos K C, Burton P W 1985: *Tectonophysics*, 117, 205–257.

McCann M W 1981: A Bayesian geophysical model for seismic hazard. PhD Thesis. Department of Civil Eng., Stanford University, Stanford, California

Mortgat C P, Shah H C 1979: *Bull. Seis. Soc. Am.*, 69, 1237–1251.

Papadopoulos G A, Kijko A 1991: *Tectonophysics*, 185, 277–294.

Papadopoulos G A, Voidomatis P 1987: *Pure Appl. Geophys.*, 125, 613–628.

Papazachos B C 1988: Seismic hazard and long-term prediction in Greece. European School of Earthquake Science, Course on Earthquake Hazard Assessment, Athens

Rao S P, Rao R B 1979: Estimated earthquake probabilities in north-east India. Andaman-Nicobar Island, Mausam, 30, 267–273.

Shanker D, Sharma M L 1997: Estimation of seismic hazard parameters for the north east India and its vicinity. In: Proc. Workshop on Earthquake Disaster Preparedness, University of Roorkee, Roorkee, India, 49–57.

Shanker D, Sharma M L 1998: *Pure and Apply Geophys.*, 152, 267–279.

Shanker D, Singh V P 1997: Seismic risk analysis for the occurrence of medium size earthquake in Kangra region of Himachal Pradesh, India. Indian National Science Academy (INSA), 63A(2), 71–76.

Singh V P, Shanker D 1993: *Geophysical Transactions*, 38, 135–149.

Srivastava H N 1973: *Him. Geol*, 3, 381–393.

Weichert D H 1980: *Bull. Seis. Soc. Am.*, 70, 1337–1356.

Wesnousky S G 1986: *J. Geoph. Res.*, 91, 12587–12631.

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 333–350 (2000)

DETERMINATION OF SITE FACTOR IN AND AROUND RHINE GRABEN, FRANCE IN THE 1–11 Hz FREQUENCY BAND BY NON-REFERENCE NETWORK AVERAGE TECHNIQUE

T PICQ[1], S K NATH[2], P SENGUPTA[2]

[Manuscript received September 3, 1999]

The greatest challenge in estimating site response from instrumental recordings of seismic events is removing the source and path effects. Several methods have been suggested in the past in that direction which may be divided into two main categories, depending on whether or not they need a reference site, with respect to which the particular effects at other sites can be estimated. An alternative approach is to determine the site response relative to the network average. In the present paper we propose one such technique for estimating the site factor in the Rhine graben and its surrounding provinces. Judging the higher signal thresh-hold level ($\geq$3 times the background) only sixteen earthquake events with local magnitude varying between 3 and 4.5 are used from the catalogue of 8 years during 1988-1995 recorded by the Fossé Rhénan network using short period single component seismometers belonging to the Ecole et Observatoire des Sciences de la Terre, Strasbourg, France and Survey Baden-Wurttenberg, Freiburg, Germany. Sixteen stations in and around Rhine graben are considered for the site factor computation in the 1–11 Hz frequency band. The contour maps prepared from the spatial distribution of non-reference network average site factor estimates at 1, 3, 5, 7, 9 and 11 Hz exhibit high values along the major faults and the minor tectonic features to the western (Vosges) and eastern (Foret-Noire) provinces. Irrespective of the magnitude of the earthquake, the Rhine graben is comparatively stable with low site factors at all the frequencies. As is expected from Coda site amplification studies, the network average contour patterns remained unchanged showing minor or no variation at all with frequencies.

Keywords: Fossé Rhénan network; non-reference network average; Rhine graben, seismic hazards; site factor; site response; source and path effects;

1. Introduction

Seismic microzonation involves division of a region into sub regions in which different safeguards must be utilised to reduce, and/or prevent damage, loss of life and societal disruptions during future earthquakes. The need to microzone urban areas was demonstrated by large spatial variations in damage distribution during big earthquakes like San Francisco earthquake of 1906, Alaska earthquake of 1964, Armenian earthquake of 1988, Loma Preita earthquake of 1989, Uttarkashi earthquake of 1991, Latur earthquake of 1993, North Ridge earthquake of 1994 and Turkish earthquake of 1999. Microzonation accounts for the diversity of earthquake on a

[1]Ecole de Physique du Globe, Strasbourg, France
[2]Department of Geology and Geophysics, Indian Institute of Technology Kharagpur-721302, West Bengal, India, e-mail: nath@gg.iitkgp.ernet.in

1217-8977/2000/$ 5.00

local scale depending on the information of local geology, soil profile topography, depth of water table, nature of earthquake source, characteristic of strong ground motion and their interaction with man made structures.

On the basis of the geological and seismological information of an area maps are prepared forecasting the probable risk. On these maps the risk is conceived as either relative or explicitly probabilistic. Maps made to give relative risk, demarcates zones with an arbitrary numerical or alphabetical scale ranging from no hazard to most hazard areas. Maps of probabilistic risks give an idea of the underlying statistical uncertainty. These maps give the odd at which a specified earthquake intensity would be exceeded at the site of interest within a given time span. In short, seismic zonation map of a country is a guide to the seismic status of a region and its susceptibility to earthquakes.

The following principles (Bolt 1993) generally help in avoiding the drawbacks in the zoning maps:

a) The map should not only take into account the size of an event but also the frequency of the earthquake in the region.

b) The zoning should be based on the historical seismicity, major tectonic trends, intensity attenuation curves, and intensity reports.

c) Contours should preferably be smooth so that expected accelerations can be read from the map even by extrapolating.

The Rhine rift being a part of the Alpine orogeny is one of the seismic risk zones in the Northwest Europe. Some of the earthquakes that affected this area were very destructive in nature. The man made buildings were destroyed in the city of Bale in Switzerland due to the 'big' earthquake of 1356. Even if the Rhine rift is associated with a moderate level of seismicity (less than 10 events with $M_L > 3.0$ each year), several seismic events could be observed recently, especially in the faulted hills edging the graben. In these faulted zones some noticeable earthquakes were recorded in the last fifteen years. Their magnitude reached to even 4.5 on the Richter scale. Hence the need for demarcating the most hazard areas in the graben and its neighbourhood has been felt and a seismic microzonation analysis is undertaken in the Rhine Valley and its surroundings.

Seismic hazard zonation can be done using peak ground acceleration, intensity of ground shaking, site response estimation, etc. Site effects in seismic regulations, land use planning or design of critical facilities is an important goal of earthquake hazard reduction programs wherein the effects of surface geology on ground shaking related to wave propagation phenomenon must be measured or estimated. There are two approaches for the estimation for site effects: either through direct in situ observation and/or measurements, or through numerical modelling based on the available geotechnical information. Field and Jacob (1995) performed detailed comparisons of various site response estimation techniques. The most common procedure consists in comparing records at nearby sites through spectral ratios if the reference site is free of any site effect. In practice, adequate reference sites are not always

available. Methods have, therefore, been developed that do not need reference sites. Langston (1979) used the "receiver function" technique to determine the velocity structure of the crust from the horizontal to vertical spectral ratio (HVSR) of teleseismic P-waves. Nakamura (1989) proposed to use this HVSR ratio on ambient noise recordings. Another attractive approach is to determine the site response relative to the network average. Different versions of this method have been proposed by several authors. In the present analysis, in the absence of a reference site and 3-component recordings, we have designed a non-reference network and event average technique for the estimation of site factors in and around the Rhine graben.

The present paper does not provide an exhaustive survey of the seismic hazard in the area but presents a new approach using network average for the zonation of the Rhine rift. The contour maps at different frequencies highlight the higher and lower site factors (SF) thereby indicating the seismicity and probable hazard in various zones.

Geology of the area

Our area of investigation comprises the Rhine graben and its neighbouring Vosges and Foret-Noire regions limited between the latitudes 47 – −49° N and the longitudes 6 – −9° E. The Rhine graben is a part of the rift system of the North West Europe between France and Germany. It is in the hinterland of the Alpine orogeny due to the collision between the Eurasian and African plates during the Cenozoic time. The major geological features in and around the Rhine graben are shown in the Fig. 1.

The Rhine graben is covered by Tertiary fluvial deposits, the thickness of which has reached several hundred meters at places. This is overlain by superficial Quaternary deposit transported by the Rhine River. The cities of Strasbourg and Mulhouse are situated in the graben. To the western part of the graben lies the Vosges Mountain comprising gneiss to sandstone ranging in age from Precambrian to Permian. Between this rocky mountain and the Rhine valley, small hills composed of Palaeozoic rocks, mainly sandstone are lined up paralleling the graben axis. Triassic marl areas are also exposed at places. This region is largely faulted and a major normal fault represents the boundary between these hills and the Rhine plain. Eastern part of the N-S trending normal fault, separating the valley from the "Foret-Noire" Mountain, is covered with marl and Permian sandstone rocks.

Data source

The "Fossé Rhénan" earthquake-monitoring network (FOSR) has been operating in the Rhine graben and its surroundings for the last 20 years. The present analysis is based on the recordings of 16 earthquake events by 16 seismic stations. At each station is installed one short period single component seismometer belonging to the "Ecole et Observatoire des Sciences de la Terre", Strasbourg, France and the "Survey Baden-Wurttenberg", Freiburg, Germany. These data are monitored, managed and put in the public domain by the "Réseau National de Surveillance Sismique" (RéNaSS), Strasbourg, France.

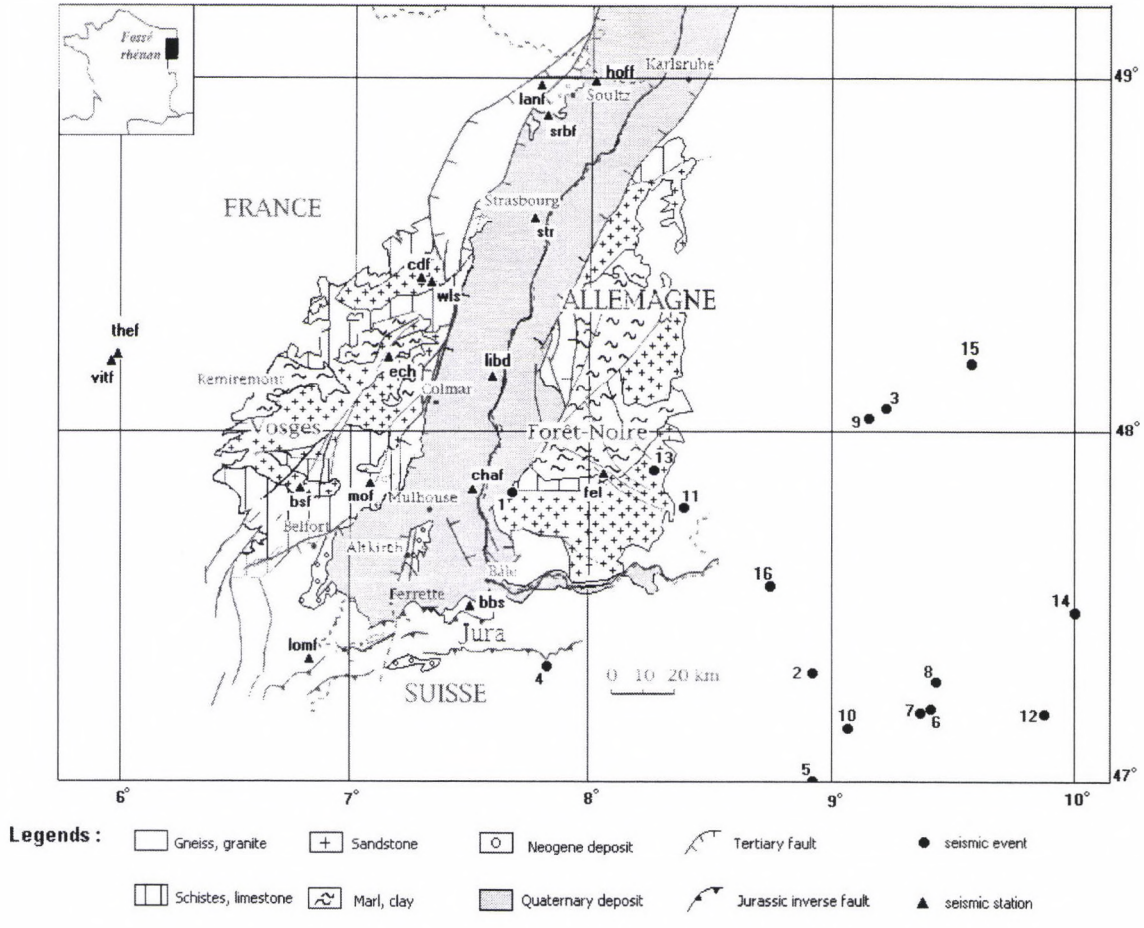


Fig. 1. Geological map of the Rhine graben and its surroundings with the stations and events marked

The stations are scattered on various geological formations in the area, in the Rhine valley or on the edges of the graben as shown in the Fig. 1. Of many hundreds of local earthquakes recorded during 1988–1995, only 16 events could be identified whose local magnitude lies between 3 and 4.5 with good signal-to-noise ratio. The distribution of the epicentres of these events is given in Fig. 1 with the details in Table I. Table II presents the earthquake recording history by the FOSR network (not necessarily each station recorded every event).

Data processing and site factor estimation

Suppose J events have been recorded by a network of I stations (each event may not be recorded by all I stations). Then the amplitude spectrum of the jth event recorded at the ith station for the kth frequency, $O_{ij}(f_k)$ can be written in the frequency domain as a product of a source term $E_j(f_k)$, a path term, $P_{ij}(f_k)$, and a site effect term, $S_i(f_k)$ (Field and Jacob 1995):

$$O_{ij}(f_k) = E_j(f_k) \cdot P_{ij}(f_k) \cdot S_i(f_k)\,. \tag{1}$$

It is to be noted that the term $E_j(f_k)$ depends not only on the radiation characteristics of the source, but also to a minimum extent on the station position as the energy is radiated with different intensities in different directions. However, in the present context the station dependency is negligible as network average is considered.

Taking the natural logarithm, Eq. (1) becomes:

$$\ln O_{ij}(f_k) = \ln E_j(f_k) + \ln P_{ij}(f_k) + \ln S_i(f_k)\,. \tag{2}$$

This linear expression often forms the basis of separating the source, path, and site effects.

The traditional spectral ratio is estimated with respect to a 'reference site'. Suppose, we have a reference site ($i = R$) that we assume to have a negligible site response ($\ln S_R = 0$). If the interstation spacing is small compared to the epicentral distances, so that $P_{ij} \sim P_{Rj}$, then the site response at each site can be estimated from,

$$\ln S_i^{SR}(f_k) = \frac{1}{J}\sum_{j=1}^{J} \ln\left(\frac{O_{ij}(f_k)}{O_{Rj}(f_k)}\right) = \frac{1}{J}\sum_{j=1}^{J}(\ln O_{ij}(f_k) - \ln O_{ij}(f_k))\,, \tag{3}$$

where J is the number of events recorded at both site I and the reference site (which may vary between sites), and f_k is the frequency at which we are computing the spectral ratio. Equation (3) constitutes the geometric average spectral ratio. If the reference site has a non-negligible site response, then the spectral ratios become relative site-response estimates.

In the present investigation an appropriate reference site was not available. Hence we followed an alternative approach wherein the network average amplitude spectrum representing all the I sites for a particular event was taken as the

reference value for that event. We will now outline the basic processing steps of this scheme for the computation of site factors in the 1–11 Hz frequency band from the analysis of 16 events recorded by 16 stations.

The processing starts with the butterworth bandpass filtering of each seismogram within 0.5 to 15 Hz frequency range for each event. In Figs 2, 3, 4 (a) and (b) the sample raw and filtered seismograms at stations FEL, CDF and BSF are presented for the event FOSR199011221551 ($M_L = 3.5$). Next the S-wave envelope (S) and the background (B) windows are selected from the filtered signal. Estimation of site factor (SF) starts with the Fast Fourier Transform (FFT) of both the S-wave envelope and the background noise. The sample Fourier spectra at FEL, CDF and BSF for the S-wave window and the background are shown in Figs 2, 3, 4 (c) and (d) respectively. Let the S-wave amplitude and the background noise amplitude be $A_{ij}(f_k)$ and $B_{ij}(f_k)$ respectively. Then the signal amplitude spectrum at the frequency f_k is expressed as,

$$O_{ij}(f_k) = A_{ij}(f_k) - B_{ij}(f_k)\,. \tag{4}$$

Only those events whose S-wave spectra are greater than or equal to three times the background are chosen for the computation of site factor so that the influence of the background is completely eliminated.

The network average for I sites and for each j event is computed at the frequency f_k as,

$$NAv(f_k) = \frac{1}{I}\sum_{i=1}^{I} O_{ij}(f_k)\,. \tag{5}$$

The conditional SF at each i site for the jth event can be written as,

$$SF_{ij}(f_k) = \frac{O_{ij}(f_k)}{NAv(f_k)}\,. \tag{6}$$

Figure 5 represents the sample background corrected S-wave amplitude spectra, the network average spectra and site factor at two stations CDF and WLS for the event FOSR199011221551. These are intermediate outputs from Eqs (4), (5) and (6) respectively. Figures 5(a) and (d) represent the S-wave envelope spectra within 0.5–15 Hz (the frequency band of our interest) re-sampled to 30 FFT points at 0.5 Hz frequency interval for the stations CDF and WLS respectively. The network average spectra within the same frequency band and 30 FFT points is depicted in Fig. 5(c). The computed SF values at the stations CDF and WLS prior to frequency smoothening are given in Figs 5(b) and (e) respectively.

Next, a 5-point central frequency mean is taken to compute the mean-conditional SF at the frequencies of interest 1, 3, 5, 7, 9 and 11 Hz to reduce the randomness in the Fourier spectra. The mean conditional $SF(f_k)$ can be written as,

$$SF_{ij}(f_k) = \mathrm{Mean}(SF_{ij}(f_{k-2}),\, SF_{ij}(f_{k-1}),\, SF_{ij}(f_k),\, SF_{ij}(f_{k+1}),\, SF_{ij}(f_{k+2}))\,. \tag{7}$$

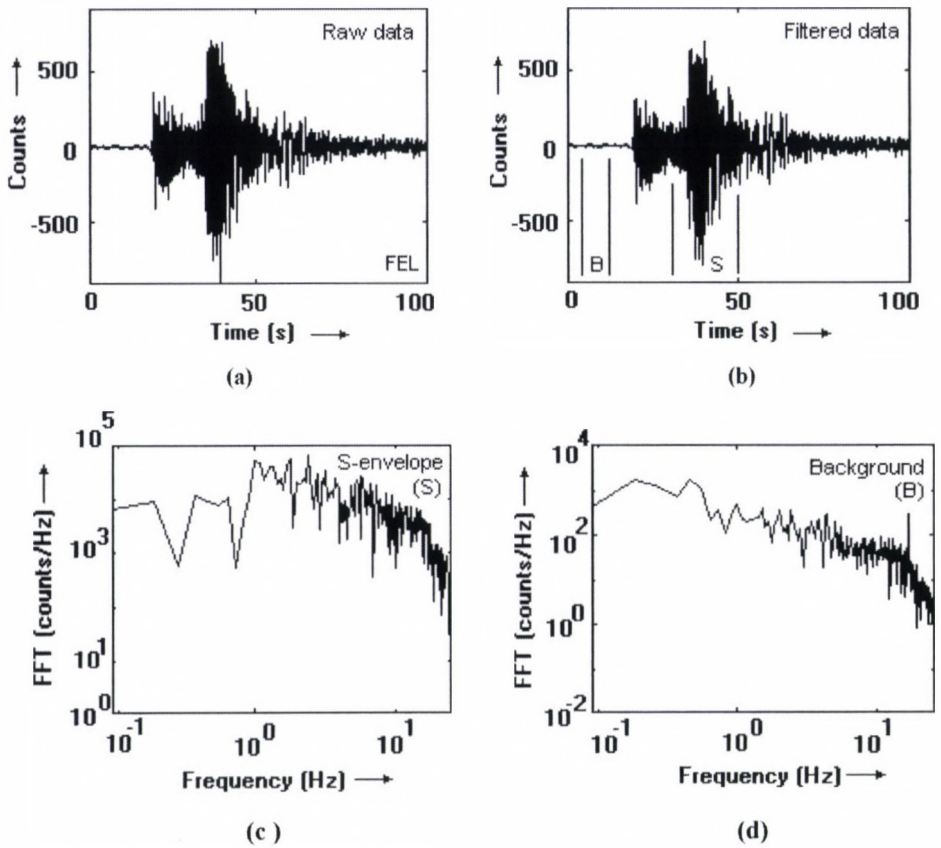


Fig. 2. Seismic event FOSR199011221551 with $M_L = 3.5$ recorded by the station FEL: a) Raw data, b) Filtered data within 0.5–15 Hz frequency band. Fast Fourier Transform of : c) S-wave envelope of the event, d) Background noise of the event

For 1 Hz, we averaged between 0.5 and 1.5 Hz only. The frequency averaging heavily smoothens the site factor values and the network average estimate eliminates the path effect. The sample mean conditional SF value plots against the recording stations at 1, 5 and 9 Hz for the events FOSR199011221551 ($M_L = 3.5$) and FOSR199212280208 ($M_L = 3.0$) are shown in Figs 6 (a, b, c) and 7 (a, b, c) respectively.

Finally the event average is computed to remove the source effect and construct the SF contour maps at each frequency. Hence, the event average SF can be written as,

$$\text{Final}\, SF_i(f_k) = \frac{1}{J} \sum_{j=1}^{J} SF_{ij}(f_k)\,, \tag{8}$$

where J is the number of earthquakes used for the analysis. In the present investigation J is equal to 16.

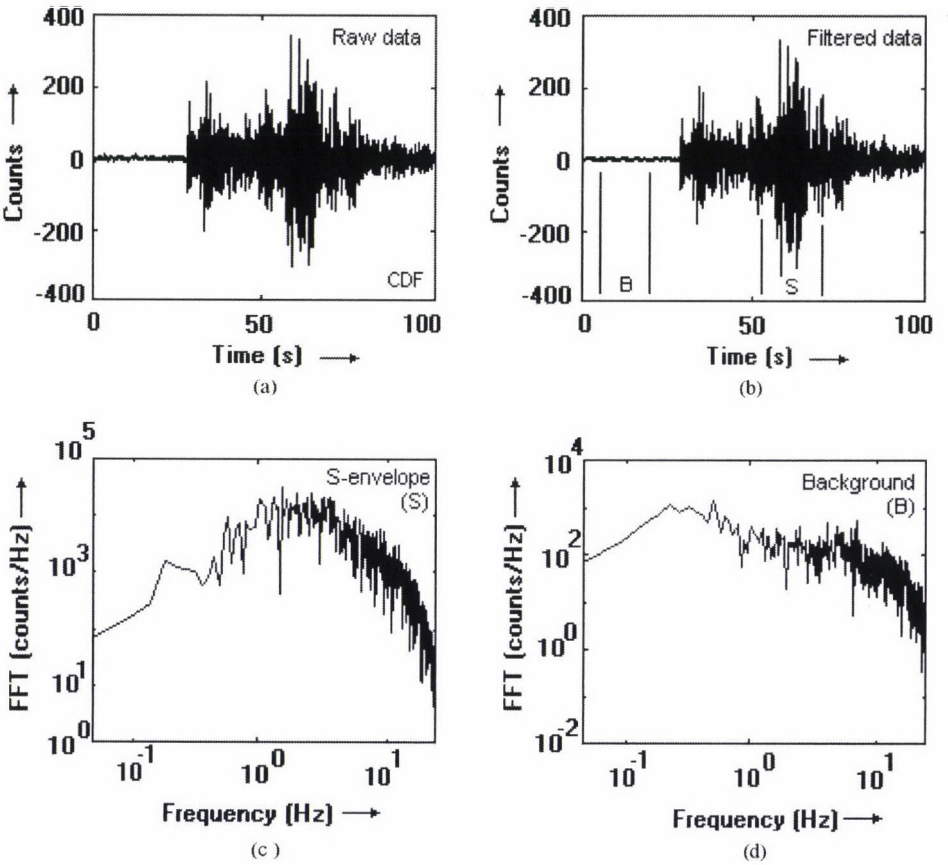


Fig. 3. Seismic event FOSR199011221551 with $M_L = 3.5$ recorded by the station CDF: a) Raw data, b) Filtered data within 0.5–15 Hz frequency band. Fast Fourier Transform of : c) S-wave envelope of the event, d) Background noise of the event

Using Eq. (8) we computed the event average site amplification factor at each site for the 1–11 Hz frequency band and plotted the final SF values on the map of the Rhine graben and its surroundings. The SF values are contoured and maps are prepared at the central frequencies 1, 3, 5, 7, 9 and 11 Hz which are presented in Figs 8 a, b, c, d, e and f respectively.

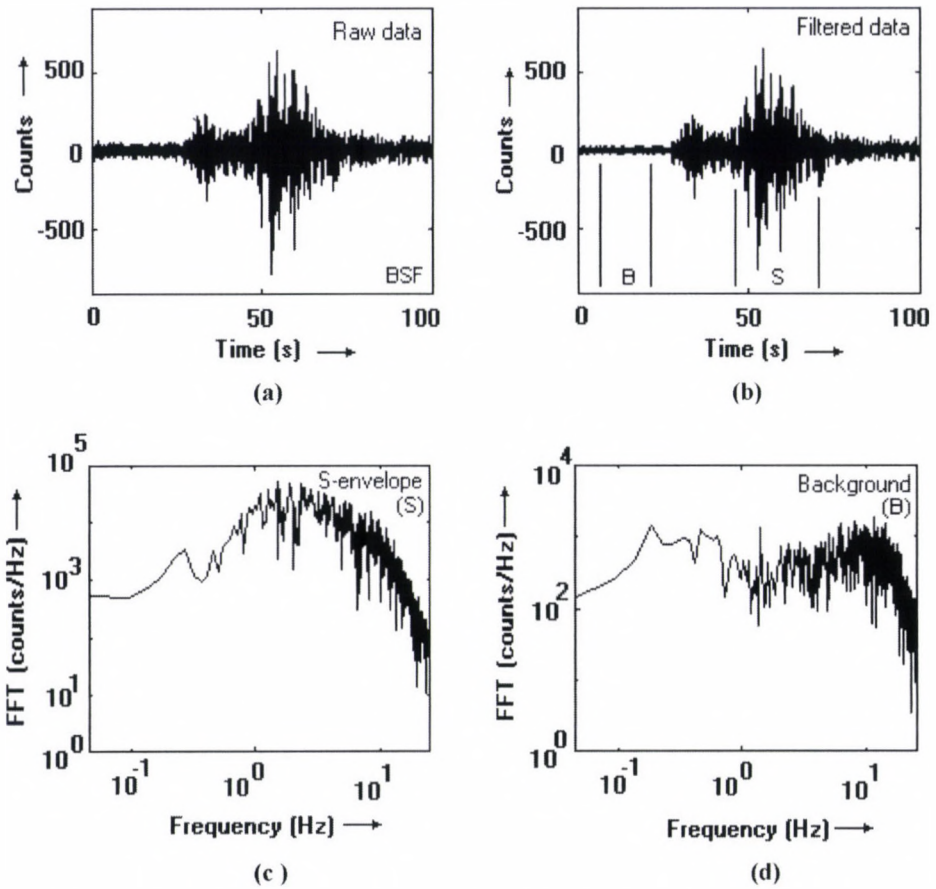


Fig. 4. Seismic event FOSR199011221551 with $M_L = 3.5$ recorded by the station BSF: a) Raw data, b) Filtered data within 0.5–15 Hz frequency band. Fast Fourier Transform of : c) S-wave envelope of the event, d) Background noise of the event

Discussion and conclusion

The stations HOFF, SRBF, STR, LIBD and CHAF are in the Rhine valley. Station FEL is situated in the "Foret-Noire" hills on the exposed marl deposit. To the West, the stations CDF, WLS, ECH, MOF, BSF, THEF and VITF are on different geological formations in the Vosges Mountain range and at the western edge of the graben. To the North there is only one station, LANF, while to the South of the graben there are two stations LOMF and BBS, with the later being very close to the city of Bale, which had an earthquake episode in 1356 of a major magnitude. The epicentres of the events are mostly distributed to the eastern side of the study area. Both the western and the eastern provinces (France and Germany respectively) are tectonically disturbed due to the Alpine orogeny and the rift tectonics. There are local and major faults criss-crossing the areas as shown on the geological map of Fig. 1.

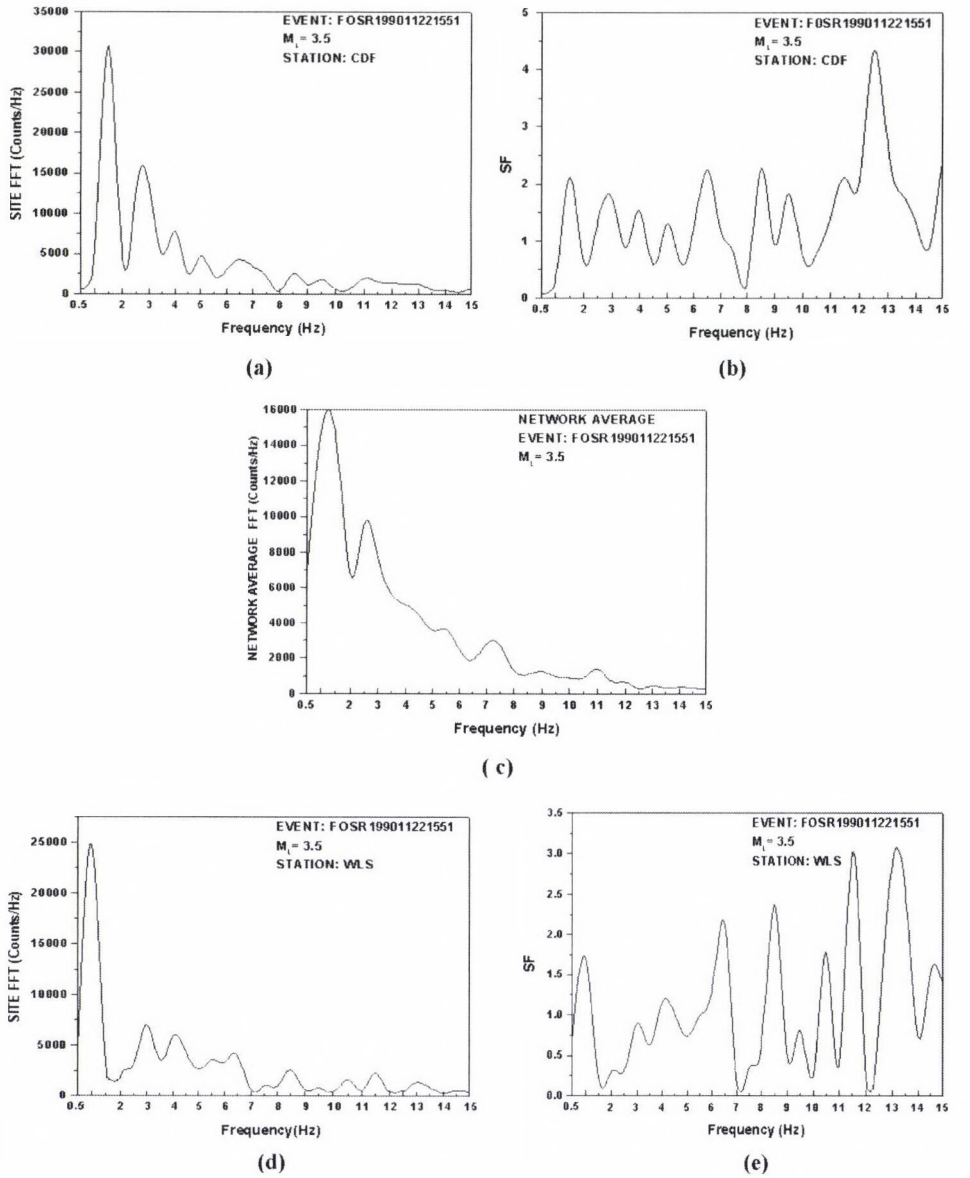


Fig. 5. a) Background corrected S-wave amplitude spectra within 0.5–15 Hz at site CDF for the event FOSR199011221551, b) Site factor at CDF prior to frequency smoothening, c) Network average S-wave spectra within 0.5–15 Hz, d) Background corrected S-wave spectra within 0.5–15 Hz at site WLS, e) Site factor at WLS prior to frequency smoothening

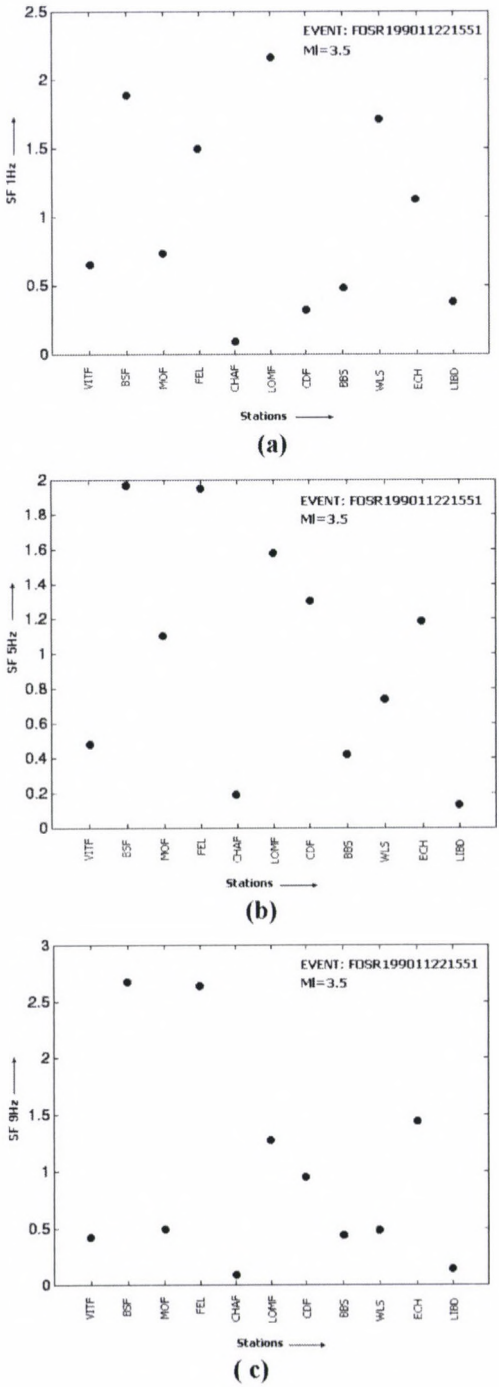


Fig. 6. Representative network average site factor versus station plots for the event FOSR199011221551 ($M_L = 3.5$) at a) 1 Hz, b) 5 Hz and c) 9 Hz

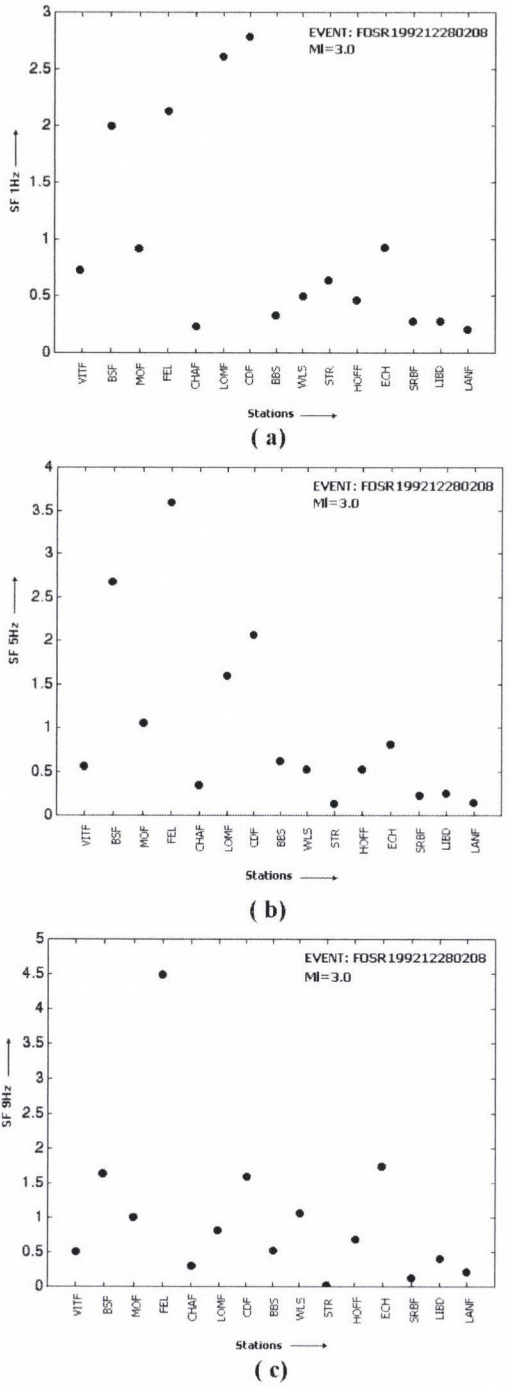


Fig. 7. Representative network average site factor versus station plots for the event FOSR199212280208 ($M_L = 3.0$) at a) 1 Hz, b) 5 Hz and c) 9 Hz

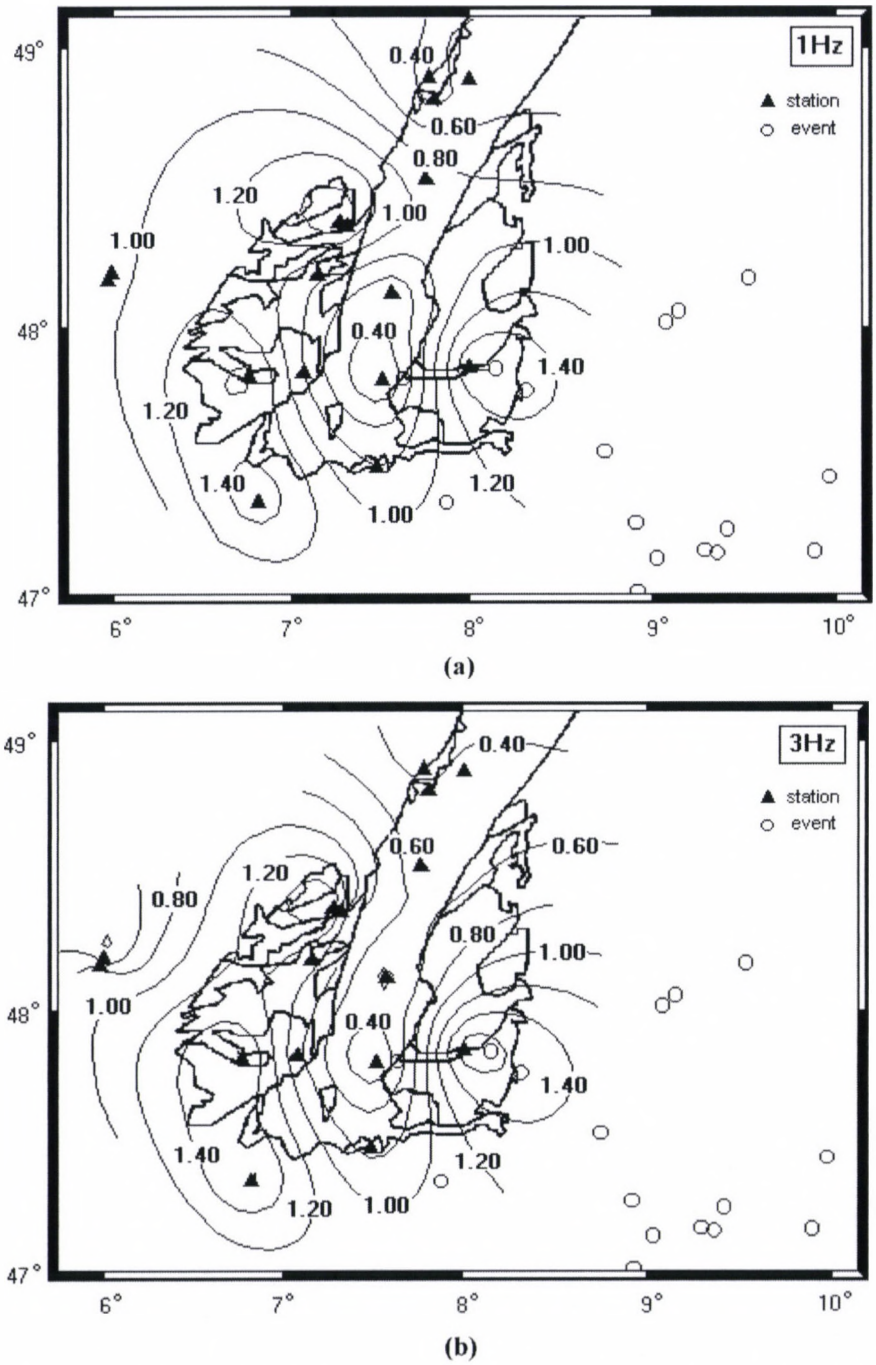
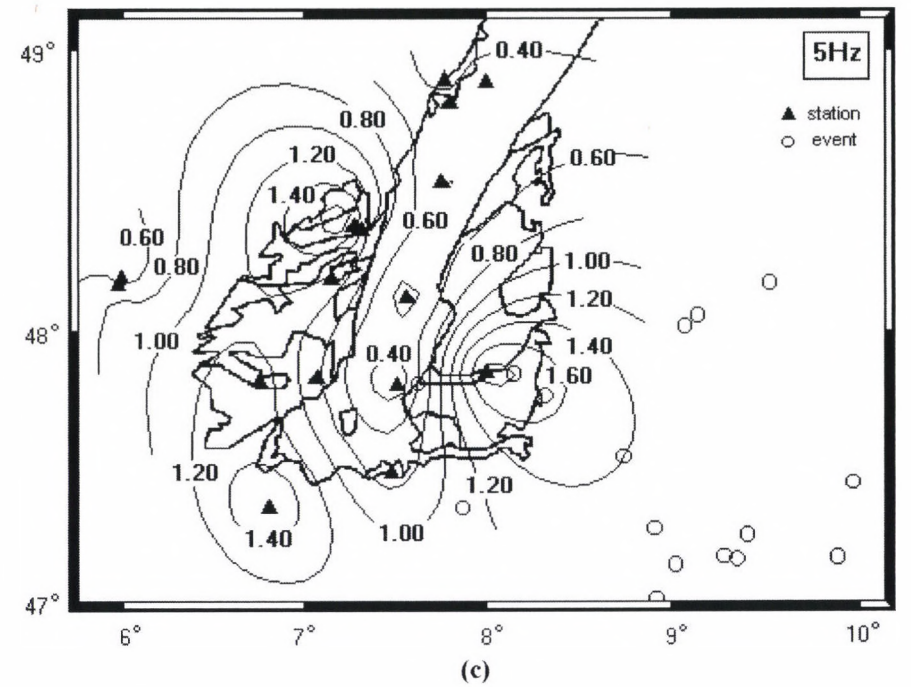
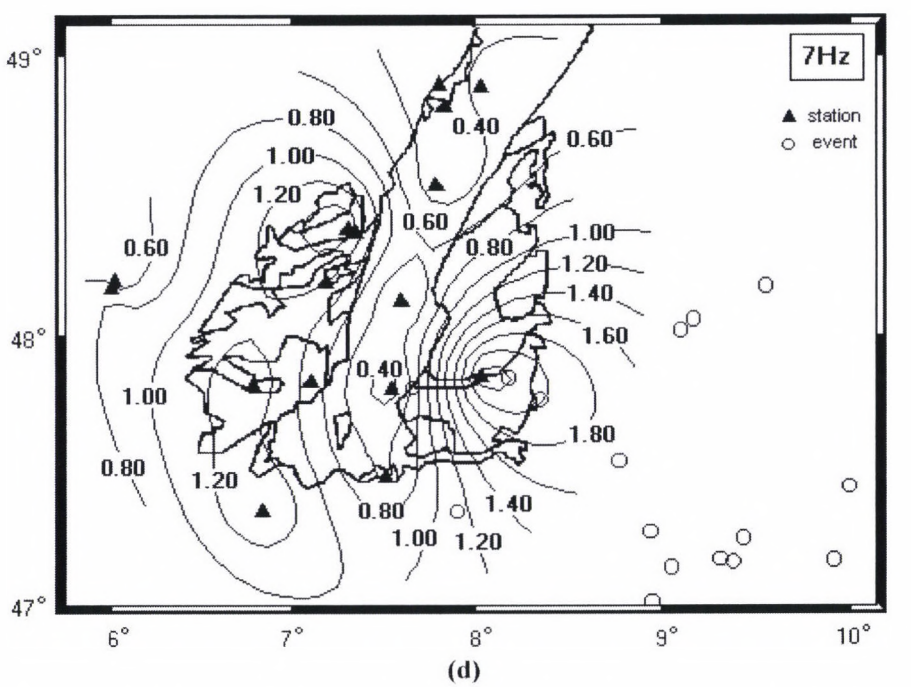


Fig. 8a. Final site factor (network and event average) contour maps in and around the Rhine graben at a) 1 Hz, b) 3 Hz

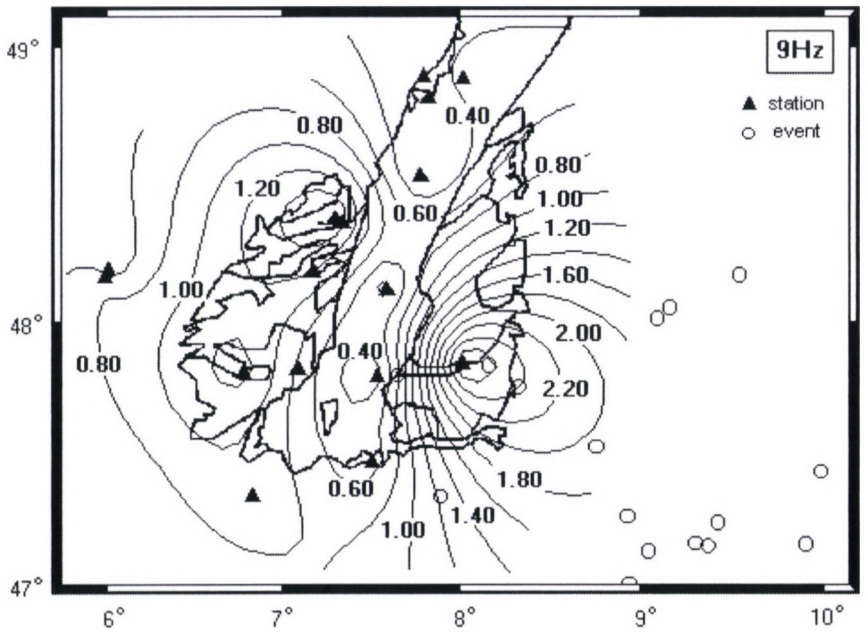


(c)

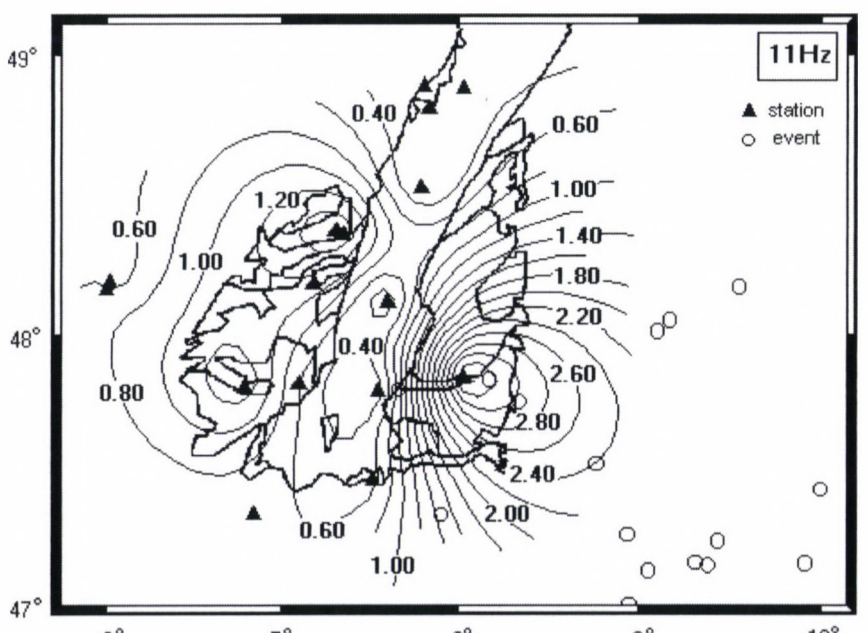


(d)

Fig. 8b. Final site factor (network and event average) contour maps in and around the Rhine graben at c) 5 Hz, d) 7 Hz

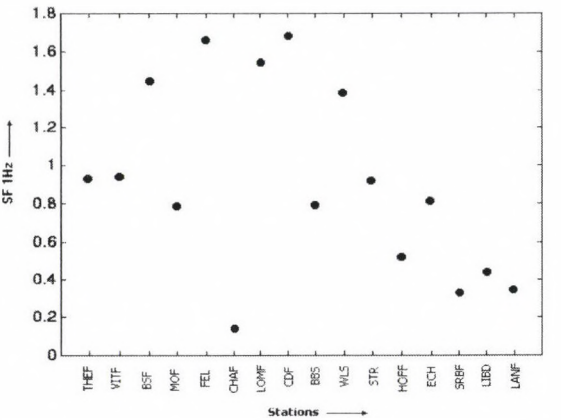


(e)

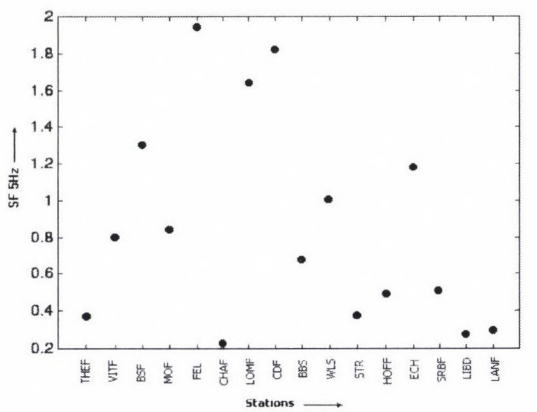


(f)

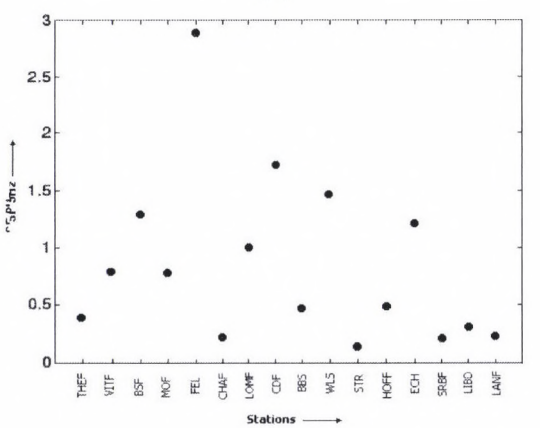
Fig. 8c. Final site factor (network and event average) contour maps in and around the Rhine graben at e) 9 Hz and f) 11 Hz



(a)



(b)



(c)

Fig. 9. Final site factor (network and event average) versus station plots for 16 earthquakes at a) 1 Hz, b) 5 Hz and c) 9 Hz

At 1 Hz (Fig. 8a (a)), the SF is found to be less than 1.0 in the graben but slowly increasing towards both the Vosges Mountain to the West and the Foret-Noire Mountain to the East. At site FEL which is situated on the marl and clay deposit, the SF value is found to be 1.6. At CDF also SF is more than 1.5. At other stations, except for those in the graben, the values are more than 1.0. It is interesting to note that site response values peaked along the major and the minor faults. In the graben we observe low SF values even though it is covered with thick layers of sediments. Near Strasbourg City, the site factor is 0.8, while near Mulhouse it is less than 0.4. At the station BBS, the value is 0.8. The station site factors at 1Hz are given in the plot of Fig. 9 (a). At 3 Hz (Fig. 8a (b)) the overall scenario remains the same. The contour patterns are also of similar nature with minor variations in the SF values. Station FEL is associated with the highest value of 1.6 and in the Rhine graben the lowest SF values are observed. The major faults and the minor tectonic features are associated with high site factors. At 5Hz (Figs 8b (c) and 9 (b)) the site effect at FEL is 1.8 with the overall pattern remaining the same as we obtained for 1 and 3Hz. At 7 (Fig. 8b (d)), 9 (Figs 8c (e) and 9 (c)) and 11 Hz (Fig. 8c (f)) also the site factor contour patterns remain unchanged with few variations in the site factor values. The amplification at the station FEL kept on increasing. The Rhine graben stations showed least SF values and the faults and fractures fringing the graben exhibited high site effects. As observed by Su et al. (1992) the network average heavily smoothens the site factor values thereby making them frequency independent.

The main variations of the SF values are observed along the eastern and the western faults demarcating the Rhine graben. Further the SF value increases on the tectonically disturbed provinces with the history of faulting and fracturing. The Rhine graben, on the other hand, exhibited lower site effects at all the frequencies. Since site factor is a representative of the response spectrum of the surficial geology (Mohraz 1976), the contour maps presented in this analysis can be treated as the seismic hazard maps at different frequencies. As it is obvious from the maps, the Rhine graben is seismically least hazardous compared to its surrounding eastern and western provinces. Irrespective of the magnitude of the earthquake, the ground motion in the Rhine graben is hardly amplified. The site factor in the present analysis is controlled mainly by the tectonism of the Alpine orogeny and the rift formation and least by the surficial deposits. However, the eastern exposed marl deposit on the Foret-Noire Mountain is associated with the highest site amplification values. From the above quantitative analysis we can, therefore, conclude that in the absence of reference site, non-reference network average estimate can be treated as an alternative method of computing the site factor and hence prepare the seismic zonation map of the area.

Acknowledgements

One of the authors (TP) is grateful to the Director, IIT Kharagpur for allowing him to do summer internship at IIT Kharagpur during the summer of 1999 and the Head of the Department of Geology and Geophysics for providing all the facilities to carry out this investigation.

References

Bolt B A 1993: In: Earthquakes. W H Freeman and Company, New York, 239-240.
Field E H, Jacob K H 1995: *BSSA*, 85, 1127–1143.
Langston C A 1979: *J. Geoph. Res.*, 84, 4749–4762.
Mohraz B 1976: *BSSA*, 66, 915–935.
Nakamura Y 1989: A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface, QR of R. T. R., 30–1.
Su F, Aki K, Teng T, Zeng Y, Koyanagi S, Mayeda K 1992: *BSSA*, 82, 580–602.

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 351–365 (2000)

THE TRADITIONAL RELATIONS BETWEEN FREIBERG AND SOPRON DURING 45 YEARS OF RESEARCH IN MAGNETOTELLURICS

G PORSTENDORFER[1]

[Manuscript received October 14, 1999]

Keywords: Freiberg, Germany; history; magnetotellurics; tellurics; Sopron, Hungary

The meeting on occasion of the 70th anniversary of Tóni Ádám, Pali Bencze and Ákos Wallner gives me the opportunity to remember to the roots of the investigations and productive exchange of ideas on the topic of telluric currents between Sopron and Freiberg.

As a guide the history of magnetotelluric investigations in Sopron was used (Ádám and Verő 1995). It was supplemented in a similar way by the corresponding investigations in Freiberg (Appendix).

In Sopron Professor Kántás recognized at the beginning of the fifties the possibility to use well-logging Picard-galvanometers with sensitivities of 10^{-8} to 10^{-9} A/mm/m for the recording of telluric currents. Already in 1954 he presented a telluric profile, measured in Hungary by an experimental instrument (Kántás 1954).

In 1952 Professor Meisser came back to Freiberg after some years of experience with self-potential measurements in Yugoslavia. In 1953 he introduced telluric investigations by the diploma-work of G Porstendorfer (Porstendorfer 1954). Also in Freiberg a high sensitivity mirror galvanometer of the VEB Askania Berlin was used as the heart of the two-channel recorder for the first telluric measurements in the Erzgebirge (Ore mountains), where a strong canalization and a high level of about 200 mV/km was found.

The first contacts introducing the friendly relations between Sopron and Freiberg result from meetings of Professors Kántás, Tárczy-Hornoch and Meisser on congresses at the beginning of the fifties. In the spring of 1957 Tárczy-Hornoch presented Toni Ádám at the Berg- und Hüttenmännischer Tag in Freiberg with a talk about a modified telluric prospecting — instrument and its application for telluric measurements of a great scale (Ádám 1958). In the harvest of 1957 scientists of Freiberg visited in the frame of a student excursion Sopron and the first contacts could be extended with the strong Hungarian team of young people, concerned in this time with telluric research, some of them already abroad in China. The epoch from 1954–1960 was the time, when both groups were occupied by investigations of telluric relative and absolute ellipses and their connection to horizontal inhomogeneities and anisotropy. In the high-ohmic surrounding of Freiberg we found a

[1]Technische Universität, Bergakademie Freiberg, D–09599 Freiberg, Gustav-Zeuner-Strasse 12

1217-8977/2000/$ 5.00

"micro-telluric" behaviour which came back in our days in the form of the "static shift"-problem and the corresponding distortion — effects during magnetotelluric investigations.

Ádám advanced in this time looking for a tensor — relation between the electric and magnetic field components, coming to his pioneer — work of magnetotelluric anisotropy (Ádám 1964). The years 1960/61 were another milestone in the history of our relations. Using the basic ideas of Tikhonov and Cagniard both groups started with magnetotelluric measurements. In Sopron the Schmidt-balance with photoelectric transformer was used as sensor for the magnetic field while in Freiberg high permeable coils were constructed. Later on both groups preferred small autocompensation variometers, using the span band technique of the high sensitivity galvanometers with small magnets.

Now the way was free for the beginning of large scale MT measurements in connection with the industry and other scientific institutions in Hungary and East Germany. Both groups were more and more concerned with questions of digitalization and using computers for processing and modelling but also with the interpretation of the conductivity anomalies meanwhile found in both countries.

Especially the year 1966 should be mentioned when our bilateral co-operation was enlarged in the direction of a multilateral co-operation of the countries of the eastern hemisphere in the frame of the KAPG. Ádám was the co-ordinator of the effective MT-group and already 10 years later the international noticed monograph "Geoelectric and Geothermal Studies" (editor Ádám 1976) could be presented on occasion of the meeting of the IAGA-group "Electromagnetic induction of the Earth and the Moon" in Sopron.

All people of the international electromagnetic community remember this meeting with its hearty atmosphere introducing the melting process in the relations between the two hemispheres leading 13 years later to the opening of the "iron courtain" also just here in Sopron.

It was the time when preparing our East-German contribution to the Sopron-workshop, when we joined our MT-results with the results of the group of Ritter of Geomagnetic Deep Sounding at the Geomagnetic Observatory in Niemegk. This group was founded by the pioneer of GDS, Horst Wiese. The map of Wiese arrows, which we showed at this time now can be completed by the results of numerous other German groups of Berlin, Potsdam, Braunschweig, Hannover, Göttingen, Münster, Frankfurt, München, Leipzig beside of Freiberg, but also of the Czech colleagues.

With the help of Ádám and Verő I had the opportunity to publish this map of the scalar amplitude of the transferfunction $|Z/H|$ in Acta Geod. Geoph. Hung. (Porstendorfer 1997). It belongs to a period range of 10–20 min, where imaginary parts are small or evanescent. By data of Rokityansky (1982) and Pěčová and Praus (1996) it could be continued in the direction of South-East-Europe until to the Carpathian region (Fig. 1).

Beside the typical known structures of the North-German-Polish and the Carpathian Conductivity Anomalies with the typical inversion of the Wiese-arrows along an axis with vanishing $|Z/H|$-values a further anomaly is visible in the central part of Germany. It shows only partly the Wiese-arrow inversion, but a distin-

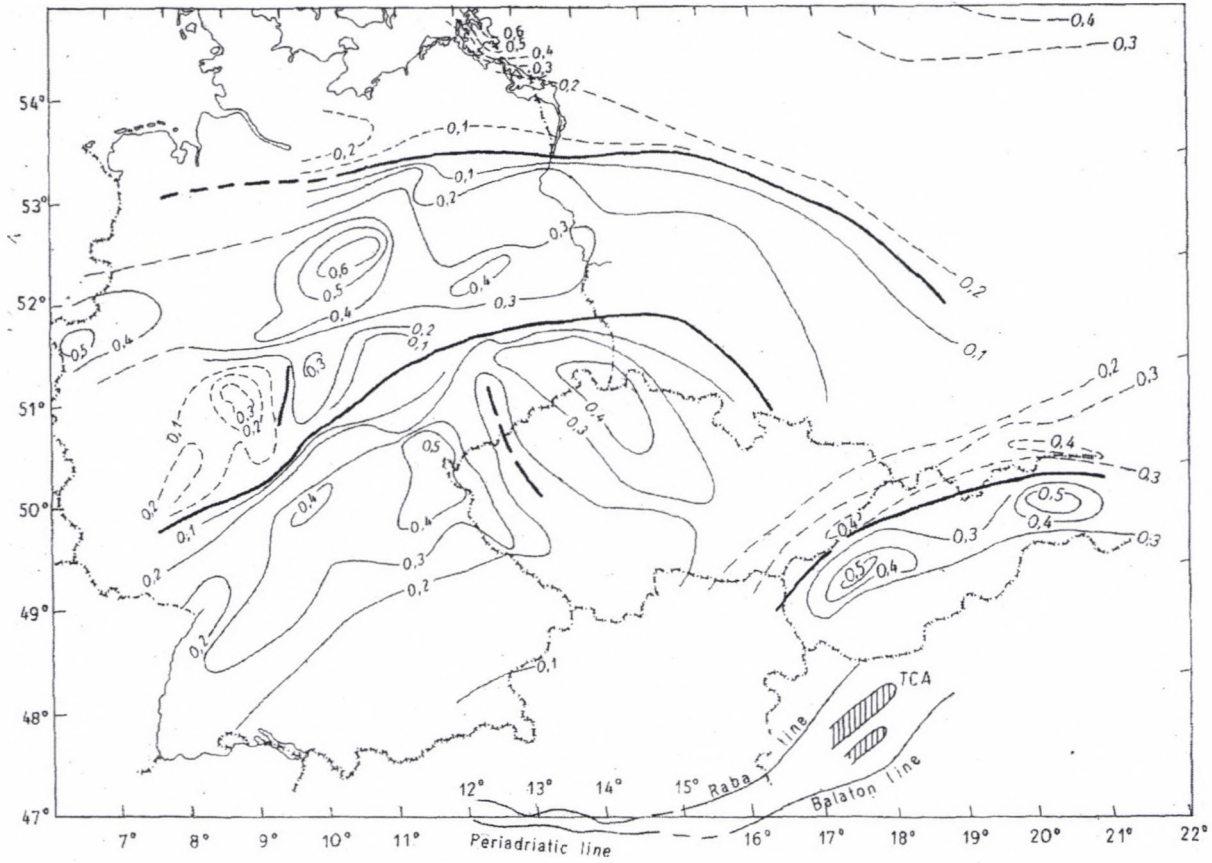


Fig. 1. Map of the scalar amplitudes of the transfer function Z/H for the period range 10–20 min (Porstendorfer 1997) combined by data of Rokityansky 1982 and Pěčová and Praus 1996. Broken isolines mark zones of significant change in the direction of the induction arrows implying a zero crossing along a detectable strike direction. Thick lines mark zones of minima of $|Z/H|$

guished minimal zone of $|Z/H|$ and a renewed growing of the $|Z/H|$-values to the South. After a maximal zone with values of 0.5 it slowly decreases into to direction of the Alps. To find the connection with the activities of the Sopron group during the last years the known structure of the Transdanubian Conductivity Anomaly and its probable continuation along the Periadriatic Line in the Austrian East Alps have been added (Ádám and Verő 1995).

Along the European Geotraverse from the Baltic Sea to the Alps selected MT-results from the eastern and western parts of Germany on two profiles are shown (Fig. 2). Starting from the Baltic Sea firstly the "old" MT-curve from 1964 is demonstrated (Fig. 3) in position A (Fig. 2). The curve shows the typical steplike form which was found meanwhile at many places of the North-German basin. Important is the introduction of a high conducting layer in the depth of 11 km. Figure 4 shows a new curve (ϱ_a and phase) at this place from measurements of the BGR-group (Hoffmann et al. 1994) in the ninetees with digital equipment and processing. It should be remarked that the phase can be used as additional indicator of the high conducting zone in the middle crust. Already 1968 we speculated about a possible cause of this layer by graphitic material (Naake et al. 1968) and indeed at the hole G14 in the Baltic Sea recently a layer of a thickness of 30 m Silurian graphitic blackshales (Alaunshales) could be found in a depth of 1600 m. The most recent interpretation of the BGR-group of the Rügen-profile (Hoffmann et al. 1998) (Fig. 5) connects the high conductive layer of Rügen with these blackshales. They belong to the Baltic shield and possibly they are now the thrust zone where formations of the Caledonian orogenesis have been shifted over the Baltic Shield.

But the behaviour of these typical MT-curves is also found in the southern part of the North-German basin (Fig. 6) which is shown in the section of position B (Fig. 2) (Hoffmann et al. 1998). In this case the interpretation of the layer in a depth of about 10 km is also graphitic blackshales, belonging to the Rhenoherzynian part of the Variscan Orogenesis. They were found in the hole Münsterland 1 in a depth of 5400–5500 m in the Lower Carboniferous. Possibly they assist as a horizon where the Rhenoherzynicum has been overthrusted above the Pre-Permian Foreland Basin. These blackshales play meanwhile an important role as possible mother rocks for gas generation (Hoffmann et al. 1998).

Going further to the South we come now in the district of the border between Rhenoherzynicum and Saxothuringicum of the Variscan Orogenesis. It is a region, characterized by a minimum-zone of the scalar $|Z/H|$-amplitudes, followed by a renewed increasing of these values to the South along a belt from NE to SW in the middle part of Germany.

Adding the values of Rokityansky, Pěčová and Praus this zone possibly continues to SE in the direction of the Carpathian Anomaly. A comparison with the schema of geologic zones of the Varisticum of Kossmat indicates that the minimum zone of $|Z/H|$ corresponds with the Mid German Crystalline High-Zone (MKZ), possibly an old terrane inside of the Mid-European Ocean before the Variscan orogenesis set in during Upper Carboniferous time.

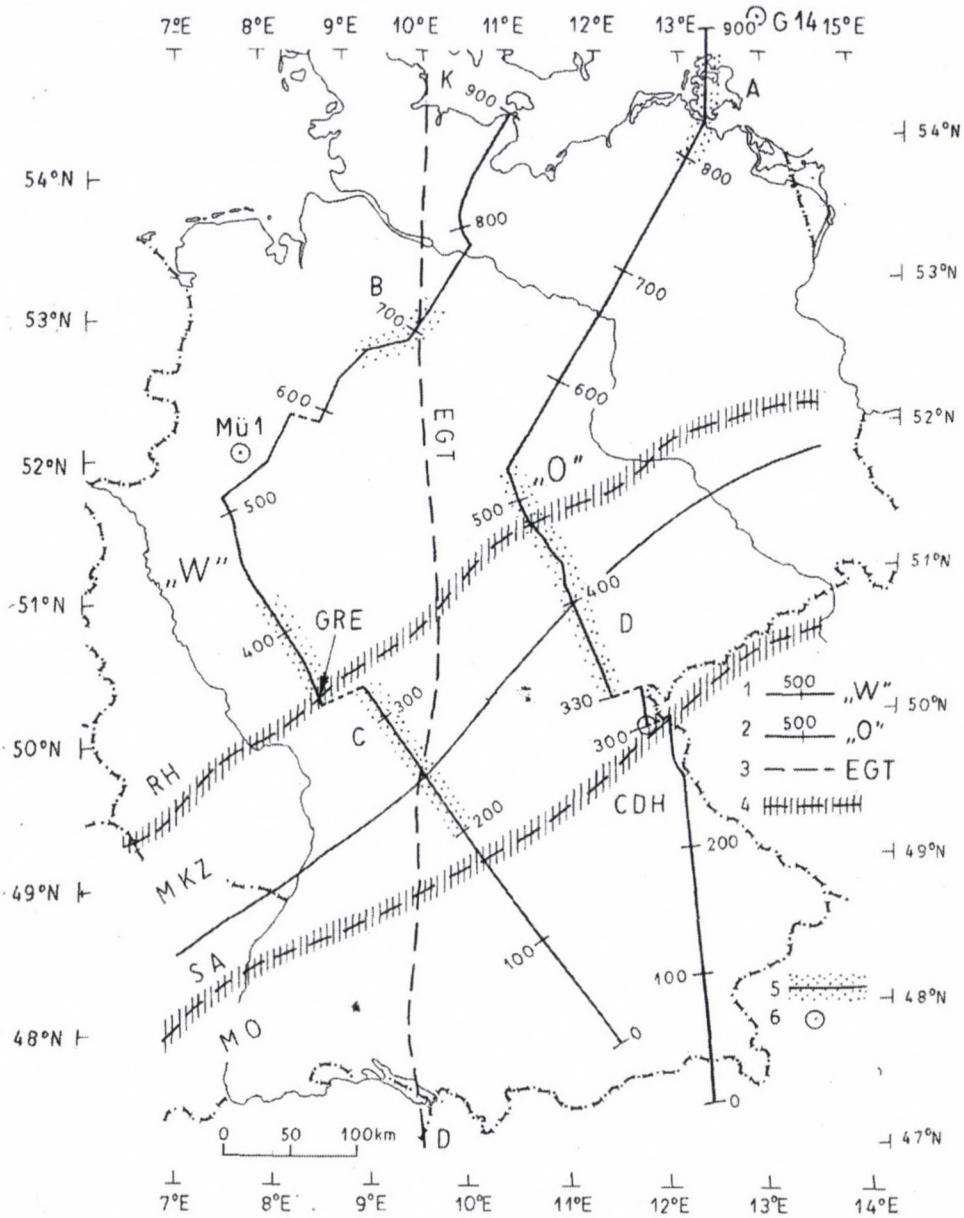


Fig. 2. MT-profiles along the European Geotraverse EGT. 1 – Western profile "W" of the ERCEUGT-group (Bahr et al. 1990), 2 – Eastern profile "O" by Porstendorfer (1997) with km-values, 3 – European Geotraverse, 4 – Border zone of the Variscan Kossmat-Zones, RH = Rhenoherzynicum, SA = Saxothuringicum, MO = Moldanubicum, MKZ = Mid German Crystalline Zone, 5 – MT-test areas A, B, C, D, 6 – Sites of deep holes, G14 = Baltic Sea hole, Mü1 = Münsterland 1, CDH = Continental Deep Hole, GRE = MT-site GRE

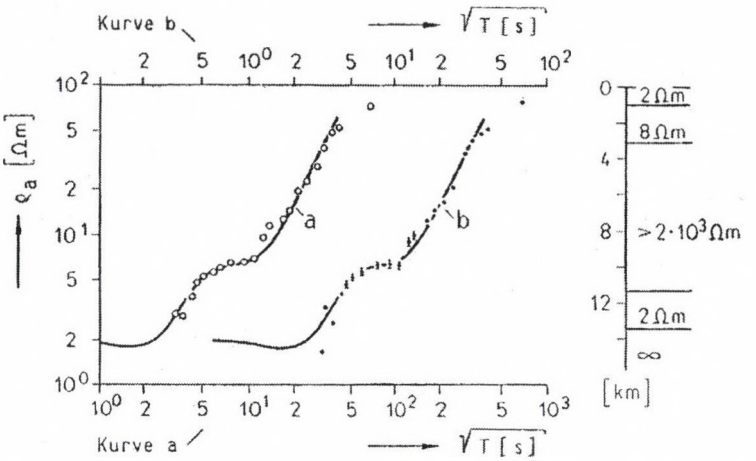


Fig. 3. MT sounding curve at Gingst/Island of Rügen and the interpretation by a five layer case (Porstendorfer 1965). Curve a: geometrical mean-values $\sqrt{\varrho_x \cdot \varrho_y}$ of the apparent resistivities in the azimuths NW-SE (x) and NE-SW (y), curve b: apparent resistivity in the Azimuth N 91° E of the predominant direction of the electrical field (≙E-polarization) (Porstendorfer 1965)

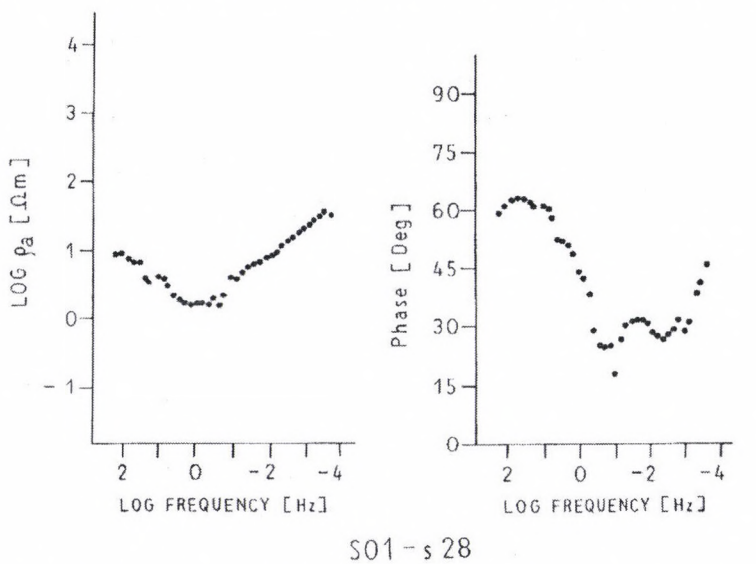


Fig. 4. MT curve in the neighbourhood of the sounding place of Fig. 3, taken by digital equipment and processing (Hoffmann et al. 1994)

The question arises, why the high $|Z/H|$-values in the South of the MKZ decrease so slowly to the South. Eisel (1990) gave an interesting interpretation by a conductive layer starting at the KTB (CDH) in a depth of about 10 km, which becomes more and more conducting to the North, reaching values of 10.000 S in the region of the MKZ. Figure 7 shows a discrete model for the section C (Fig. 2) of Hofer (1990) in the western part of Germany with a typical MT sounding curve

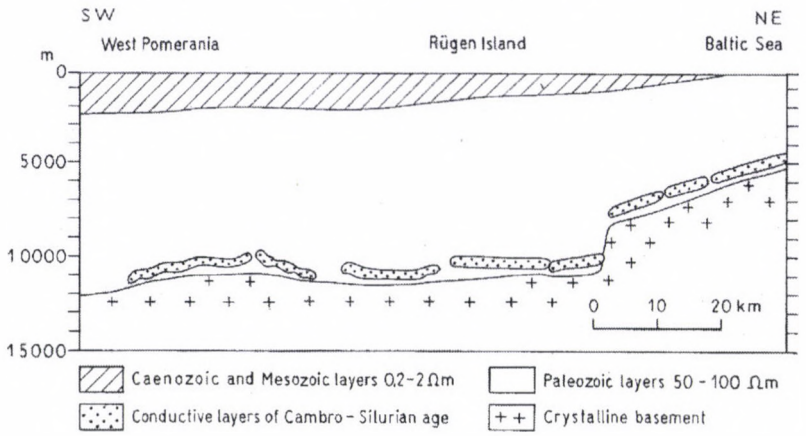


Fig. 5. MT-profile of Zone A (Fig. 1) on the Island of Rügen (Hoffmann et al. 1998)

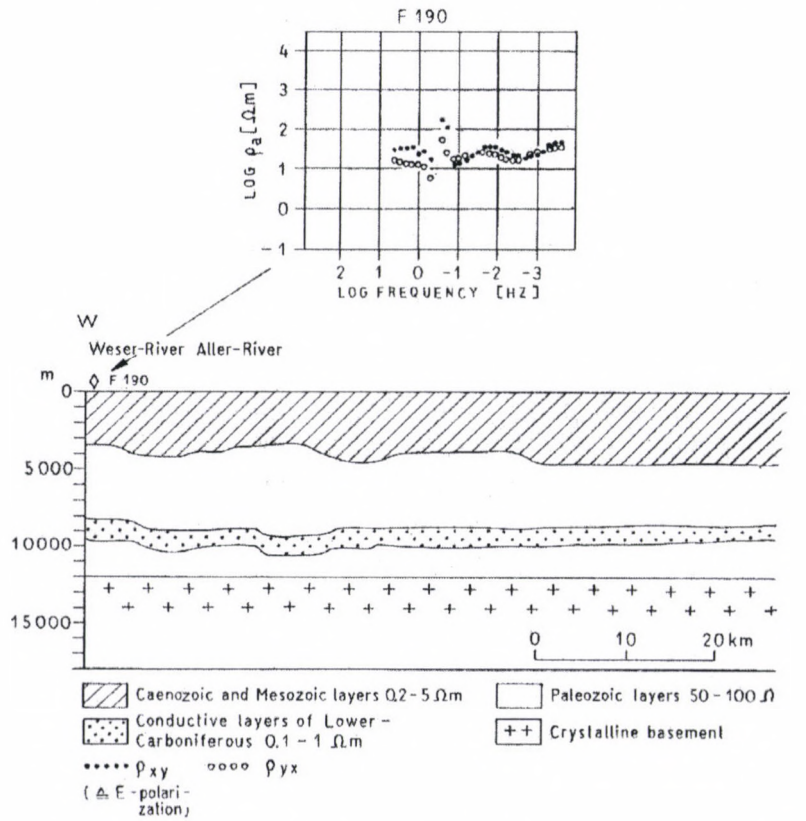


Fig. 6. MT-profile of Zone B (Fig. 1) on the South region of the North-German Sediment Basin after Jödicke (Hoffmann et al. 1998)

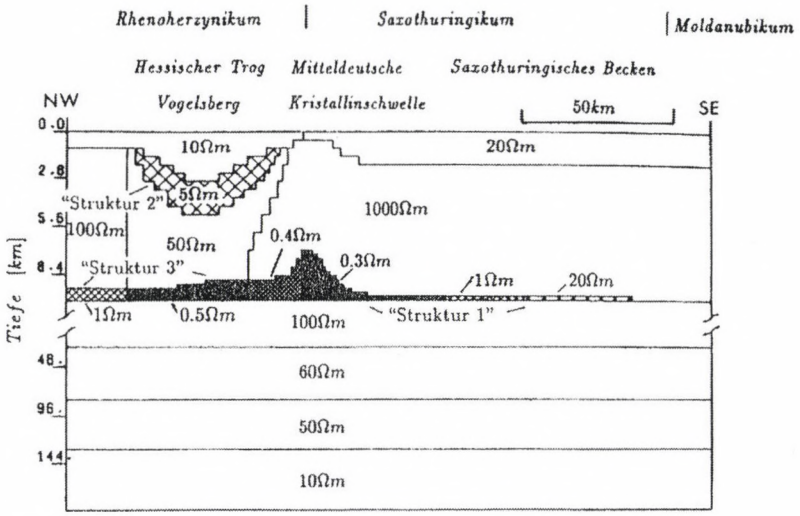


Fig. 7. Possible final FD-model along a profile crossing the Varistican Zones Rhenoherzynicum and Saxothuringicum in West-Germany (Hofer 1990) (Zone C in Fig. 1)

GRE (position Fig. 2) above the MKZ (in comparison to a similar MT curve M41 within the TCA in Hungary) (Fig. 8). Figure 9 shows a result of Börner (1995) for section D using the inversion technique of Smith and Booker (RRI) in the eastern part of Germany. Two interpretations are imaginable for the high conductivities inside of the Saxothuringian Crust. During the closing process of the Variscan orogenesis processes of dehydration within the amphibolithic facies could happen (Porstendorfer 1997). But in this case we have to introduce very high values of mineralization of the corresponding fluids. Petrologists think, that such high values of mineralization could not be stable in the lower crust (Bucher 1999). So the interpretation by layers of high carbonification until to the graphitic stage with electronic conduction remains as an alternative. The process of carbonification could have been supported by the additional thermal impulses of the granitic intrusions 340 to 330 millions years B.P. (Early Carboniferous). Imaginable could be also in this case the process of overthrusting of Saxothuringicum over Rhenoherzynicum along the graphitic black shale horizons. As a conclusion we can postulate that in the wells G14, Münsterland 1 and KTB but also at the surface in the Ziegenrücker Mulde (southern part of section D (Fig. 2)) (Pretzschner 1993) graphitic layers have been found which show subsurface conditions high conductivity and characteristic frequency dispersions. Until now only attempts exist to distinguish between electronic and electrolytic conduction by including processes of induced polarization into the theory of MT (Pretzschner 1993).

If we go further to the South in the direction of the Alps we find another conducting zone, caused by porosity and fluids in the Molasse-basin north of the Alps (Berktold 1969), but the corresponding S-values do not reach the high values of the Saxothuringian region.

So finally we find the connection to Ádám's investigation along the Transdanu-

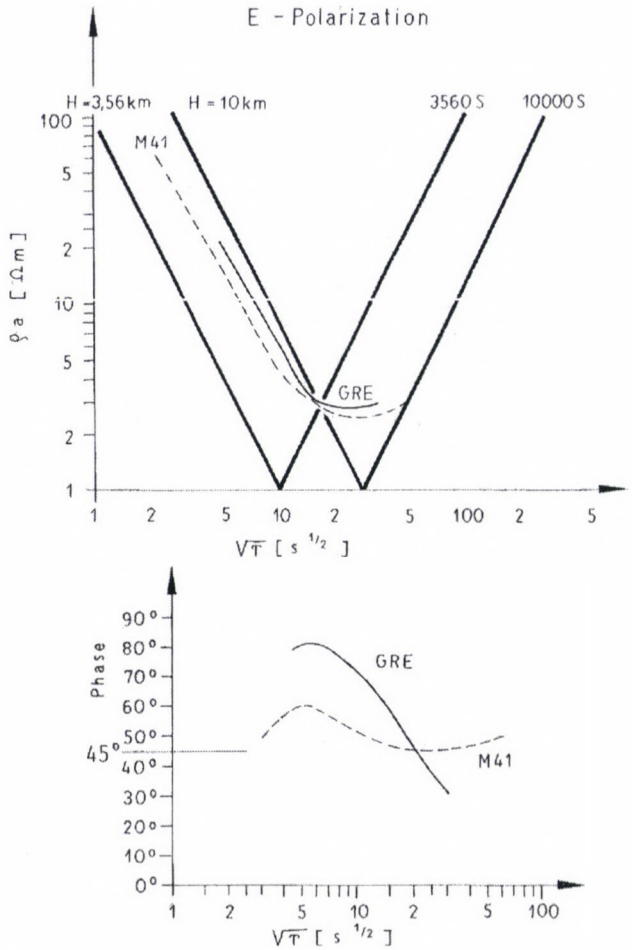


Fig. 8. MT-sounding curve of site GRE (Fig. 1) above the Mid German Crystalline High in comparison to a similar MT-sounding curve M41 of the TCA (after Ádám)

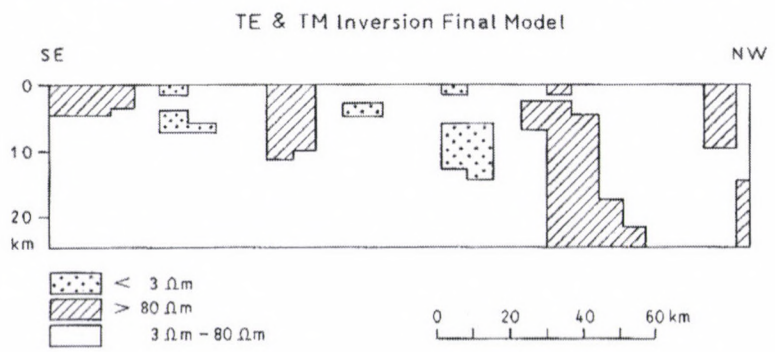


Fig. 9. MT-Inversion model along a profile crossing the Varistican Zones Rhenoherzynicum and Saxothuringicum in East-Germany (Börner 1995) (Zone D in Fig. 2)

bian Conductivity Anomaly and its probable continuation along the Periadriatic line, where the interpretation preferes now also graphitic blocks with its interesting consequences concerning the question of earth-quake risk.

Appendix

Events in telluric-magnetotelluric research at the Bergakademie Freiberg, Germany

1829–1830: Simultane D-variation-measurements in Freiberg (F Reich), Berlin (A v Humboldt), Paris and Kasan (Russia)

1953: Compilation of the bases of Tellurics (G Porstendorfer)

1954: First measurements of earth currents with an experimental equipment in Holzhau/ Erzgebirge (Tellurics, Bases and Applications, Diploma-work of G Porstendorfer)

1956: Start of active collaboration with G. Kunetz (Paris), M N Berdichevsky (Moscow) and A Ádám (Sopron)

1959: Utilization of stray currents by telluric vector diagrams. Application in mine districts

1960: Automatic circular soundings by 4-point-methods and in the electrolytical trough in comparison with "absolute ellipses" of Tellurics. Tellurics, Bases, Measurement technique and New Applications. (Thesis of G Porstendorfer)

1961: Start of magnetotelluric investigations in Mecklenburg by using high-permeable coils. Demonstration of Hungarian experiences in Tellurics and Magnetotellurics during a study trip of G Porstendorfer to Sopron, Budapest and Hungarian field parties.

1962: Telluric investigations in Mecklenburg (Diploma-work of G Thieme) Demonstration of Russian experiences of Magnetotellurics during a study trip of G Porstendorfer to Moscow (M N Berdichevsky) and Leningrad

1963: Construction of a 4-channel telluric recorder at the VEB Geophysikalischer Gertebau Brieselang after a basic concept of the Institute of Applied Geophysics Freiberg. Golden medal at the Leipzig Fair. Telluric measurements in Thuringia in collaboration with VEB Geophysik Leipzig (G. Porstendorfer). Magnetotelluric investigations on the Island of Rügen by using of autocompensation variometers of new construction (with G. Miersch). Detection of a high conducting zone in a depths of 11 km on the Island of Rügen. First magnetotelluric sounding in the crystalline-metamorphic region near Freiberg with indication of a high conducting zone in a depth of 80 km (Asthenosphere region). First magnetotelluric sounding near Auma/Thuringia with indications of conducting zones within the Varisticum.

1964: Magnetotelluric regional profile Arkona-Mid German Main Fault in collaboration with VEB Geophysik Leipzig

1965: Magnetotelluric soundings on the Island of Zanzibar (Tanzania/Africa) (G Porstendorfer and P Kuehn) and on Spitzbergen (Ch. Oelsner). Methodical and instrumental development of magnetotelluric methods with applications for deep-soundings in the region of the North-German-Polish-Conductivity Anomaly. (Habil. Thesis of G Porstendorfer). Demonstration of the results of Magnetotellurics in the GDR by a colloquium-talk of G Porstendorfer for Westgerman scientists at the Institute of Geophysics, TU Braunschweig (invitation by Professor Kertz)

1966: Extension of magnetotelluric investigations of the VEB Geophysik in NE- Mecklenburg. Demonstration of French experiences in Magnetotellurics by H Fournier during a study-trip of G Porstendorfer in France. Start of a comprehensive investigation of all deep well logs in the North of GDR with the aim to determine the longitu-

dinal conductivities and longitudinal resistivities for magnetotelluric interpretations (E Porstendorfer)

1968: Publication of the first work explaining the conducting crust regions by thin graphitic layers (L Naake, E Porstendorfer, G Porstendorfer)

1969: Methodical investigations for applying magnetotelluric methods in marine districts (Thesis P Kuehn). KAPG-meeting on Magnetotellurics in Freiberg. Presentation of the first magnetotelluric North-South profile of the GDR. Presentation of the first digital magnetotelluric equipment of the Institute of Applied Geophysics (Manthey). Start of digital processing of magnetotelluric sounding curves by statistical frequency analysis

1970: Talk of G Porstendorfer during the KAPG-meeting in Leningrad on "high conducting layers in the crust"

1971: Publication in "Gerlands Beiträge zur Geophysik" on "high conducting zones of the crust" (G Porstendorfer together with Berdichevsky, Vanyan and Feldmann). Guest lecture of G Porstendorfer on Magnetotellurics in Havanna/Cuba with first experiments on the application of radio waves

1974: Investigation of very deep holes in Mecklenburg with respect of material of high qualification (Ing.-work Wolf). Tellurics from running ships in the Baltic Sea (Thesis F Kuehn)

1975: Publication of the book "Principles of Magnetotelluric Prospecting" (G Porstendorfer) Start of regional compilations of results of MT-sounding in Europe (Diploma-work Seidel). Representation of a new FM-MT-equipment of the Institute of Applied Geophysics Freiberg with possibilities of Remote Reference MT by wireless transmission in Brandenburg (Thesis F Donner)

1976: Common elaboration of the results of Tellurics, Magnetotellurics and Geomagnetic Deep Sounding in the district of GDR together with the Niemegk group of GDS. Presentation on the JAGA-workshop in Sopron and publication within the KAPG-Monograph "Geoelectric and Geothermical Studies" (ed. A Ádám)

1977: Development of the methodics of finite-difference modelling. Modelling of a North-South-profile through the North of GDR enclosing anisotropic layers in the region of the deeper crust (Thesis Tanzer). Works to the influence of high conducting layers of the mantle to the magnetic secular variation (together with Mundt)

1978: Complex interpretation of magnetotelluric results along a North-South-profile through the GDR (together with Brause)

1979: Utilization of Radio Broadcasting Stations and VLF-fields by magnetotelluric principles and construction of a VLF-equipment (F Donner). KAPG-meeting in Dresden and Freiberg on the topic Magnetotellurics and Geothermy

1980: Works on possible connections of magnetic Satellite residual fields (MAGSAT) with magnetic secular variation anomalies and their interpretation by high conducting layers within the asthenosphere

1981: Estimation of the longitudinal conductivity of the North-German Sediment Basin from regional paleohydrogeological data (together with Hähne and E Porstendorfer)

1982: Regional MT-measurements in the Erzgebirge (Thesis G Legler)

1983: The connection of Magnetotellurics with the Transient-method in the near zone (together with F Kamenetsky). Studies to the theory of transfer-functions and impuls-answers of geophysical methods above layered media

1984: Initiative proposal for the new beginning of MT-measurements in Mecklenburg using digital technique

1985: Elaboration of a project proposal for the magnetotelluric investigation of the planet Mars and its moon Phobos (G Porstendorfer for the UdSSR and the Institute of Cosmical Research of the GDR)

1986: Complex interpretation of seismical ZENTROSEIS profiles using MT-results. Methodical investigations for the utilization of transient-soundings in the prospection of lignite (together with Goethe and Kamenetsky). Utilization of 16 2/3 Hz-Railway fields using MT-principles (Thesis Pawlick). Spectral IP-measurements in lignite open pit mining by using of a Phoenix digital equipment (together with Niesner/Leoben)

1987: Possibilities of a Digital-MT for the prospection of pre-Zechstein in the North of GDR (Thesis Lohse)

1988: New interpretation of MT-results in Mecklenburg – island of Usedom – as a base for new MT-measurements (together with Goethe and diploma-work Gentemann). Extension of the VLF-profiling-equipment of the Institute of Applied Geophysics for VLF distance sounding in the near zone of a VLF-transmitter (Thesis Salloum)

1989: Start of common MT-investigations along profiles from the Continental Deep Hole across the Frankenwald to the Harz (together with Schwarz/FU Berlin and Ritter/Niemegk) using a new digital MT-equipment of the Institute of Geophysics Freiberg and digital MT-equipment of the Institute of Geophysics/FU Berlin

1992: Compilation of all MT activities of the Institute of Geophysics/Freiberg in the region of the North-German sedimentary Basin (Case history) as base of a project of the "Bundesanstalt für Geowissenschaften und Rohstoffe" /Hannover concerning condenced MT-measurements by METRONIX digital equipment and GEOTOOL-processing in North-Germany with the aim to get information of high conducting black shales as possible sources of gas mother rocks.

1993: Connection of spectral IP-results in the near zone with possible applications of the IP in the Magnetotelluric (Thesis Pretzschner)

1994: Project-proposal to use the 1333 Ampere-BALTIC CABLE for transient prospecting in the region of the Baltic Sea. Demonstration of MT-results in East Germany at US colleagues of the US National Laboratory in Livermore/Cal. (invitation of G Porstendorfer by Professor A Duba)

1995: Inversion of magnetotelluric dates along a profile from the Frankenwald to the Harz (Thesis R U Börner). Compilation of the East- and Westgerman results of Magnetotellurics and Geomagnetic Deep Sounding with respect to a "Mid German Conductivity Anomaly" within the Varisticum (G Porstendorfer)

1996: Magnetotelluric investigations by digital MT-equipment along a profile crossing the Ural/Russia (Oelsner, Donner, Börner)

Events in relative telluric-magnetotelluric frequency soundings at the Geodetic and Geophysical Research Institute in Sopron, Hungary

1952: First telluric measurements in the Sopron Mts
1956: Comparative earth current measurements in Beijing geomagnetic observatory (China) and Sopron (Hungary)
1957: Earth current (telluric) recording started at the Nagycenk Observatory
1959–1969: Regional earth current measurements
1959–1964: Methodical and instrumental development of the telluric method (absolute ellipse from time derivatives, stabilization of absolute ellipses; total variation counter, analogue telluric models etc.)
1960: First magnetotelluric (MT) measurement near Sopron. Indication of a deep fault in the crystalline basement
1961: Methods for the determination of the MT anisotropy
1962: Indication of the Transdanubian crustal conductive anomaly (CA) by relative frequency soundings
1963: One of the first indications of the conductive asthenosphere by MT, below the hot Pannonian Basin it lies much shallower than below the cold East European platform
1963: "Study of the Electrical Conductivity of the Earth's Crust and Upper Mantle: Methodology and Results" (Ádám's PhD thesis)
1964–1970: Instrument development for telluric and MT measurement a) T-20 total variation variometer, b) MTV-2 high sensitivity variometer, c) Induction coils
1967–1969: Study of the dimension of the MT field. Synchronous electromagnetic measurements in Eastern Europe, organized by KAPG
1967: First detailed MT (resistivity) sounding of the Earth's crust and upper mantle down to about 400 km in Nagycenk.
1967: Start of the systematic study of the Transdanubian CA
1968: "Electrical Structure of the Upper Mantle in the Hungarian Basin. Questions of its determination and its Peculiarities" (Ádám's DSc thesis)
1969: Proof of a regional anisotropy in the Pannonian Basin
1971: "Spreading tectonics investigated by magnetotelluric anisotropy" (Nature)
1971: Proposal for the "first upper mantle MT E-W profile across Europe" (with Fournier in Acta Geod. Geoph. Mont. Hung.)
1972: Tensorial MT processing program
1976: Publication of a Monograph: "Geoelectric and Geothermal Studies", Akadémiai Kiadó, Budapest
1976: 3rd IAGA Electromagnetic induction Workshop in Sopron
1976: Relation between conductivity anomalies in the crust and upper mantle-asthenosphere and surface heat flow (improved 1978, 1980, 1987)
1977: Numerical models of 2D structures by finite difference method (Tátrallyay's PhD thesis)
1977: Analogue modelling starts
1978–1983: MT measurements in the Eastern Alps in co-operation with University Vienna and Institute of Meteorology and Geodynamics, Vienna
1980: MT measurements in the Baltic Shield, Finland along the SVECA profile in co-operation with Oulu University
1980: Review on the electric structure in an orogenic and in an ancient tectonic area (Carpathians and Russian Platform)
1982–1985: Time variation of MT resistivity at the Komló coalfield

1983–1985: On-line audiomagnetic instrument in Finnish-Hungarian co-operation
1984: Method for the determination of the underground structure by means of the analysis of artificial EM fields (Hungarian patent No. 186 678)
1985: "Role of the Magnetotelluric Method in Geophysical Prospecting", by Ádám, Hyderabad, India
1985–1990: MT measurements in the area of the Paks nuclear power plant
1988–1989: Study of EM noise
1988–1992: Audiomagnetotelluric study in the continuation of the Transdanubian CA in the Bakony-Drauzug collision zone (3 profiles)
1987–1989: MT measurements along the Pannonian Geotraverse
1987: Map on the depth distribution of conductive crustal formations in the Pre-Austrian basement of the Pannonian Basin
1988–1989: Relation between MT phase and current distribution and electric charges in 2D structures
1988–1993: 1D, 2D MT inversion and 2D, 3D forward programs (Steiner's PhD thesis)
1990: Ádám's inaugural lecture at the Hungarian Academy of Sciences, as corresponding member: "Transdanubian Conductivity Anomaly. Its Geological and Methodological Significance"
1991–1994: Study of the low viscosity graphitic conductor and its effect on seismicity
1993: Ádám's inaugural lecture at the Hungarian Academy of Sciences as ordinary member: "Electrical resistivity as Indicator of Physics of the Earth"
1994: Geophysical imaging of 3D structures by EM methods (Szarka's DSc thesis)
1994: Geomagnetic risk (Wesztergom's PhD thesis)

References

Ádám A 1958: *Freiberger Forschungshefte C*, 45, 52–61.
Ádám A 1964: *Freiberger Forschungshefte C*, 168, 1–48.
Ádám A ed. 1976: Geoelectric and Geothermal Studies (East-Central Europe, Soviet Asia) KAPG Geophysical Monograph, Akadémiai Kiadó, Budapest
Ádám A, Verő J 1995: History of EM-induction research at the Geodetic and Geophysical Research Institute in Sopron. Mitteilungen des Arbeitskreises Geschichte der Geophysik der Deutschen Geophysikalischen Gesellschaft, Vol. 14, No. 1
Berktold A 1969: Tiefensondierung mit Hilfe der Variationen des erdmagnetischen und erdelektrischen Feldes längs eines Profils vom Oberpfälzer Wald durch das nördliche Alpenvorland bis zu den Kitzbühler Alpen. Diss. Naturwiss. Fak. der Universität München
Börner R U 1995: Die Leitfähigkeitsverteilung der Erdkruste zwischen Frankenwald und Harz – abgeleitet aus der Inversion magnetotellurischer Messungen. Dissertation, Fak. f. Geowiss., Geotechnik und Bergbau, TU Bergakademie Freiberg
Bucher K 1999: Fluids in the Lithosphere. Festkoll. für Professor Dr. V. Haak, Potsdam
Eisel M 1990: Über die Superposition von lokalen und regionalen Leitfähigkeitsanomalien, untersucht anhand magnetotellurischer Messungen entlang eines Nord-Südprofils der Oberpfalz, Diplomarbeit, Universität Frankfurt/Main
Hofer S 1990: Ableitung der Leitfähigkeitsverteilung in der Erdkruste im Bereich der großtektonischen Grenze Rhenoherzynikum-Saxothuringikum östlich des Rheingrabens, aus einer gemeinsamen Interpretation flächenhafter Messungen der Magnetotellurik und der Erdmagnetischen Tiefensondierung. Dissertation Fak. für Geowissenschaften der Universität München

Hoffmann N, Fluche B, Müller W 1994: Erste Ergebnisse neuer magnetotellurischer Messungen in Nordostdeutschland. Kolloquium "Erdmagnetische Tiefenforschung", Höchst, 307–321.

Hoffmann N, Jödicke H, Fluche B, Jording A, Müller B 1998: *Zeitschr. f. Angew. Geol.*, 44, 140–158.

Kántás K 1954: *Bányászati Lapok*, 12, 1–6.

Naake L, Porstendorfer E, Porstendorfer G 1968: *Freiberger Forschungshefte C*, 233, 55–57.

Pěčová J, Praus O 1996: *Studia geoph. et geod.*, 40, 50–76.

Porstendorfer G 1954: *Freiberger Forschungshefte C*, 16, 1–104.

Porstendorfer G 1965: Methodical and instrumental development of magnetotelluric methods with application of the deep prospecting of the North-German Conductivity Anomaly. Veröff. des Instituts für Geodynamik Jena, H. 3, 1–76.

Porstendorfer G 1997: *Acta Geod. Geoph. Hung.*, 32, 135–150.

Pretzschner C 1993: Quantitative Modellierung des Einflusses der Polarisierbarkeit auf elektromagnetische Nah- und Fernfeldmethoden. Diss. Fak. für Math. und Naturwiss., TU Bergakademie Freiberg

Rokityansky I I 1982: Geoelectromagnetic Investigation of the Earth's Crust and Mantle. Springer-Verlag

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 367–372 (2000)

THE IMPORTANCE OF GEOPHYSICAL EXPERIMENTS FOR FUNDAMENTAL PHYSICAL RESEARCH

W SCHRÖDER[1] and H-J TREDER[2]

[Manuscript received February 10, 1999]

The authors discuss the interdependence between geophysics and fundamental physics both from the historical and from the methodological points of view.

Keywords: absorption of gravity; cosmic rays; figure of the Earth; gravity; tidal force

Aristotle made a distinction between 1. the ideal laws of heavens (translunar-sphere), 2. "$\tau\varepsilon\chi\nu\eta$", the knowledge of "simple machines" and their combinations and 3. "physics", the science which considers everything "sublunar"; i.e. all what occurs on and in the Earth itself — without human intervention.

Thus according to Aristotelian natural philosophy, "physics" and "geology" describe one and the same geophysics. Nicolaus Copernicus, Galileo Galilei and Johannes Kepler first postulated the the concept of the uniformity of laws of nature throughout the whole cosmos, and with the idea of a cosmic physics which includes astronomy, geology and technology. Galilei pioneered around 1630 the development of experimental physics, and one half century later Isaac Newton did the same for theoretical physics. According to the Aristotelian conception on the essence of physics, as the science of "natural occurrences" whose laws can be deduced from statements about the properties of terrestrial bodies and thus also from the generalisation of geophysical (and of cosmic-physical) statements i.e. "a principle of the generalised induction". On the other hand according to Galilei Physics is based on the "principle of exact induction". Its basis includes not only those occurrences taking place by themselves and on "evolving structures", but "ideal cases" are produced in laboratories where all terrestrial and cosmic influences are as far as possible excluded that they can be effectively neglected. Cosmic and terrestrial influences are then considered to be stochastic disturbances which in principle "cannot be explained and need no explanation" (Pauli).

Independently from the possibility to carrying out this idealistic programme of a purely local laboratory physics as "if it was isolated from the cosmos", geophysics, planetary physics and astrophysics can be considered as special applications of physical laws discovered in the laboratory by exact induction on special, natural objects of high complexity. Thus geophysics and astrophysics are in principle not able to deepen our theoretical insight into basic physical laws, and are simply fields of application. Geophysics and astrophysics only yield, in the sense of Newton's axiomatics,

[1]Hechelstrasse 8, D-28777 Bremen Roennebeck, Germany
[2]Rosa-Luxemburg-Strasse 17a, D-14282 Potsdam, Germany

1217-8977/2000/$ 5.00

initial and final conditions to the action of physical laws which are discovered in the laboratory.

However, this idealistic separation of the study of physical laws on the basis of exact induction from their application to given cosmic objects and their use for technical purposes does not correspond to the actual historical development of physics.

Even if we neglect all cosmological and cosmogonic problems which refer to universal initial and final conditions, some of which are not arbitrarily prescribed, geophysics has been important since Galilei and Newton in the history of physics in giving opportunity for critical physical insights in many fields.

Newton's theory of gravity originated from an effort to unify the Galilean laws of free fall and Kepler's laws of planetary motions in a single universal gravity law. According to Newton, the Galilean acceleration of free fall, g, is the field intensity of gravity of the mass M on the Earth's surface; thus

$$g = \frac{fM}{R^2}\,, \tag{1}$$

where R is the Earth's radius, M its mass and f the Newtonian gravity constant. The "active Earth mass" fM scaled by this law has to determine the acceleration of the Moon on its path around the Earth; according to the third Keplerian law it follows:

$$\frac{v^2}{r} = \frac{fM}{r^2} \rightarrow v^2 r = -fM\,, \tag{2}$$

where v is the velocity of the Moon relative to the Earth and r the Earth-Moon distance.

For many years Newton was not convinced by his gravity law (deduced from Kepler's laws), as the values from Eqs (1) and (2) were not compatible. The Earth radius R, as he knew it, was too small. A later value of the Earth radius from geodetic measurements made by C Picard resulted in a correction of the value of g and yield the values of fM from the Galilean acceleration g and from the acceleration of the Moon, v^2/r. Having resolved this, Newton was ready to let his "Principia Mathematica Philosophiae Naturalis" be printed.

Galilei's and Newton's notions about inertia and weight of matter, i.e. the notion of mass, led to theoretical considerations by C Huygens and later by Newton about the geometrical structure of rotating bodies which could only checked on a "cosmic scale". The determination of the flattening of the terrestrial body offered itself as an "experimentum crucis". According to the Newtonian principles, the terrestrial body is (to a first approximation) an ellipsoid flattened in the direction of the axis of rotation. The magnitude of this flattening, the deviation from the idealistic spheric form, can be *a priori* estimated according to Newton on the basis of the inertial forces of the rotation.

In opposition to the Galilean inertial principle and the Newtonian law of force, Cartesian physics was still advocated by some in the second half of the 17th century and in the first half of the 18th century. From its ideas, the Cassinis concluded a spindle-form for the Earth, contrary to Newton's thesis.

Around 1740, on initiative of the French Academy, geodetic expeditions were organised led by Ch M de la Condamine and P S Maupertuis, respectively, to measure the radius of curvature of the Earth in South America and in Lapland. Their results confirmed the Newtonian flattening of the Earth.

According to the Newtonian principles, it is possible to prove the rotation of the Earth. The axioms of mechanics tell us that the astronomical value for the angular velocity of the Earth corresponds exactly to its value computed from Huygens' centrifugal force. This postulate is known as the Huygens-Mach principle in General Theory of Relativity and of Cosmology ("Mach-Einstein doctrine"). The astronomer Foucault demonstrated in 1851 that the angular velocity of the rotation of the Earth can be determined in a closed laboratory. This is the same effect which is known in physical geography as "von Baer's law". The quantitative coincidence of the astrometric, geodetic and experimental determinations of the Earth's rotation has become, since Kammerlingh-Onnes (1879) and from ever improved geodetic and astrometric data, more and more acceptable, even if the classical problem of the "rotation of the cosmos" (Gödel-cosmos) which should be deducable from the rotation of the Earth, remained basically open.

The theory of the figure of the Earth, being a problem specially of geodesy and generally of geophysics, was formulated by A C Clairaut in 1743 on the basis of Newtonian principles. Clairaut, following Newton considered the motion of the Moon as the decisive test for the exact validity of the Newtonian principles of celestial mechanics and especially of the gravity law. Clairaut at first found a deviation of the actual lunar motion from the computed one. Euler concluded from this a necessary correction to the Newtonian gravity law, namely a screening effect. This gravity force $\overline{K}^*_{I,II}$ (Laplace, H von Seliger) is for the masses m_I and m_{II}:

$$\begin{aligned} \overline{K}^*_{I,II} &= \tfrac{f m_I m_{II}}{r^3_{I,II}} \cdot \overline{r}_{I,II} \cdot \exp(-\lambda \int \varrho dr) = \\ &= -f \tfrac{16}{q} \pi^2 \varrho_I \varrho_{II} R_I^3 R_{II}^3 \left(1 - \tfrac{3}{4}\lambda \cdot \varrho_I R_I\right) \cdot \left(1 - \tfrac{3}{4}\lambda \cdot \varrho_{II} R_{II}\right) \cdot \tfrac{\overline{r}_{I,II}}{r^3_{I,II}} , \end{aligned} \tag{3}$$

where ϱ_I and ϱ_{II} are the mass densities and R_I, R_{II} the radii of the two bodies. λ is a new universal constant, the even today debated, absorption constant of the "gravitational flux":

$$\Delta m = -\frac{3}{4}\lambda \varrho R m \,. \tag{4}$$

The more accurate computation of the figure of the Earth according to the exact theory proved, even at Clairaut's time within the limits of measurements accuracy, the complete validity of the Newtonian law. This validity was supported by an argument which went back to Laplace and was confirmed by geophysicists after 1900 with a very high accuracy:

Corrections to the Newtonian law of gravity in the form given by Laplace tell us that bodies in the gravity field fall with different velocities; the larger and more massive a body is, the slower it falls. On the well-known question to children: "Which is heavier: one kilogram of feathers or one kilogram lead?" the response would be: "one kilogram of feather" (see Thiessen and Treder 1982).

Geophysical methods for the determination of the figure of the Earth, with a reversion pendulum since F W Bessel (around 1840) and from the study of the relation of weight and inertia with a torsion pendulum as initiated by L Eötvös (around 1900), confirm with an increasing accuracy that the absorption coefficient λ has to be very small. According to present estimations (Steenbeck and Treder 1981):

$$\lambda \leq 10^{-15}\,\mathrm{g}^{-1}\,\mathrm{cm}^2\,. \tag{5}$$

Eötvös (1909), O Hecker (1907) and later A A Michelson and H L Gale (1920) confirmed this estimation by another, purely geophysical method: The gravity force of the Sun and of the Moon are not directly measurable according to the Newtonian principles, but the difference of the gravity fields, is:

$$G = \frac{fM_S}{r^3} \cdot R_{\mathrm{Earth}}\,, \tag{6}$$

where M_S is the mass of the Sun, r the Sun-Earth distance.

These are the tidal forces. If instead of the Newtonian law, the modified law, given by Eq. (3) is valid, then modified, bigger solar tides would result namely, using Eq. (4):

$$G^* = G \cdot (1 + \lambda \varrho_{\mathrm{Earth}} r) \tag{7}$$

(Steenbeck and Treder 1981). The actual determination of the solar tides (and especially their relation to lunar tides) yields again an upper limit for λ; nevertheless, problems of geodesy and gravimetry still emerge (Treder 1977).

The role of geophysics (and especially of gravimetry and geodesy) in experimental gravity research is complicated by the relatively large mass of the terrestrial body in comparison to all bodies at our disposal in a laboratory and the fact that the gravity field on the surface of a circular body increases linearly with its radius R. That is why further fine details of the Newtonian and Einsteinian gravity theories and the search for possible fine corrections to Einstein's general-relativistic gravity theory lead to geophysical problems, such as, on the one hand, to the measurement of the dependence of the value of the Galilean acceleration of free fall g on time of the day and year, and on the other hand, on the type of measurement method (whether "spring balance" or "free fall"): Steenbeck and Treder (1981) discussed this in their book. The difficulties are here connected less with the accuracy of the measurements themselves than with the estimation of other geophysical and geodetic influences on the measured values which are not of a purely stochastic character (hence the proposition to carry out such measurements on satellites).

A further important characteristic of the terrestrial body is its relatively large age of 4 to 5 billion years (practically the same, as that of the solar system). This age is comparable with the so-called age of the cosmos, namely 10 to 20 billion years, so that even if some fundamental physical constants or other elementary structures of the cosmos do change — as certain cosmological hypotheses predict — during the evolution of the universe, it should be reflected in the geologically and paleogeophysically understandable history and in the proportions and properties of the

Earth. Ertel (1935a,b) discussed such problems (and for the present status of this discussion see Watznauer 1985 and Lauterbach 1985).

Geology and paleo-geophysics are coupled here with cosmology and cosmogony. Critical tests in the fields of paleo-climatology, geotectonics and geochemistry yield at least upper limits for the secular change of the gravity constant (or more generally, the proportionality factor between heavy and inertial mass). These have been much discussed since Dirac and Jordan (e.g. by Ertel 1931, 1935a,b, 1936, 1938). The enormous increase in the accuracy of age determination due to the substitution of astronomical and astronomical-geodetic methods by micro-physical ones has made very small age differences detectable in the laboratory, too. (Thus the Einstein-shift in gravity potential can be confirmed by the motion of an atomic clock around the Earth, and laboratory measurements with high resolution spectroscopy — Pound-Rebka experiment — have an even higher degree of accuracy).

In addition to the gravity field, geophysics deals also with the geomagnetic field in which W Gilbert, a contemporary of Galilei's, was a pioneer of experimental research. Later, the mathematical and technical developments due to Gauss and Weber (in about 1840) followed from the world-wide research programme of geomagnetic measurements suggested by A von Humboldt. The problems of understanding geomagnetism remained even later following the development of the Faraday-Maxwellian electrodynamics, Lorentz's electron theory, the first steps of the quantum theory and relativit, coupled to basic problems of the physical theory of fields. Hypotheses of an uniform field theory of gravity and electromagnetism led to different ideas about the origin of various characteristics of the geomagnetic field and linked geomagnetism with the mass and moment of inertia of the Earth. That is how the research into the geomagnetic (and with it, the geoelectric) field gave at least estimates for the "gravo-electric" or "gravo-magnetic" effects, looked for ever since Faraday's time. This research included also aspects of the unitary field theory as put forward by Einstein and Schrödinger and led to negative results about the hodogetic ideas by H A Wilson, A Picard and P M S Blackett about the experimental limits of the unity of gravity and electromagnetism.

Treder (1977, 1979) discussed in more details these connections between geophysics, the theory of relativity and quantum theory. Here again, originally geophysically and cosmologically based hypotheses led to experimental insights into the elementary structure of matter and the universality and invariability of the constants of nature (Rompe and Treder 1985). It was an important advance to carry out and to put at our general disposal such experiments of increasing accuracy in nuclear physics, microelectronics and high resolution spectroscopy to enable us to apply the results of modern laboratory physics to problem of the "Earth and cosmos".

The pioneering role of geophysical methods in fundamental physics is exemplified in the first half of this century at the discovery and research of cosmic rays. The "Höhenstrahlung" (height radiation) discovered by V Hess, W Kolhörster et al. in 1911/1912 resulted at first in common problems for physical, meteorological serological research facilities.

The interaction of hard cosmic rays with the terrestrial atmosphere (under the

influence of the geomagnetic field) led to early results and ideas in high-energy physics, an in the physics of elementary particles. Corresponding to the dependence on the "optical density" of the terrestrial atmosphere towards cosmic rays, new physical reactions (shower and super-showers) were discovered and prototypes of new elementary particles (positrons, myrons, mesons etc.) were detected. The recognition of reactions which create showers and in which "new elementary particles" develop, necessitated and yielded ideas about the structure of the magnetosphere and ionosphere.

The significance of the research into cosmic rays was that it has became only possible in the 1950s to produce similar particles of sufficiently high energy in the laboratory (even "industrially"), in sufficient quantity, quality and intensity. Nevertheless, even today research is being undertaken in the "ultra-hard domain" of the spectrum of cosmic rays to find exotic elementary particles such as "free magnetic poles" (according to Dirac) and "free quarks".

In order to screen ultra-hard cosmic rays, very sensitive physical experiments are carried out in deep mine shafts etc., where knowledge about geotectonic structures and about the geochemical composition of the rocky roof is necessary to evaluate the accuracy of the measurements and to estimate eventual "dirt effects".

Such knowledge is also necessary for checking the theory of the limited lifetime of protons (as a result of unitary field theories). The average lifetime of a proton should be extremely long; theories led to estimates of 10^{34} to 10^{35} years i.e. to times against which the lifetime of Earth and cosmos namely 10^9 to 10^{11} years is negligibly short. Because of the magnitude of the Loschmidt number, $N \approx 6 \cdot 10^{23}$, in order to be able to record one event of proton decomposition per year, masses of more than 10000 tons have to be included in the experiment. This is only possible with large scale "geophysical" experimental arrangements which then need exact information about the local structure of the lithosphere and its geochemistry.

References

Ertel H 1931: *Gerl. Beitr. Geophys.*, 32, 38–46.
Ertel H 1935a: *Gerl. Beitr. Geophys.*, 43, 327–330.
Ertel H 1935b: *Naturwiss.*, 23, 36–38.
Ertel H 1936: *Physik. Zeitschr.*, 37, 138–139.
Ertel H 1938: *Naturwiss.*, 26, 463.
Lauterbach R 1985: *Ann. Phys.*, 42, 489–495.
Rompe R, Treder H-J 1985: *Ann. Phys.*, 42, 559–576.
Steenbeck M, Treder H-J 1981: Möglichkeiten der experimentellen Schwerekraftforschung. Akademie-Verlag, Berlin
Thiessen P A, Treder H-J 1982: *Gerl. Beitr. Geophys.*, 91, 97–107.
Treder H-J 1977: In: 3rd International Symposium Geodesy and Physics of the Earth, H Kautzleben (ed.) Potsdam
Treder H-J, Woodward J F, Yourgrau W 1979: *Gerl. Beitr. Geophys.*, 88, 11–22.
Watznauer A 1985: *Ann. Phys.*, 52, 533–536.

Acta Geod. Geoph. Hung., Vol. 35(3), pp. 373–378 (2000)

GAUSS AND THE THEORY OF GRAVITATION

W SCRÖDER[1] and H-J TREDER[2]

[Manuscript received April 22, 1999]

A historical review is given on various aspects of Gauss' research and on his interpretation of the theory of gravitation.

Keywords: Gauss; geomagnetism; theory of gravitation

1. Gauss and the theory of gravitation

The most important contribution by C F Gauss (1777–1855) to the theory of gravitation was based on his, at first very general proof deduced independently from S D Poisson (1781–1840) that the Newtonian potential

$$\varphi = -f \int \frac{o\, dx\, dy\, dz}{r}$$

only in vacuum fulfils Laplace's (1749–1827) homogeneous potential equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \varphi = \Delta\varphi = 0\,.$$

In border areas which are filled by masses, the Laplace-operator $\Delta\varphi$ defines the mass density ϱ:

$$\Delta\varphi = 4\pi f \varrho\,,$$

where f is the Newtonian gravitational constant. Based on this idea, Gauss proved in 1840 his great potential theses (among others, the Gaussian integral thesis) over the Newtonian potential function φ which enabled him to present the Newtonian gravitational force mathematically as a potential field in spite of its remote effect-character. An energy density belongs to this potential field:

$$\frac{1}{2} \cdot \left[\left(\frac{\partial\varphi}{\partial x} \right)^2 + \left(\frac{\partial\varphi}{\partial y} \right)^2 + \left(\frac{\partial\varphi}{\partial z} \right)^2 \right].$$

Thus Gauss succeeded in making an important step toward a field theory of gravitation.

Gauss studied the analogies of this potential field φ with, and its deviations from, electricity and magnetism which he studied both theoretically and experimentally

[1]Hechelstrasse 8, D-28777 Bremen Roennebeck, Germany
[2]Rosa-Luxemburg-Strasse 17a, D-14282 Potsdam, Germany

1217-8977/2000/$ 5.00

together with his Göttingen physical colleague W Weber (1803–1891). In connection with this, he considered the idea of a finite propagation velocity of electric and magnetic forces, and as first, looked for a correct mathematical formulation of such finite propagation velocities of physical effects. Gauss deduced from the finite velocity the fact of a temporal retardation, the retarded appearance of the effects of a cause due to the spatial distance between the events.

Gauss presented a mathematical initial idea in a letter from 1845 to Weber. Gauss stimulated his and Weber's student B Riemann (1826–1866) who outlined in 1850 the first field theory of gravitation and electricity which are congenial in their physical intuition with M Faraday's (1791–1867) contemporary basic ideas on the electromagnetic field.

Gauss' idea could also be transferred to gravitation, as shown by Riemann and by other mathematicians and astronomers influenced by Gauss. Gauss' idea led — in contrast to Laplace's opinion expressed in his famous "Mécanique céleste", 1805 — to the result that a retardation (with about the velocity of light c) of the gravitational effects is acceptable in the frame of celestial mechanics of the solar system.

As basis for the celestial mechanical consequences of the retardation of the gravitational action according to Gauss, Riemann and C Neumann (1832–1925) developed the difference between the gravitational potential φ and the "effective interaction potential of gravitation" Φ. The latter interaction potential Φ appears in the Lagrangean action-function L (and with it, in the action-integral $\int L dt$, too) for the planetary motion. If T is the kinetic energy of a planet with mass m, then:

$$L = T + m \cdot \Phi \tag{1}$$

and the equations of motion are the "Euler-Lagrangean equations" to Eq. (1):

$$\frac{d}{dt} \cdot \frac{\partial L}{\partial v^I} - \frac{\partial L}{\partial x^I} = 0 \,.$$

The gravitational potential $\varphi \approx \frac{1}{r}$ is — with the exception of the sign — identical with Riemann's interaction potential Φ only for potential functions which do not depend on the velocities $v^I = \frac{d}{dt} x^I$ of the masses similarly to the Newtonian potential, i.e. in such cases $\varphi = -\Phi$. This is, however, generally not valid, and first of all, terms in Φ which are of the form of a total differential of the time as $\frac{d}{dt} S = \dot{S}$ have no importance as such terms do not contribute to the equations of motion of the masses.

The Gaussian retardation of the gravitational action with a finite velocity of propagation $c^* = \frac{c}{\sqrt{\delta}}$ of the gravitation — which has to be very large with respect to the velocity of motion of the planets relative to the Sun v^I, so as $c^{*2} \gg v^2$ is valid — it yields for the Newtonian potential function $\Phi = \frac{fM}{r}$ a series after the radial velocity of the planet $\frac{d}{dt} r = \dot{r}$ relative to the Sun:

$$\Phi' = \frac{fM}{r\left(1 - \frac{r}{c^*}\right)} = \frac{fM}{r} \cdot \left(1 + \frac{\dot{r}}{c^*} + \frac{\dot{r}^2}{c^{*2}} + \ldots\right) . \tag{2}$$

This series was deduced by Neumann in 1868 directly from Gauss' ideas on the importance of the finite propagation velocity of physical forces. In Eq. (2), the expression

$$\frac{fM}{c^*} \cdot \frac{\dot{r}}{r} = \frac{fM}{c^*} \cdot \frac{d}{dt}(\log r) \tag{3}$$

is a complete differential. Thus, Neumann's series, Eq. (2), yields with $c^{*2} = \frac{c^2}{d} \gg \dot{r}^2$ the effective action-function of the planetary motion:

$$L' = T + m\Phi' \approx T + f\frac{mM}{r}\left(1 + \frac{\delta \dot{r}^2}{c^2}\right). \tag{4}$$

An analogous expression had been intuitively introduced into electrodynamics by Weber in 1846 on an initiative in one of Gauss' letters.

The form of the kinetic term T (the kinetic energy of the planet) is in Eq. (4) independent from kinematics. In the Galilean and Newtonian classical mechanics it is valid for the motion of a planet m around the Sun in rest:

$$T = \frac{m}{2}v^2 \tag{5a}$$

while in Einstein's relativistic mechanics it is approximately:

$$T \approx m\left(\frac{v^2}{2} + \frac{v^4}{8c^2}\right). \tag{5b}$$

It is evident from these formulas that the retardation effects of the gravitation are of the order of magnitude $\delta\frac{v^2}{c^2}$ according to Gauss, Riemann and Neumann. There are no correction terms of the order of magnitude v/c to the Newtonian force laws; all non-Newtonian corrections in the gravitational law are proportional to $\frac{v^2}{c^2} \cdot \frac{fM}{r}$, if the propagation velocity of gravitation $c^* = \frac{c}{\sqrt{\delta}}$ is of the order of magnitude of the velocity of light.

This result implies Gauss' criticism to Laplace's thesis. Laplace supposed namely that the series development of the retarded gravitational potential after the radial velocity $\dot{r}$ of the planet:

$$\varphi' = -\frac{fM}{r}\left(1 + \frac{\dot{r}}{c^*} + \frac{\dot{r}^2}{c^{*2}} + \ldots\right) \tag{6}$$

should have to result in a correction term to the gravitational force being linear in the velocity $\dot{r}$, namely as $\approx \frac{\dot{r}}{c^*} \cdot \frac{fM}{r^3} \cdot x^1$. In order to make such a correction compatible with the facts of celestial mechanics, the estimation $c^* \geq \frac{c}{v}c$ should have to be valid. Thus according to Gauss the propagation velocity of the gravitation c^* could be of the order of magnitude of the velocity of light c, while according to Laplace it should have to be at least of the order of magnitude

$$c^* \div \frac{c^2}{v} \gg c$$

in order to get disturbances in the planetary motion due to retardation effects in acceptable magnitude. Laplace finally supposed that even $c^* \geq \left(\frac{c}{v}\right)^2 c$ should hold.

It was a consequence of the studies made by Gauss and by his students that Laplace's error could be corrected and the fact proved that celestial mechanics includes the possibility that gravitational action propagates with a velocity c^* being of the order of magnitude of the velocity of light c in accordance with the field physics of gravitation. The result of such a propagation of gravitation $c^* = \frac{c}{\sqrt{\delta}}$ yields then empirically just detectable disturbances in the Keplerian motion of two- and three-body problems of celestial mechanics, which result in a secular motion of the apsid-lines as described by F Tisserand's (1845–1896) formula:

$$\Delta\varphi = \frac{2\delta\pi f M}{c^2\alpha(1-e^2)}\,. \tag{7}$$

Here α is the major axis and e the eccentricity of the Keplerian ellipse. Tisserand and F Zöllner (1834–1882) have already brought the motion of the perihelion as expressed by Eq. (7) in connection with the unexplained rest term found by U Leverrier (1811-1877) in the celestial mechanical theory of the motion of Mercury.

2. Later discussions and applications

These insights into the connections between the retardation of the gravitation action in Gauss' sense and the "non-Newtonian" corrections to the classical Keplerian problem of the motion of a planet in the gravitational field of a central body were reconsidered around 1870 by follower-mathematicians and -astronomers of Gauss and Laplace as F Tisserand in Paris, C Neumann, H Scheibner and F Zöllner in Leipzig. In consequence of the re-formulation of all problems of the gravitational theory within Einstein's theory of relativity, these older problems were discussed again around 1915/1920 by H v Seeliger, P Painlevé and E Wiechert among others.

If the retardation velocity c^* is as according to Gauss of the order of magnitude of the velocity of light c (being the fundamental constant of electrodynamics and special theory of relativity), then the non-Newtonian corrections to the two- and three-body problems of celestial mechanics are, according to Gauss', Riemann's and Weber's potential functions, of the same order of magnitude as the relativistic effects in celestial mechanics. Such effects were deduced from the special theory of relativity by H Poincaré (1906) and by W de Sitter (1911). Einstein found in 1915 the formula

$$\Delta\psi = \frac{6\pi f M}{c^2\alpha(1-e^2)} \tag{8}$$

for the motion of the perihelion from his general theory of relativity as the only "non-Newtonian" correction of the Keplerian motion. This result was confirmed by K Schwarzschild in 1916 by the determination of the exact form of the centrally symmetric gravitational field in the framework of the general theory of relativity. Einstein, Infeld and Hoffmann expanded this in 1938 to the two-body problem.

The same rotation of the perihelion as given by Eq. (8) would be obtained from Tisserand's Eq. (7) from the retardation of the Newtonian gravitational action according to Gauss, if the effective gravitational potential is supposed to be

$$\Phi' = \frac{fM}{r}\left(1 + 3\frac{\dot{r}^2}{c^2}\right)$$

(Gerber 1898) what yields a propagation velocity c^* of the gravitational action less than the velocity of light c:

$$c^* = \frac{c}{\sqrt{\delta}} = \frac{c}{\sqrt{3}}\,. \tag{9}$$

Thus, Eq. (9) is physically very satisfactory: the perihelion motion of Mercury is really no consequence of a finite propagation velocity of gravitation (H v Seeliger 1917/1918, M V Laue 1917/1920).

The perihelion motion (Eq. 8) is deduced in Einstein's general relativity theory from the geodetic motion of the planet in Schwarzschild's space-time-metrics, and this spherical symmetric metrics is essentially independent of time; it is the equivalent of the Newtonian gravitational potential $-\frac{fM}{r}$ in the general theory of relativity. According to Einstein's general theory of relativity the retardation effect of the gravitational field of the Sun is purely an effect of the co-ordinates which is produced by the choice of the reference system and which could be also eliminated by an appropriate choice. This is the same as shown by Riemann (around 1859) that the retarded action of the gravitational potential in Neumann's approximation fulfils a time-dependent generalisation of the Laplacean potential equation. The effective interaction potential

$$\Phi' = \frac{fM}{r}\left(1 + \frac{\dot{r}^2}{c^*2}\right)$$

is namely a solution of the bi-harmonic potential equation:

$$\Delta\Delta\Phi' = 0\,.$$

In this problem Einstein and his general theory of relativity closed a series thoughts initiated by Gauss. This conclusion led simultaneously to the new problem of gravitational waves as possible analogue of the electromagnetic waves. Physical processes are looked for here which transport the energy of the gravitational field with the velocity of light (as anticipated by Gauss) (Treder 1975, Schröder and Treder 1997).

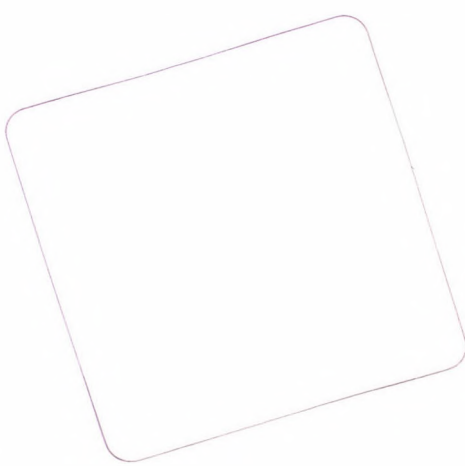
References

Einstein A 1929: Die Grundlagen der allgemeinen Relativitätstheorie. 5. Aufl., Leipzig

Einstein A 1969: Grundzüge der Relativitätstheorie. 7. Aufl., Berlin, Braunschweig

Gauss C F 1889: Allgemeine Lehrsätze in Beziehung auf die im verkehrten Verhältnisse des Quadrates der Entfernung wirkenden Anziehungs- und Abstossungskräfte. Ostwalds Klassiker, Nr. 2, Leipzig

Gauss C F 1893: Die Intensität der erdmagnetischen Kraft, auf ein absolutes Mass zurückgeführt. Ostwalds Klassiker, Nr. 53, Leipzig
Gauss C F 1927: Anziehung eines Ringes. Ostwalds Klassiker, Nr. 225, Leipzig
Gauss C F, Werke Bd V Darin 1867: Allgemeine Theorie des Erdmagnetismus. Göttingen
Laplace de P S 1880: Mécanique céleste. In: Oeuvres, P S de Laplace, Bd. 4, Paris
Laue M v 1961: In: Gesammelte Schriften und Vorträge, M v Laue, Bd. III, Braunschweig
Neumann C 1898: Allgemeine Untersuchungen über die Newtonsche Theorie der Fernwirkungen. Leipzig
Riemann B 1880: Schwere, Elektrizität und Magnetismus. 2. Aufl., Hannover
Schröder W, Treder H-J 1997: *EOS*, 78, 479.
Tisserand F 1896: Mécanique céleste. Bd. IV, Paris
Treder H-J 1972/1975: Die Relativität der Trägheit. Berlin/Moskau
Treder H-J 1974: Über die Prinzipien der Dynamik bei Einstein, Hertz, Mach und Poincaré, Berlin
Treder H-J 1975: *Die Sterne*, 51, 69–81.
Weber W 1890/1892: Gesammelte Werke. Bd. III, IV, Leipzig
Wiechert E 1925: In: Physik in der Kultur der Gegenwart. E Lechner (ed.), 2. Aufl., Leipzig
Zöllner F 1883: Dei Natur der Cometen. 3. Aufl., Leipzig



Acta Geod. Geoph. Hung., Vol. 35(3), pp. 379–381 (2000)

BOOK REVIEWS

W SCHRÖDER ed.: *Geomagnetism (Research: Past and Present).* (Containing papers presented at IAGA historical symposia in Birmingham, August, 1999.) Interdivisional Commission on History of the IAGA – European Section, Science Edition, D-28777 Bremen-Rönnebeck, 2000, 248 pp, 20 $

This book is like flashlight on certain periods of the science branch Geomagnetism (in a very wide sense of the word) on a less illuminated background of the general development of human knowledge. This illumination is given by Gregori (following Moritz's introductory considerations on history of science) in a philosophical paper on Geomagnetism and Fundamental Science where he discusses the earliest history and use of geomagnetism in navigation and mentions several less known very early Italian sources, as e.g. Burri's (lost) first isogonic map. Halley first conserved isogonic map, with the first isolines, called at that time Halleyan lines which are also discussed, followed by experiments to use geomagnetism for positioning purposes on sea. In the 19th century, the central problem was the source of the geomagnetic field. Gauss' discovery on the internal origin, controversies about the possibility of material origin, models of the Earth's interior led to Gregori's own hypotheses about the geomagnetic field, reconnection etc.

Halley's role is well documented in two papers, one (Cook) explains his role in our ideas about the geomagnetic field, while Clark describes his Atlantic voyages which resulted in the isogonic map of the ocean. Mandea adds to them the French activity, mainly by Nautonier, who also tried to use the magnetic field for positioning (longitude determination) purposes. In a further paper she lists declination measurements in France in the 17th century, together with their supposed accuracy. Barraclough surveys the development of geomagnetic maps and modelling, starting from the 17th century till present maps deduced from satellite measurements.

Beckman's paper on a very important geomagnetic storm, in a certain sense the first identified as such is a most interesting one. The Swedish Hiorter and the English Graham were the first to observe a geomagnetic storm simultaneously in Greenwich and Uppsala, with the silent result that extrema occurred simultaneously at both places. Curiously they did not mention this fact. Hiorter was a pupil of Celsius, who inspired such measurements. He was also among the pioneers of the connection aurora-geomagnetism.

The next jump is quite big in time: Yoshida discusses Tanakadate's activity in connection with the supposed non-potential part of the geomagnetic field, the effect of vertical currents in the air. Tanakadate conserved scientific integrity even if results contradicted his ideas. Julius Bartels, his somewhat younger contemporary is best known from the big volumes of the Chapman-Bartels Geomagnetism, and for the older generation by his (manual) methods of the study of cyclic phenomena. The Bernhardts survey his activity in the latter field. Collacino, Ferrante and Valensise refute claims about Nobile's negligence in preparation of his tragic polar expedition based on unpublished correspondence between him and Eredia about the meteorological side of the preparation.

The last two papers are of a more general character: Scalera discusses the different reactions of the Church to Medieval developments in Earth sciences (shape and motion of

the Earth, fossils) and finds the first appearance of modern thinking in questions put for discussion by Medieval scholars. Gregori and Gregori survey possibilities of archeastronomy, especially orientation and position of windows in ancient buildings. Light through asymmetrically placed openings illuminate in some cases e.g. the image of the Patron Saint enabling astronomic calculations. The paper is followed by a list of monuments with different characteristics, mostly with divisions of the circle into parts of integral numbers. Some abstracts conclude the very interesting collection, both from a historical and from a geomagnetic-geophysical point of view.

J Verő

W SCHRÖDER ed.: *Long and Short Term Variability in Sun's History and Global Change.* (Containing papers presented at IAGA historical symposia in Birmingham, August, 1999.) Interdivisional Commission on History of the IAGA – European Section, Science Edition, D-28777 Bremen-Rönnebeck, 2000, 364 pp, 20 $

The first part of the papers in this book deals with different long period variations in the solar activity and with the effects of this variations on terrestrial climate and weather, partly also on biological systems. The impression is having read all the corresponding papers that there is no consensus in these questions, apart from the fact that such changes exist in the solar activity and they may influence terrestrial processes, too. A certain theoretical-experimental background is given by McIntyre who discusses angular momentum transfer from waves to linear motion, both in solar and terrestrial (Quasi-Biennial Oscillation) aspects. Egeland reports on three important periods in Norwegian auroral research: first mentions in Viking times, Celsius and Hiorters's discovery of the connection with geomagnetic activity and 19th century achievements, including Birkeland's basic results. Schröder and Treder found that the number of auroras was not significantly less during the Spörer and Maunder minima than else, and list possibilities for further support of this funding by other means.

The following papers can be distributed according to the effects studied. In the first group, only solar events are included, in the second, long period effects are looked for in terrestrial phenomena, in the third, direct connections are searched between solar and terrestrial phenomena.

Gavryusheva and Gavryushev describe connections between different indicators of solar activity and conclude that the correlations between them confirm their common origin. Tsirulnik et al. performed frequency analysis by the Method of Global Minimum and found many but unstable periods in the sunspot numbers and they also tried to predict the present cycle. Paterno concludes from the history of the solar neutrino problem that new properties of the neutrino are to be found. Noci discusses SOHO's results about the extended solar corona, mainly in the ultraviolet part of the spectrum. Current ideas about coronal holes, streamers and slow solar wind are explained and illustrated, including the structure of neutral sheets in the solar wind

In the second group, Sneyers proposes the method of Change-Point Detection for the study of the homogeneity of data series. Using this method, he found such changing points in the Northern Hemisphere temperature data and identified several such points in each season. Yan suggested the existence of periods of 4–5, 22 and longer than 50 years in temperature, emphasizing the quasi-bi-decadal wave supported by several data sets. Sofia summarizes possibilities and future tasks in the study of the solar effect on climate. Schuurmans confirms his previously published double solar cycle (22 years) effect

in temperatures in The Netherlands. Brunetti et al. analyzed Italian mean temperature and precipitation data, and found significant trends and periodicities. The amplitudes of the different waves, however, changed significantly, but conserved phase. Correlations with sunspot and geomagnetic data are significant, too, and they concluded to – at least partly – solar forcing of the present warming. A lag in Northern hemisphere temperatures by about ten years is explained by the thermal inertia of the oceans.

In the third group Hanna demonstrates a connection between smoothed sunspot numbers and global temperature with a correlation as high as 0.83. Baranyi and Ludmány found different terrestrial effects in the case of parallel and antiparallel years. They emphasize their use of proxy data for all variables and the significance of historic data in future studies. O'Brien discusses the atmospheric response to solar particle events, and gives a suitable model for it, even in the case of a big event (increased ionization on September 29–30, 1989). Bucha looks for connections between geomagnetic, solar and climate variations, too. In his opinion, geomagnetic forcing plays an active role in climate history, especially in polar regions.

Cornelissen et al. present effects of the eleven year solar cycle in biological systems. As this paper is supposedly not published in full length, the understanding of the interesting results is difficult in certain points. These are discussed in more detail by Halberg who presents research in connection with the solar effects on biological systems in longer study, including bibliography with 151 items. With a certain simplification, the aim is to find the cause of the weekly and semi-weekly period detected in biological systems. It is supposed that the cause lies in geomagnetism, in weekly periods of the solar wind, geomagnetic activity etc. Nevertheless, these periods are there seldom significant and very variable, thus such periodicity in the paleogeomagnetic field is suggested.

In the last part of the book, some topics are discussed which are somehow connected to the terrestrial effects of the solar activity.

Mosalam Shaltout describes current research in solar-terrestrial physics in Egypt. He mentions the correlation between high water of the Nile and solar activity, too. The early history of meteorological observations in Europe, especially in Italy is presented by Colacino. It is interesting to note that the network at the end of the 18th century included Buda in Hungary, too. Tankosic et al. discuss the properties of heavy elements in hot stars. Hörz reviews Helmholtz's idea about shrinking as source of the Sun's heat, instead of the previously proposed meteoric origin by Robert Mayer and Lord Kelvin. Filling presents historical illustrations about solar eclipses, Rendtel describes available material for the August 11, 1999 eclipse. Gaab lists two interesting cases when in the 16–17th centuries comets and auroras could be confused. A dozen short summaries and several photos of the lecturers conclude the book.

Gregori et al's paper again summarizes the aim of he study of Earth sciences. These aims are well served by the two books published on the basis of the Birmingham symposia, not only by presenting results, but also by explaining much debated fields from different sides, thus by showing fields which need further research. Papers in these fields are often published in different, less known journals, thus a lot of them remains nearly unknown, even experts are less acquainted with others' methods and results. In addition to other merits of the books, this simultaneous publication of the results of different groups may help to find consensus in the debated questions.

J Verő

INSTRUCTIONS TO AUTHORS

Manuscripts should be sent to the Editorial Office (address see above). Articles not published or submitted for publication elsewhere are only accepted.

Manuscripts should be preferably on disc (LATEX *.tex files or Word for Windows documents or text-only ASCII files) with a written copy, or via e-mail (actagg@ggki.hu). Files will be accepted, after previous agreement, via *anonymous ftp*, too. Typewritten manuscripts (in duplicate) are also accepted.

Manuscripts should include the following components which should be presented in the order listed (tables and illustrations should be separated in case of manuscripts on disc, too).

1. Title, name(s) of the author(s), affiliation, dateline, abstract, keywords
2. Text, acknowledgements
3. References
4. Footnotes
5. Legends
6. Tables
7. Illustrations

1. The *affiliation* should be as concise as possible and should include the complete mailing address of the authors. The *date of receipt* will be supplied by the editors. The *abstract* should not exceed 250 words and should clearly and simply summarize the most important methods and results. 5-10 significant expressions describing the content are used as *keywords*. Authors may recommend these keywords.

2. The *text* should be in English and as short and clear as possible. In case of typewritten manuscripts, please note the following:

- avoid possible confusion between o, O (letters) and 0 (zero), l (letter) and 1 (one), ν (Greek nu) and u, v (letters), etc.
- explain ambiguous and uncommon symbols by making marginal notes in pencil
- formulas are to be numbered consecutively with the number in parentheses to the right of the formula.
  Text references to the equations may then be made by the number in parenthesis. The word equation in this context is to be abbreviated to Eq. and Eqs in the plural
- the International System of Units (SI) should be used.

3. *References* are accepted only in the Harvard system. Citations in the text should be as:

... (Bomford 1971) ... or Bomford (1971) ...
... (Brosche and Sündermann 1976) ...
... (Gibbs et al. 1976b) ...

The list of references should contain names and initials of all authors (the abbreviation et al. is not accepted here); for journal articles, year of publication, title of the journal in abbreviated form, volume number, first and last page. For books or chapters in books, the title is followed by the publisher and place of publication.

All items must appear both in the text and in the list of references.

4. *Footnotes* should be typed on separate sheets.

5. *Legends* should be short and clear. The place of tables and figures should be indicated in the text on the margin.

6. *Tables* should be numbered serially with Roman numerals. Vertical lines are not used.

7. All the *illustrations* should contain the figure number and author's name in pencil on the reverse. The most important point with figures is clearness. Photographs and half-tone illustrations should be sharp and well contrasted. Colour photographs will be accepted, but the extra cost of reproduction in colour must be borne by the authors. Encapsulated postscript (EPS) files will be accepted on disk or via e-mail, too.

Only original papers will be published and a copy of the Publishing Agreement will be sent to the authors of papers accepted for publication. Manuscripts will be processed only after receiving the signed copy of the agreement. Information is sent to the first author if no other wish is expressed.

Acta Geodaetica et Geophysica Hungarica

VOLUME 35, NUMBER 4, 2000

Akadémiai Kiadó, Budapest

AGG 35 (4) 383–480 (2000) HU ISSN 1217-8977

ACTA GEODAETICA ET GEOPHYSICA HUNGARICA

A Quarterly Journal of the Hungarian Academy of Sciences

Acta Geodaetica et Geophysica Hungarica (AGG) publishes original reports and reviews on geodesy and geophysics in English.

AGG is published in yearly volumes of four issues by

AKADÉMIAI KIADÓ
Prielle K. u. 4, H-1117 Budapest, Hungary
http://www.akkrt.hu

Manuscripts and editorial correspondence should be addressed to

AGG Editorial Office
Geodetical and Geophysical Research Institute
P.O. Box 5, H-9401 Sopron, Hungary
Phone: (36-99) 508-340
Fax: (36-99) 508-355
E-mail: actagg@ggki.hu
http://www.ggki.hu/agg/

Subscription information

Orders should be addressed to

AKADÉMIAI KIADÓ
P.O. Box 245, H-1519 Budapest, Hungary
Fax: (36-1) 464-8221
E-mail: kiss.s@akkrt.hu

Subscription price for Volume 35 (2000) in 4 issues US$ 180.00, including normal postage, airmail delivery US$ 20.00.

Printed in Hungary
PXP Ltd., Budapest

AGeod 35 (2000) 4

CONTENTS

Acta Geod. Geoph. Hung., Vol. 35(4), pp. 383–396 (2000)

THE NUMBER OF HOTSPOTS IN MANTLE CONVECTION: EFFECT OF DEPTH-DEPENDENT VISCOSITY AND INTERNAL HEATING IN TWO-DIMENSIONAL MODELS

A GALSA[1] and L CSEREPES[1]

[Manuscript received January 13, 2000]

Two-dimensional numerical models of mantle convection have been calculated for a range of high Rayleigh numbers, depth-dependent viscosity and basal plus internal heating. Large aspect ratio boxes have been used in the calculations in order to estimate the expected areal density of upwellings in infinite fluid layers. The results are analyzed with regard to the number of the Earth's hotspots which are assumed to be surface imprints of cylindrical upwellings in the mantle. For a pure whole-mantle situation, 6–7 upwellings can be expected. If the upper mantle convects separately above the 660 km discontinuity (allowing a second convective layer below 660 km depth), the theoretically estimated number of upper-mantle plumes can be as high as 250. Given the number of real hotspots (42 to 117 according to different compilations), it is suggested that the flow regime of the mantle is intermediate between the pure whole-mantle or pure two-layer circulation.

Keywords: convection; mantle plumes; numerical simulation

1. Introduction

Convection in the Earth's mantle is of thermal origin: it is the consequence of buoyancy resulting from thermal expansion and the heterogeneous distribution of temperature. Beyond the fundamental fact that convection is the cause of plate motion, its most important tectonic manifestations are the subduction zones, mid-oceanic ridges and hotspots.

According to present knowledge, dipping sheets marked by deep seismicity beneath subduction zones are the main downwellings of mantle circulation. This is clearly proven by their positive velocity anomaly of 5-10% found in seismic tomographic measurements (Mitronovas and Isacks 1971). Recent results of global tomography report downwelling slabs to at least 1700 km depth, and some slab segments (e.g. the Farallon anomaly beneath Central America) can be detected as deep as the core-mantle boundary (CMB), which is a strong argument for whole mantle circulation (van der Hilst et al. 1997).

The deep root of mid-oceanic ridges is a place where one can assume sheet-like upwellings. These ascending currents rising up to the surface in rifts, though producing high surface elevation in large extents, are only secondary phenomena in all likelihood: these upwellings are a passive consequence of the divergence of two adjacent lithospheric plates. This conclusion is confirmed by tomographic maps

[1]Department of Geophysics, Eötvös University, Ludovika tér 2, H–1083 Budapest, Hungary

1217-8977/2000/$ 5.00

of the upper mantle which show that the negative velocity anomaly below the ridge systems can be tracked down to the depth of about 350 km (Woodhouse and Dziewonski 1984).

The most intriguing traces of convection on the surface are the hotspots. According to the hypothesis suggested by Morgan (1971), they are located above relatively narrow upwelling pipes, 'plumes', which rise across the mantle. The hotspots, as their name indicates, are characterized by high heat flow and volcanism. In addition, the elevation of the ambient ocean floor (in the case of oceanic hotspots) and the corresponding gravity anomaly (Robinson et al. 1987) suggest strongly that the hotspots are results of convective upwellings, probably the main upwellings of the mantle. Recent tomographic procedures have already been able to delineate narrow warm structures below hotspots in the upper mantle (Kennett and Widiyantoro 1999), and the root of some of them can be traced down to the boundary layer at the CMB, as in the case of the plume below Iceland detected by 'classical' travel time tomography (Bijwaard and Spakman 1999) and the one below the Hawaiian hotspot by diffraction tomography (Ji and Nataf 1998a,b).

Numerical hydrodynamic modelling is an accepted tool of studying possible convective structures in the mantle. Although laboratory experiments can also yield valuable information on the physics of thermal convection, a series of aspects of the real mantle circulation (spherical geometry, presence of internal heating, nonlinear creep mechanisms) are hardly reproducible under laboratory conditions, i.e. they can only be examined by solving the relevant hydrodynamic equations. This approach is used here for studying the development of upwellings.

Various features of the upwelling currents could be chosen as a basis for checking the observed properties of mantle plumes, but here we will focus on a single and easily observable property, namely the number (or areal density) of them. If plumes are identified by hotspots, it is a question of simply counting them on the Earth's surface and their number can be compared with the number of upwellings produced in theoretical models. There is no universal agreement how many or which hotspots could correspond to individual plumes, since clusters of close hotspots can originate from a single large upwelling, but various hotspot statistics show that there may be 42 (Crough and Jurdy 1980) to 117 (Vogt 1981) plumes in the mantle.

The actual purpose of the present study is to investigate how the number of upwellings in theoretical calculations depends on the physical parameters prescribed in the models. Our modelling has been restricted to selected cases which might correspond to the physical properties of the mantle. The effect of heating modes as well as the depth-dependence of viscosity has been studied under isothermal and stress-free boundary conditions at high Rayleigh numbers (higher than 105). In order to obtain reliable results on the 'density' of upwellings, the calculations had to be carried out in wide boxes (with large aspect ratios up to 16). The enormous computational requirements did not allow such calculations in realistic, three-dimensional conditions, therefore our study is restricted to two dimensions. The basic structural features, e.g. characteristic length scales of convective cells must not depend significantly on this simplification, thus we may expect that the results can be compared to the real mantle convection.

2. Model description

The Boussinesq approximation (Chandrasekhar 1961) of the equations of convection, governing the conservation of mass, momentum and energy, can be written as:

$$\frac{\partial u_j}{\partial x_j} = 0\,, \tag{1}$$

$$\varrho_a \cdot \frac{du_i}{dt} = \varrho_a g e_i \left[1 - \alpha(T - T_a)\right] \frac{\partial}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\eta\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)\right]\,, \tag{2}$$

$$\frac{\partial T}{\partial t} + u_j \frac{\partial T}{\partial x_j} = \kappa \frac{\partial^2 T}{\partial x_j^2} + \frac{Q}{\varrho_a c_p}\,, \tag{3}$$

where i, j are 1 and 2 for two-dimensional modelling; x_j are the x and z coordinates, the latter being positive downwards, towards the centre of the Earth; t is time; $e_i = (0, 1)$ is the downwards pointing unit vector; ϱ_a is the average density; g is the gravitational acceleration; Q is the volumetric rate of internal heating; T_a is the average temperature. The thermodynamical parameters are the following: α is the thermal expansion coefficient; c_p is the specific heat; and the transport parameters: η is the dynamic viscosity; κ is the thermal diffusivity. $u_j = (u, w)$, T and p denote the unknown velocity, temperature and pressure distributions, respectively. The parameters α, c_p, κ and g are considered constant both in time and space, while the viscosity is allowed to vary with depth.

It is suitable to non-dimensionalize the equations for the numerical modelling. With some algebra, the Eqs (1–3) can be reduced to

$$\left(\frac{\partial^2}{\partial z^2} - \frac{\partial^2}{\partial x^2}\right)\left[\eta\left(\frac{\partial^2}{\partial z^2} - \frac{\partial^2}{\partial x^2}\right)\Psi\right] + 4\frac{\partial}{\partial z}\left(\eta \frac{\partial^3}{\partial x \partial x^2}\Psi\right) = R\frac{\partial T}{\partial x}\,, \tag{4}$$

$$\frac{\partial T}{\partial t} + u_j \frac{\partial T}{\partial x_j} = \frac{\partial^2 T}{\partial x_j^2} + H\,, \tag{5}$$

where

$$R = \frac{\alpha \varrho_a g \Delta T d^3}{\kappa \eta_a} \tag{6}$$

$$H = \frac{Q d^2}{K \Delta T} \tag{7}$$

are the Rayleigh number and the non-dimensional internal heating, respectively. The new symbols in Eqs (4–7) are as follows: d is the layer depth (this can be assumed as the total depth of the mantle); η_a is the reference viscosity at the surface; K is the thermal conductivity; ΔT is temperature drop across the layer and Ψ denotes the stream function. Ψ describes the velocities as

$$u = -\frac{\partial \Psi}{\partial z}; \qquad W = \frac{\partial \Psi}{\partial x}\,. \tag{8}$$

For the non-dimensionalization of the above equations, we applied the following primary scale units: d for length, d^2/κ for time, ΔT for temperature and η_a for the viscosity.

As mentioned above, we will show cases which probably approximate well the main features of the mantle. This means stress-free boundaries at the top and bottom, which is the best approach to treating both the surface and the mantle-core interaction. The boundary conditions for the Eq. (4) are $\Psi = \partial^2\Psi/\partial z^2 = 0$. The non-dimensional thermal boundary conditions for the Eq. (5) are $T(z = 0) = 0$ at the top of the domain and $T(z = 1) = 1$ at the bottom. Reflecting boundary conditions were imposed along the sides of the model. Where we need to scale back the non-dimensional quantities to physical dimensions, we will use the coefficients proposed by Leitch and Yuen (1989), namely (to mention only those which will be really needed later in this paper): $d = 2900$ km, $\Delta T = 2900$ K, $K = 5$ W/(m K), $\kappa = 10^{-6}$ m^2/s.

The Eqs (4), (5) and (8) have been solved with a hybrid, partly finite-difference, partly spectral, scheme (Cserepes 1985). We have used a rectangular model domain with a large aspect ratio of 16, in order to avoid the disturbing influence of the side walls to the plume density. The numerical grid has 128 evenly spaced grid-points in the vertical and 1024 grid-points in the horizontal direction. In all cases the computations were started from conductive initial conditions with a small horizontal random perturbation to start the evolution of plumes.

3. Results and discussion

3.1 Testing the code

The accuracy of the numerical algorithm must be tested. The benchmark results of Blankenbach et al. (1989) has been used for this purpose.

Three cases are compared to the benchmark tests. For all these, constant viscosity and no internal heating has been imposed in a square domain. The top and the bottom boundaries are isothermal and stress-free, and reflecting conditions are assumed along the sides similarly to all our subsequent models presented in this study. The three test cases differ in the Rayleigh number (Fig. 1). The comparison of our solutions to the benchmarks of Blankenbach et al. is given in terms of the Nusselt number and the root-mean-square velocity. The former is the average heat flux across the surface and the latter is

$$v_{rms} = \sqrt{\frac{1}{l}\int_0^l \int_0^1 (u^2 + w^2) dz dx}\,, \tag{9}$$

where l is the non-dimensional length of the box. Table I presents the comparison. It is apparent that, although the inaccuracy of results increases with the Rayleigh number, its magnitude remains acceptable throughout.

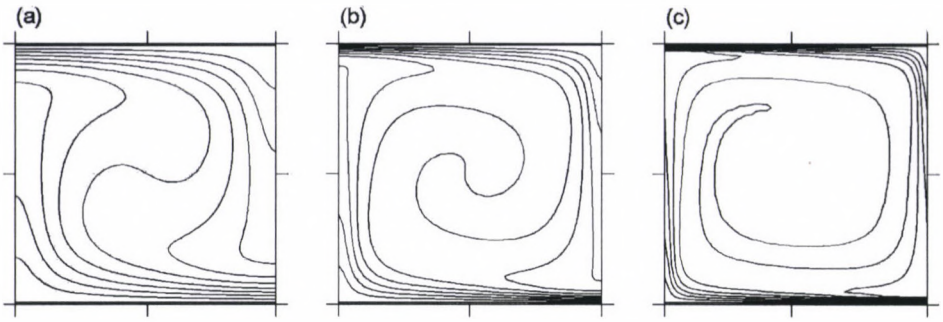


Fig. 1. Convective half-cells in a square box with isothermal and stress-free boundary conditions for three Rayleigh numbers. The boxes are discretized with 64×64 grid-points. The steady-state temperature distribution is shown with uniform contouring between 0 (top temperature) and 1 (bottom temperature). The Rayleigh numbers: (a) $R = 10^4$, (b) $R = 10^5$ and (c) $R = 10^6$

Table I. Comparison of the results of Blankenbach et al. (1989) and the present study. Nu is the Nusselt number and v_{rms} is the root-mean-square velocity

		Blankenbach et al. (1989)	Present study	Deviation %
Fig. 1a	Nu	4.884	4.876	0.16
	v_{rms}	42.865	42.823	0.1
Fig. 1b	Nu	10.534	10.466	0.65
	v_{rms}	193.215	192.59	0.32
Fig. 1c	Nu	21.972	21.442	2.41
	v_{rms}	833.99	824.59	1.13

3.2 The effect of depth-dependent viscosity

The simplest case to be shown is the constant viscosity model without internal heating. In this case we studied a broad range of the Rayleigh number from a low value ($R = 10^3$), which is somewhat higher than the critical Rayleigh number ($R_{cr} = 658$), up to the highest value ($R = 10^7$) possible with the chosen numerical grid without encountering numerical instabilities.

As it is known, if the Rayleigh number increases, the convection will be more and more vigorous and the surface heat flow will rise. At the same time, the diameter of the upwellings and downwellings will decrease (Figs 2a-2c). When $R \geq 10^5$, the plumes have the expected narrow 'stems' and wide 'mushroom heads'. The upwellings and downwellings are equally well-developed features. At the value of $R = 10^5$, the symmetry of the isotherms, so obvious at lower Rayleigh numbers, starts to be disrupted: the convection is no longer steady and reaches only a quasi-steady state. From about $R = 5 \cdot 10^5$ (Fig. 2b), the number of the cells in the

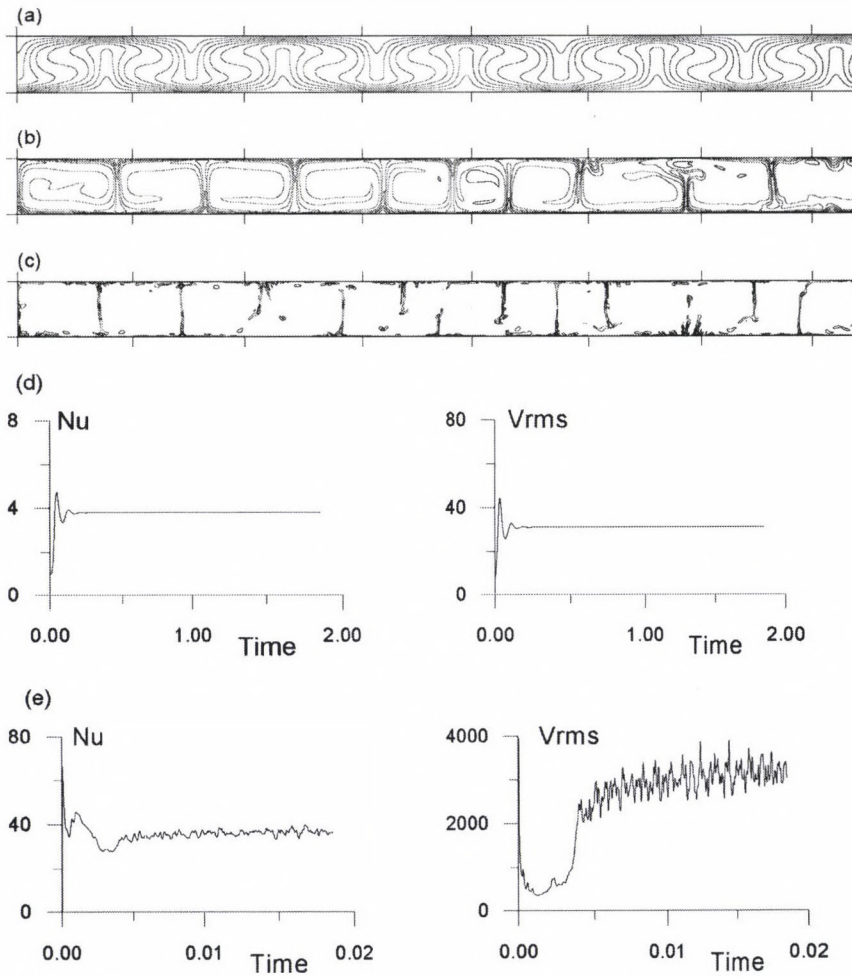


Fig. 2. Convection cells with homogenous viscosity distribution in a box with aspect ratio 16. Isotherms plotted in the same way as in Fig. 1. (a) $R = 5 \cdot 10^3$, (b) $R = 5 \cdot 10^5$, (c) $R = 10^7$. (d) The non-dimensional average surface heat flow and the non-dimensional root-mean-square velocity as a function of non-dimensional time at $R = 5 \cdot 10^3$. (e) Same as in Fig. 2d, but for $R = 10^7$

box can slightly vary in time. The circulation becomes more and more irregular at higher Rayleigh numbers, new up- and downwellings are born, while others die or migrate towards each other and merge (Fig. 2c). The temporal variation of the Nusselt number and the mean velocity for a stationary and a time-dependent system are compared in Figs 2d and 2e. We could not run experiments above $R = 10^7$ with the given grid size since the diameter of upwellings and downwellings became as small as 3–5 grid-points which did not allow accurate resolution.

The viscosity distribution of the mantle is not well constrained by observations (Pari and Peltier 1995), but theory based on the measured temperature- and pressure-dependence of creep in silicates shows that mantle viscosity should prob-

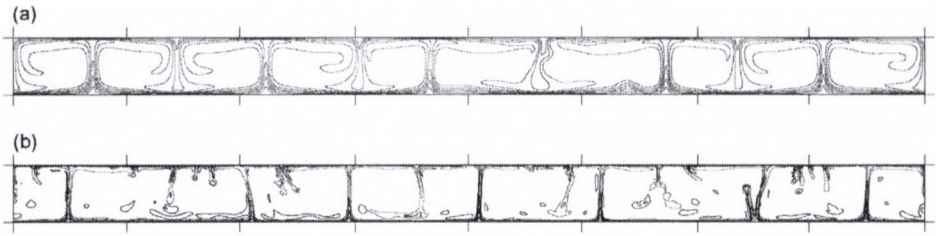


Fig. 3. Convective cell patterns for a viscosity increased exponentially by a factor of 10 from the surface to the bottom. (a) $R = 5 \cdot 10^5$, (b) $R = 10^7$

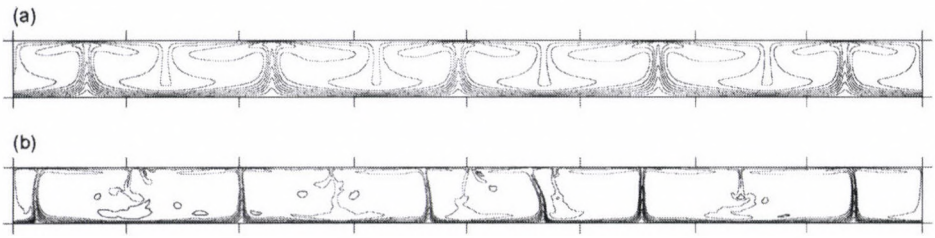


Fig. 4. Convective cell patterns for a viscosity increased exponentially by a factor of 100 from the surface to the bottom. (a) $R = 5 \cdot 10^5$, (b) $R = 10^7$

ably increase by one or two orders of magnitude from the upper mantle down to the CMB (van Keken and Ballentine 1998, Kellogg and King 1997, Ji and Nataf 1998a,b). This increase must play a great role in determining the structure of thermal convection. To explore this effect, we have run three models where the viscosity grows exponentionally by a factor of 10, 100 and 1000, respectively, from the surface down to the bottom of the box.

The quasi-steady convection patterns obtained at the same Rayleigh numbers, but for different viscosity distributions can be compared in Figs 2–5. When there is a significant viscosity increase with depth, the upwellings are broader in the lower mantle than in the upper mantle. Obviously, this is due to the fact that the viscosity which increases downwards leads to decreasing velocities, and to maintain the same vertical mass flux at all depths the plume must be wider down than in the upper part of the fluid. Figure 6 shows the average plume width as a function of the Rayleigh number and the viscosity contrast. The width of the plume is defined as the half-width of a Gaussian curve fitted to the temperature distribution at the depth of $z = 0.5$ (Ji and Nataf 1998a,b). Increasing the viscosity contrast seems to result in wider plumes, or in other words, higher Rayleigh numbers are needed to reach the regime of the same thin plumes if the viscosity contrast is high. This follows simply from the definition (6) of the Rayleigh number where we use the surface viscosity. If we used some suitably defined average viscosity in (6), the fitting straight lines of Fig. 6 could be more or less coincident.

Figure 7 summarizes the above models in terms of the number of cells (or number of plumes) as a function of the Rayleigh number. When the output of the compu-

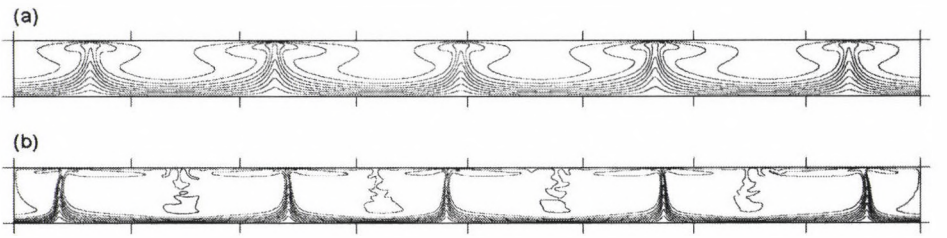


Fig. 5. Convective cell patterns for a viscosity increased exponentially by a factor of 1000 from the surface to the bottom. (a) $R = 5 \cdot 10^5$, (b) $R = 10^7$

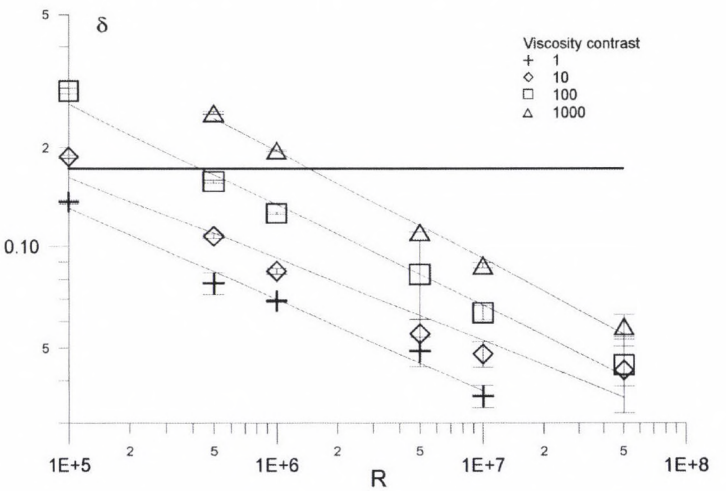


Fig. 6. Half-width (δ) of plumes as a function of the Rayleigh number in case of exponentially increasing viscosity. The viscosity contrasts between the bottom and top boundary: 1, 10, 100 and 1000. Gaussian shapes were fitted to the temperature distribution of the upwellings at the depth of $z = 0.5$. The thick horizontal line represents the width of 500 km, which is probably the maximum diameter of mantle plumes (e.g. Bijwaard and Spakman 1999). The data of the regression lines: (a) $\delta = 3.0 \cdot R^{-0.272}$, (b) $\delta = 2.69 \cdot R^{-0.244}$, (c) $\delta = 8.61 \cdot R^{-0.301}$, (d) $\delta = 17.1 \cdot R^{-0.324}$

tation is not stationary and the number of cells changes in time, we calculate an average cell number. To do this, the computation is run for a further 2000 iteration steps after reaching the quasi-steady state and the number of plumes (hotspots) is counted at every 200 steps (Fig. 8). We calculate the average from these counts. Although the relationship between the viscosity law and the number of hotspots cannot be easily established from the hotspot 'statistics' of Fig. 7, there is a clear tendency for a decreasing hotspot number with increasing viscosity contrast if the Rayleigh number is high enough ($R \geq 10^6$). It is not clear however whether the increase in the Rayleigh number (for a given viscosity distribution) has some straight and significant effect on the number of plumes.

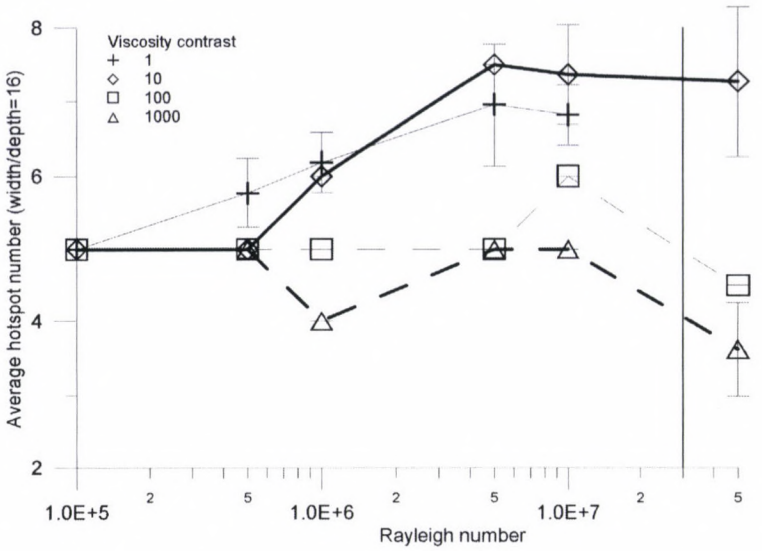


Fig. 7. Variation of the number of cells (or plumes) for increasing Rayleigh number and different viscosity distributions. When the flow is not stationary, the average cell number is indicated with its standard deviation. The typical Rayleigh number for the Earth's mantle is displayed by a vertical line. The viscosity contrast between the bottom CMB and the surface is marked in the figure by symbols

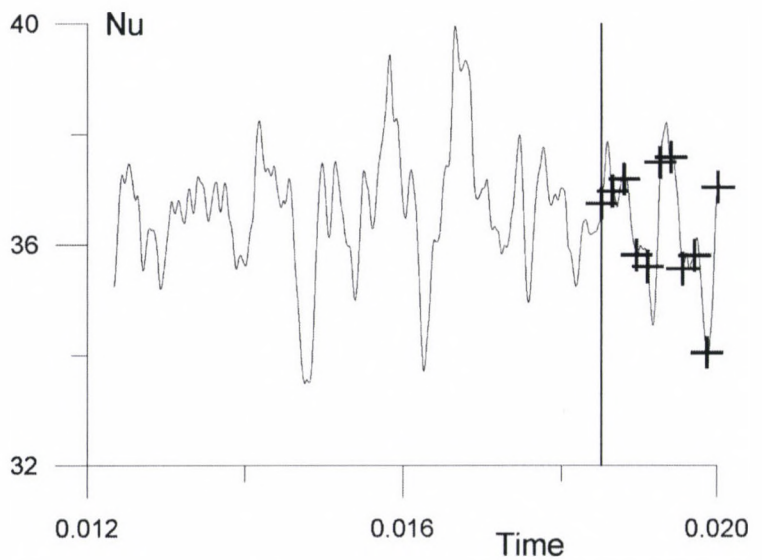


Fig. 8. Variation of the average heat flow for the case when the viscosity is constant, at $R = 10^7$. The 20000th iteration step is shown by the vertical line. From here, the number of cells were observed at every 200th steps as indicated by the crosses

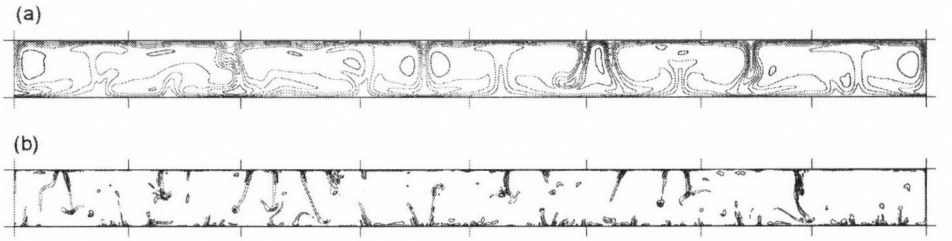


Fig. 9. Isotherms of the convection cells for the case where viscosity is constant and there is internal heating ($H = 10$). (a) $R = 10^5$, (b) $R = 10^7$

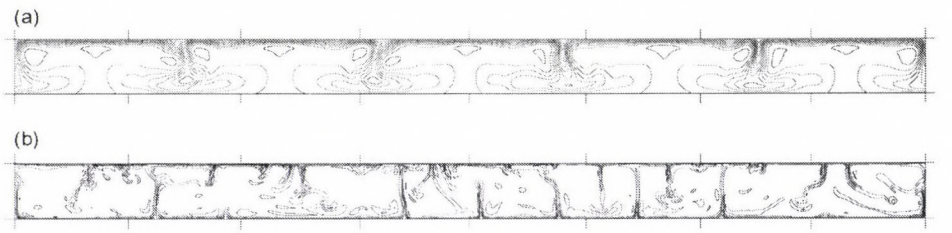


Fig. 10. Isotherms of the convection cells for exponentially increasing viscosity and internal heating ($H = 10$). The viscosity at the bottom is 10 times the surface value. (a) $R = 10^5$, (b) $R = 10^7$

3.3 The case of internal heating

The contribution of the internally produced (radioactive) heat to the total heating of the mantle must not be left out of consideration, because it is not only a major heating agent, but also it is already known to influence the character of the convection considerably (Travis et al. 1990). $H = 10$ has been chosen as the value of non-dimensional internal heating, which is generally accepted in mantle modelling and corresponds to a chondritic abundance of radioactive elements (van Keken and Ballentine 1998, Steinbach and Yuen 1998, 1999).

The case of constant viscosity is shown in Fig. 9 at two Rayleigh numbers, $R = 10^5$ and $R = 10^7$. The introduction of internal heating causes strong asymmetry. The downwellings are dominant features, their temperature differs remarkably from the surroundings, whereas the warm rising zones do not form well-developed upwellings. Therefore, at high Rayleigh numbers, we cannot see distinct convective cells.

The downwellings are even more vigorous for depth-dependent viscosity at low Rayleigh numbers (Figs 10, 11), but as R increases, plumes develop from a more and more pronounced bottom boundary layer and the downwellings become relatively weak. Another consequence of the presence of internal heating is that the final results in the models shown here are never stationary.

Figure 12 summarizes the internally heated cases in terms of the number of upwellings. Increase of the viscosity difference between the top and bottom boundary reduces the number of hotspots just as it was found without internal heating in Fig. 7.

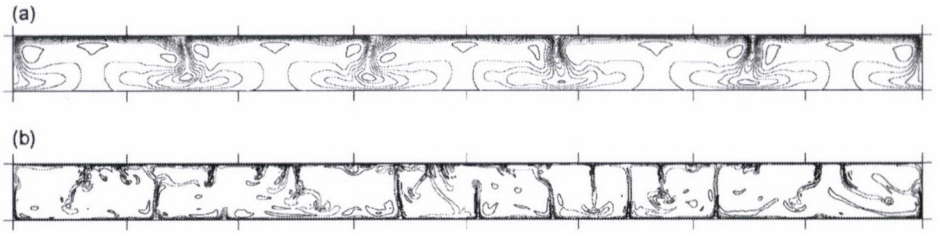


Fig. 11. Same as in Fig. 10, but the viscosity contrast between the bottom and top is 100

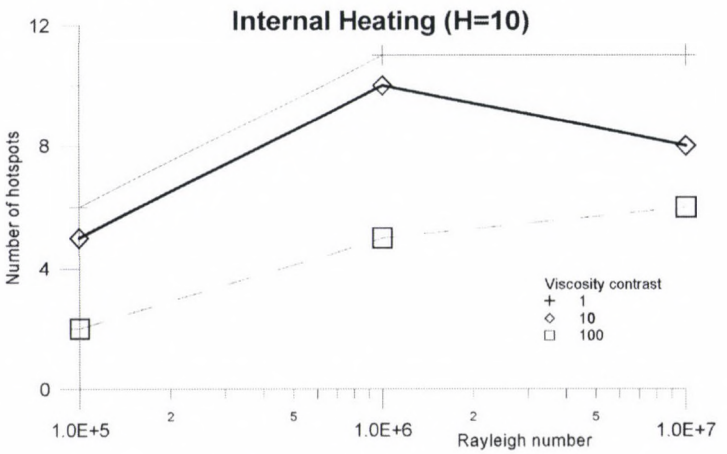


Fig. 12. Number of convection cells as a function the Rayleigh number in the presence of internal heating of $H = 10$ for exponentially increasing depth-dependent viscosity. The viscosity contrast is marked in the figure by symbols

It is worth rescaling some characteristic quantities, e.g. velocities, to real dimensional values. Of all our models, it is the case of Fig. 11b which might reproduce most closely the physical parameters of the real mantle convection. We have found in this model that the average horizontal velocity on the surface is 1–2 cm/y, but in some places it exceeds 3 cm/y. It is encouraging that the results of these relatively simple models approximate the plate-tectonic velocities very closely. By examining the plume shape, we can see that the plume is rather broad, 1200–1400 km, near the CMB, then it becomes thinner upwards and finally it has a narrow stem with a diameter of 300–400 km in the upper part of the mantle. Near the surface the upwelling material is spread over a large area and forms a mushroom shaped head about 200 km thick. These results are in perfect accordance with Bijwaard and Spakman's (1999) recent tomographic results on the Iceland plume. The average surface Nusselt number in the model of Fig. 11b is 18, which corresponds to a heat flux of 90 mW/m^2, slightly higher than the present-day terrestrial heat flow.

Conclusions

The present study has been an attempt to reconstruct the mantle flow by setting up simple two-dimensional models. It has been found that the internal heating and the depth-dependent viscosity have opposite effect on the development of ascending and descending currents. Internal heating reinforces the downwellings making them predominant vertical features of the flow, whereas the viscosity which increases significantly with depth strengthens the upwellings. If the viscosity increase is two orders of magnitude from the surface to the base of the convecting region, its influence becomes dominant at high Rayleigh numbers, even if the internal heating is strong, and this results in vigorous plume-like upwellings rooted in the bottom boundary layer. This model can be compatible with our knowledge about the mantle, and its quantitative predictions compare with basic geophysical quantities (plate velocities, surface heat flux).

Our study focuses on the number — or areal density — of upwellings. A main finding is that the average distance between two plumes increases with the bottom-to-top viscosity ratio in the cases of depth-dependent viscosity (if the Rayleigh number is high enough). This result holds for flows with internal heating as well.

The results of the models which are the closest with their input parameters to the Earth (e.g. Fig. 11b) can be applied to estimate the 'hotspot density' on a full sphere. Starting from the average plume distance of Fig. 11b, one finds that 6 or 7 plumes rising through the whole mantle should be detected in the Earth. This number is comparable to the number of cells found by Bercovici et al. (1989a,b) in full-sphere convection models. The Rayleigh number used in their study is less than the most probable value or the one applied in the model of Fig. 11b, but our results show that the increase of the Rayleigh number does not affect substantially the number of upwellings.

In contrast to these results, the various hotspot statistics enumerate several dozens of hotspots on the surface of the Earth. The difference cannot be attributed to insufficiencies of the whole-mantle models, e.g. to inaccurate knowledge on the applied physical coefficients. No model can produce such a dense plume system rising through the whole mantle as required by the observed number of hotspots. Although it is true that some hotspot clusters may reflect the presence of a single large mantle upwelling underneath the affected region, the revision of the geological data with regard to this possibility would not reduce the total hotspot number significantly. It seems that one has to assume inevitably an intermediate boundary layer within the mantle just as it has long been supposed by the advocates of upper-mantle convection (McKenzie et al. 1974). This boundary layer, very probably at the 660 km phase transition depth, can be the source of additional upwellings. It is easy to apply our models to an other 'end-member' case, i.e. where we assume a completely impermeable boundary at 660 km (the first end-member model being the one employed above where we assigned the total mantle depth to d, depth of the convective fluid). For a strict upper-mantle situation we could apply our models with pure basal heating and no significant viscosity variation. These models can fit heat flux and plate velocity data just as well as we have seen for the model of

Fig. 11b. For these models Fig. 7 suggests cca. 7 plumes in a box of aspect ratio 16. Extrapolating this again to a full sphere, we can estimate as many as 250 upper-mantle plumes. Now it seems that the reality must be somewhere half-way between the two end-member cases: to reproduce the real number of the Earth's hotspots (42 to 117, according to Crough and Jurdy 1980 and Vogt 1981), the 660 km discontinuity must not be an uninterrupted and completely impermeable boundary, and some of the plumes must be of whole-mantle character while others might come from the 660 km boundary layer.

This idea of two coexisting flow regimes is not new and has been supported by many arguments (Allcgre 1982, Solheim and Peltier 1994, Cserepes and Yuen 1997, Hofmann 1997) to which we can now add our considerations about the observed and calculated hotspot statistics. A more convincing argument could rely on full three-dimensional fluid-dynamical modelling. This should be the next step of our studies about the number of mantle plumes.

Acknowledgements

Support from the OTKA grant T026630 is gratefully acknowledged. A.G. thanks László Lenkey and Zsolt Regály for constructive remarks and useful suggestions during the modelling work.

References

Allcgre C J 1982: *Tectonophysics*, 81, 109–132.
Bercovici D, Schubert G, Glatzmaier G A 1989a: *Science*, 244, 950–955.
Bercovici D, Schubert G, Glatzmaier G A 1989b: *Geophys. Res. Lett.*, 16, 617–620.
Bijwaard H, Spakman W 1999: *Earth Planet Sci. Lett.*, 166, 121–126.
Blankenbach B, Busse F, Christensen U, Cserepes L, Gunkel D, Hansen U, Harder H, Jarvis G, Koch M, Marquart G, Moore D, Olson P, Schmeling H, Schnaubelt T 1989: *Geophys. J. Int.*, 98, 23–38.
Chandrasekhar S 1961: Hydrodynamic and hydromagnetic stability. Clarendon Press, Oxford
Crough S T, Jurdy D M 1980: *Earth Planet. Sci. Lett.*, 48, 15–22.
Cserepes L 1985: *Annales Universitatis Sci. Budapest. R. Eötvös, Sectio Geophys. et Meteor.*, 1–2, 52–67.
Cserepes L, Yuen D A 1997: *Geophys. Res. Lett.*, 24, 181–184.
Hofmann A W 1997: *Nature*, 385, 219–229.
Ji Y, Nataf H-C 1998a: *Earth Planet. Sci. Lett.*, 159, 87–98.
Ji Y, Nataf H-C 1998b: *Earth Planet. Sci. Lett.*, 159, 99–115.
Kellogg L H, King S D 1997: *Earth Planet. Sci. Lett.*, 148, 13–26.
Kennett B L N, Widiyantoro S 1999: *Earth Planet. Sci. Lett.*, 165, 145–155.
Leitch A M, Yuen D A 1989: *Geophys. Res. Lett.*, 16, 1407–1410.
McKenzie D P, Roberts J M, Weiss N O 1974: *J. Fluid Mech.*, 62, 465–538.
Mitronovas W, Isacks B L 1971: *J. Geophys. Res.*, 76, 7154–7180.
Morgan W J 1971: *Nature*, 230, 42–43.
Pari G, Peltier W R 1995: *J. Geophys. Res.*, 100, 12731–12751.
Robinson E M, Parsons B, Daly S F 1987: *Earth Planet. Sci. Lett.*, 82, 335–348.

Solheim L P, Peltier W R 1994: *J. Geophys. Res.*, 99, 6997–7018.
Steinbach V, Yuen D A 1998: *Earth Planet. Sci. Lett.*, 162, 15–25.
Steinbach V, Yuen D A 1999: *Earth Planet. Sci. Lett.*, 172, 213–220.
Travis B, Weinstein S, Olson P 1990: *Geophys. Res. Lett.*, 17, 243–246.
van der Hilst R D, Widiyantoro S, Engdahl E R 1997: *Nature*, 386, 578–584.
van Keken P E, Ballentine C J 1998: *Earth Planet. Sci. Lett.*, 156, 19–32.
Vogt P R 1981: *J. Geophys. Res.*, 86, 950–960.
Woodhouse J H, Dziewonski A M 1984: *J. Geophys. Res.*, 89, 5953–5986.

Acta Geod. Geoph. Hung., Vol. 35(4), pp. 397–414 (2000)

SLINGRAM MEASUREMENTS IN THE MECSEK MOUNTAINS, HUNGARY

E PRÁCSER[1], A ÁDÁM[2], L SZARKA[2], I MÜLLER[3], P TURBERG[3]

[Manuscript received April 21, 2000]

The so-called Slingram measurements form a family of frequency-domain electromagnetic methods, when the transmitter is a vertical magnetic dipole, and the vertical component of the magnetic field is measured, at some distance from the transmitter. The measured value is usually the imaginary (out-of-phase) part of the vertical magnetic field component. Such an equipment was manufactured at the University of Neuchâtel (Switzerland) and supplied to the Geodetic and Geophysical Research Institute in frame of a Swiss-Hungarian co-operation, supported by the Swiss National Research Foundation. As a part of test measurements, Slingram profilings and soundings were carried out for exploration of karstic water reservoirs connected to tectonically weak zones in the Mecsek Mountains, close to Pécs (Hungary). One-dimensional inversion and comparison with other geo-electromagnetic techniques as radio-magnetotellurics (RMT) and radiofrequency-electromagnetics (RF-EM) resulted in good and interpretable results.

Keywords: electromagnetic methods; karstic water reservoir; Mecsek Mountains; Slingram measurement

Introduction

The so-called Slingram measurements form a family of frequency-domain electromagnetic methods, when the transmitter is a vertical magnetic dipole, and the vertical component of the magnetic field is measured, at some distance from the transmitter. The measured value is usually the imaginary (out-of-phase) part of the vertical magnetic field component, but the real (in-phase) part, containing the primary field (that is the magnetic field in full free space) is also measured.

There are a lot of variants of Slingram measurements, e.g. airborne electromagnetic measurements, the so-called EM31 measurements, and the Apex MinMax technique. An equipment using the Slingram principle has been developed at the Hydrogeological Centre of the Neuchâtel University. The transmitter-receiver separation varies between 7.07 m – 56.56 m, in the operating frequency varies between 220 Hz and 14080 Hz.

The measured data themselves depend not only on the resistivity of the medium at depth, but they depend also on other parameters: the transmitter-receiver separation, the frequency, and the momentum of the transmitter. In order to remove

[1]Loránd Eötvös Geophysical Institute (ELGI), H–1145 Budapest, Columbus u. 17–23, Hungary
[2]Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H–4901 Sopron, POB 5, Hungary
[3]Université de Neuchâtel, Center d'Hydrogéologie, Rue Emile-Argand II, CH–2000 Neuchâtel, Switzerland

1217-8977/2000/$ 5.00

the effects of these latter parameters, the measured data are usually transformed into apparent resistivities. In Slingram measurements the apparent resistivity is usually computed from the imaginary part of the magnetic field, but it could be equally calculated from other components: from the real part, from the absolute value, and even from the phase. The apparent resistivity computation had a greater significance decades ago, when the data processing had a limited possibility. Nowadays, with the advent of fast computers, layer parameters can be directly determined from the response function. The inverted parameters (depths and layer resistivities) give of course more information about the subsurface structure, than the apparent resistivity itself.

In this paper the results of Slingram test measurements, carried out in the Mecsek Mountains (Southern Transdanubia, Hungary) are shown, where near- surface structural changes in connection of possible karstic water-reservoirs were studied.

In the first part of the paper the theoretical aspects of the Slingram modelling and inversion are summarised, then the field results are presented.

Mathematical modelling of Slingram measurements for layered earth

A widely used data processing method in geophysics is the one-dimensional inversion. It means, that an algorithm searches for layer parameters (resistivities and thicknesses), having the same electromagnetic response, which was measured in the field. The inversion algorithm needs the solution of the direct problem. Let us review briefly the basic formulae. The vertical and horizontal components of the electromagnetic field due to a vertical magnetic dipole is given by a Hankel transform:

$$H_z(\omega) = \frac{IS}{4\pi}\left\{\int_0^\infty J_0(\lambda r)\lambda^2 e^{-\lambda(z+h)}R_0(\lambda)d\lambda - \frac{1}{R^3} + \frac{3(z-h)^2}{R^5}\right\}$$
$$H_r(\omega) = \frac{IS}{4\pi}\left\{\int_0^\infty J_1(\lambda r)\lambda^2 e^{-\lambda(z+h)}R_0(\lambda)d\lambda + \frac{3r(z-h)}{R^5}\right\} \tag{1}$$

where IS is the momentum of the dipole, $R = \sqrt{r^2 + (z-h)^2}$, $\omega = 2\pi f$, and $R_0(\lambda)$ is the kernel function depending on the layer parameters, given by the recursion formula:

$$R_j(\lambda) = \frac{\dfrac{\lambda_j - \lambda_{j+1}}{\lambda_j + \lambda_{j+1}} + R_{j+1}(\lambda)e^{-2\lambda_{j+1}d_{j+1}}}{1 + \dfrac{\lambda_j - \lambda_{j+1}}{\lambda_j + \lambda_{j+1}}R_{j+1}(\lambda)e^{-2\lambda_{j+1}d_{j+1}}}$$

$\lambda_j = \sqrt{\lambda^2 + i\omega\mu\sigma_j}$, h, z are the height of the transmitter, and of the receiver correspondingly (in our case, if both the transmitter and the receiver are at the surface of a horizontal half-space $h = z = 0$), σ_j is the conductivity of the j-th layer, d_j is the thickness of the j-th layer, J_0, J_1 are Bessel functions.

In Slingram measurements only the vertical component of the electromagnetic field is measured. However, when the plane of the receiver coil is not parallel to the surface, the measured value is effected also by the horizontal component of the magnetic field.

In the case when the plane of the transmitter coil is not parallel to the surface, the momentum of the transmitter can be separated into vertical and horizontal components, therefore the magnetic field due to a horizontal magnetic dipole is also necessary to be taken into account.

The vertical component of the magnetic field due to a horizontal magnetic dipole:

$$H_z(\omega) = \frac{IS}{4\pi}\left\{-\int_0^\infty J_1(\lambda r)\lambda^2 e^{-\lambda(z+h)} R_0(\lambda)d\lambda + \frac{3(z-h)r}{R^5}\right\}.$$

The horizontal component of the magnetic field:

$$\begin{aligned} H_r(\omega) \quad = \quad & \frac{IS}{4\pi}\left\{-\frac{1}{r}\int_0^\infty J_1(\lambda r)\lambda e^{-\lambda(z+h)} R_0(\lambda)d\lambda\right\} + \\ & \frac{IS}{4\pi}\left\{\int_0^\infty J_0(\lambda r)\lambda^2 e^{-\lambda(z+h)} R_0(\lambda)d\lambda - \frac{2}{R^3}\right\}. \end{aligned} \tag{2}$$

In case of tilted transmitter and receiver, at the transmitter side both the horizontal and vertical component due to the transmitter dipole has to be taken into account, and at the receiver side both the vertical and horizontal components of the magnetic field should be computed. With a tilted system the measured value is given by the formula:

$$\begin{aligned} H_z^{(0)}(\omega) \quad = \quad & (H_z^{(v)}(\omega)\cos\Theta - H_z^{(h)}(\omega)\sin\Theta)\cos\phi - \\ & -(H_r^{(v)}(\omega)\cos\Theta - H_r^{(h)}(\omega)\sin\Theta)\sin\phi\,, \end{aligned}$$

where $H_z^{(0)}$ is the measured field component, $H_z^{(v)}$, $H_r^{(v)}$ are the magnetic field components due to the vertical magnetic dipole, $H_z^{(h)}$, $H_r^{(h)}$ are the magnetic field components due to the horizontal magnetic dipole, and Θ, ϕ are the tilt angle of the transmitter and the receiver, correspondingly.

If the misorientation angles Θ and ϕ were exactly known, there would be no problem to carry out an exact one-dimensional modelling and inversion. If the misorientation angles are not known, in the one-dimensional inversion Θ and ϕ angles should be considered as unknowns. In this latter case the inversion results are not only the layer parameters, but also the two tilt angle values. This kind of inversion is possible to be done in an appropriate induction number domain, with sufficient number of frequencies.

Apparent resistivity computation

For small induction numbers, the apparent resistivity can be computed simply from the imaginary part of the magnetic field. If the induction number $Q = r^2\mu\sigma\omega$ is small (much smaller than 1), then the recursion formula, which determines the kernel function in formulae (1) and (2) becomes simpler:

$$R_j(\lambda) \approx \frac{i\omega\mu(\sigma_j - \sigma_{j+1})}{4\lambda^2} + R_{j+1}(\lambda)e^{-2\lambda d_j} .$$

Then $\mathrm{Im}H_z$ is obtained as follows:

$$\begin{aligned} \mathrm{Im}H_z \quad = \quad & -\frac{IS}{4\pi}\left(\frac{\sigma_1\omega\mu}{4r} + \frac{(\sigma_2-\sigma_1)\omega\mu}{4\sqrt{r^2+4d_1^2}} + \frac{(\sigma_3-\sigma_2)\omega\mu}{4\sqrt{r^2+4(d_1+d_2)^2}} + \ldots \right. \\ & \left. + \frac{(\sigma_n-\sigma_{n-1})\omega\mu}{4\sqrt{r^2+4(d_1+d_2+\ldots+d_{n-1})}}\right) . \end{aligned} \tag{3}$$

The kernel function $R_0(\lambda)$ for homogeneous half-spaces is as follows:

$$R_0(\lambda) = \frac{\lambda-\lambda_1}{\lambda+\lambda_1} \approx \frac{i\omega\mu\sigma_1}{4\lambda^2} .$$

From the kernel function the vertical component of the magnetic field over a homogeneous half-space with conductivity σ_1 is obtained as by the following formula:

$$\mathrm{Im}H_z \approx -\frac{IS}{4\pi}\frac{\omega\mu\sigma_1}{4\sqrt{r^2+(z+h)^2}} .$$

Then the σ_1 conductivity can be expressed as a function of $\mathrm{Im}H_z$. In this way over a more complicated structure the measured $\mathrm{Im}H_z$ can be transformed into apparent conductivity as follows:

$$\sigma_a = \frac{4\mathrm{Im}H_z}{H_z^{(p)}\omega\mu r^2} , \tag{4}$$

where $H_z^{(p)}$ the primary magnetic field, and both the transmitter and the receiver are at the surface ($z = 0$, $h = 0$). The reciprocal value of the apparent conductivity is the apparent resistivity:

$$\varrho_a = \frac{H_z^{(p)}\omega\mu r^2}{4\mathrm{Im}H_z} .$$

After substituting formula (3) into formula (4) we obtain a very informative form for the apparent conductivity over a layered half-space:

$$\begin{aligned} \sigma_a \quad = \quad & \sigma_1 - \sigma_1 P(z_1) + \sigma_2\left(P(z_1) - P(z_2)\right) + \ldots \\ & \ldots + \sigma_{n=1}\left(P(z_{n-2}) - P(z_{n-1})\right) + \sigma_n P(z_{n-1}) , \end{aligned} \tag{5}$$

where

$$P(z_i) = \frac{1}{\sqrt{\left(\frac{2z_i}{r}\right)^2 + 1}}, \qquad z_i = \sum_{j=1}^{i} d_j .$$

It should be observed for small induction numbers that the apparent conductivity (and also its reciprocal value, the apparent resistivity) is independent of the frequency. It depends only on the layer parameters and on the transmitter-receiver separation. In this case the penetration depth of the Slingram measurement is independent of the frequency. If we want to obtain information from different depths at a measuring site, we need to change the transmitter-receiver separation.

One-dimensional inversion

The goal of electromagnetic geophysics is to determine the conductivity distribution of medium below the surface. When the conductivity does not change in horizontal direction, that is the model is layered, a so called one-dimensional inversion can be performed. Its results are the layer parameters: resistivities and thicknesses. In the inversion it is assumed that for small variations in the layer parameters, the observed data are linear functions of the layer parameters. Shortly let us look over the theoretical aspects of the inversion.

The components of the vector $\mathbf{m}$ are the $\mathrm{Re}H_z$, or $\mathrm{Im}H_z$ values, measured at a measuring site, using different frequencies or/and different transmitter-receiver separations. The components of the vector $\mathbf{p}$ are the layer parameters. The inversion is carried out in several iteration steps. Let us assume that the $\mathbf{p}_i$ estimation of the parameter vector is known. The data vector belonging to it is $\mathbf{m}_i$. $\mathbf{m}_0$ means the measured data. Let $\Delta\mathbf{m} = \mathbf{m}_0 - \mathbf{m}_i$, and $\mathbf{p}_{i+1} = \mathbf{p}_i + \Delta\mathbf{p}$. The parameter vector must be modified in such a way, that the resulting $\mathbf{m}_{i+1}$ data vector belonging to the parameter vector gets closer to $\mathbf{m}_0$. The vector $\Delta\mathbf{p}$ can be computed solving the equation

$$\Delta\mathbf{m} = \mathbf{J}\Delta\mathbf{p} . \tag{6}$$

The $\mathbf{J}$ Jacobian matrix contains partial derivatives, which are known from the direct calculation. $\mathbf{m}_i$ is known from $\mathbf{p}_i$. Since $\mathbf{m}_0$ is measured, $\Delta\mathbf{m}$ is known, too. If matrix $\mathbf{J}$ in Eq. (6) were an invertable square matrix, then we could easily obtain the unique solution. Unfortunately in most cases $\mathbf{J}$ is not an invertable square matrix. A generalized inverse is based on the singular value decomposition (SVD) of the Jacobian matrix $\mathbf{J}$ (Jackson 1972):

$$\mathbf{J} = \mathbf{U}\Lambda\mathbf{V}^T, \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}, \quad \mathbf{V}^T\mathbf{V} = \mathbf{I}, \tag{7}$$

where $\mathbf{I}$ is the identity matrix, and the Λ is a diagonal matrix containing

the eigenvalues:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix} .$$

The generalized inverse is obtained as follows:

$$\mathbf{J}^+ = \mathbf{V}\Lambda^{-1}\mathbf{U}^T .$$

In the knowledge of $\Delta\mathbf{m}$, $\Delta\mathbf{p}$ is to be computed with the following formula:

$$\mathbf{V}\mathbf{V}^T\Delta\mathbf{p} = \mathbf{J}^+\mathbf{U}\mathbf{U}^T\Delta\mathbf{m} ,$$

where $\mathbf{R} = \mathbf{V}\mathbf{V}^T$ is the resolution matrix, $\mathbf{S} = \mathbf{U}\mathbf{U}^T$ is the information density matrix.

Inasmuch the resolution matrix is the identity matrix, each element of $\Delta\mathbf{p}$ is individually given, otherwise only some linear combination of the elements can be determined. E.g. for the layer parameters it means that only the product of the thickness and the conductivity of a thin conductive layer can be given. After solving Eq. (6), a sum of $\mathbf{p}_i$ and $\Delta\mathbf{p}$ provides a new estimation for the layer parameters. Then, with the modified layer parameters, a new forward computation is performed, resulting in a new iteration step. The iteration process stops, when either the norm of $\Delta\mathbf{m}$ is small, or $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$ do not differ any more from each other. The results are reliable, if the number of the measured data (size of $\mathbf{m}$) exceeds the number of layer parameters (size of $\mathbf{p}$), and the initial model is not very far from the reality.

The multi-dimensional case

It is not guaranteed that the medium is a one-dimensional one. There are already two- and three-dimensional inversion algorithms, working on a similar principle as the one-dimensional inversion does. The main difficulty with the multi-dimensional inversions is, that they are rather time-consuming, even if the fastest computers are used. The reliability of the obtained model is often questionable, too. Despite of these problems, in some special cases they may give reasonable results. A plot of the measurements along a profile, together with some additional geological information tells us, whether the structure below the surface is two- or three-dimensional. In order to tell, whether the source of the anomaly is a special multi-dimensional structure, forward computations are used. For example the electromagnetic field due a conductive thin plate is known. The fracture zones at depths may be modelled with the plate model. In this case it is already possible to carry out an inversion with some limitations, of course. It can be determined e.g., if the anomaly observed along a profile is equivalent with the effect of a plate at a given depth with a given tilt angle and conductance.

Case histories: Measurements in the Mecsek Mountains

In the area of Mecsek Mts. Slingram profilings and soundings have been used for the exploration of the karstic water reservoirs connected to the tectonically weak zones (fracture, faults, etc.).

The Slingram instrument we used was manufactured at the Neuchâtel University (Hydrogeological Centre) and supplied to the Geodetic and Geophysical Research Institute in the frame of a grant given by the Swiss National Science Foundation. The measurements have been carried out with vertical magnetic dipole in the field at 4 frequencies (440, 1760, 7040 and 14080 Hz) with constant distance of 40 m between the transmitter and receiver loops.

Besides the Slingram measurements other electromagnetic techniques: RF-EM and RMT radiofrequency methods were also applied. These measurements enable us to reduce the ambiguity of the data.

Case history A

The main karstic water source in the Mecsek Mts. is the so-called "Tettye spring" in Triassic limestone. This cannot supply enough water for the town Pécs, therefore expensive water transport is needed from the Danube region. Secondary water resources in the Mecsek Mts. could decrease these high expenses. It is suspected that the N-S valleys running from the mountains towards the town could also represent karstic water reservoirs as the water springs on their southern ends point to that direction. Nevertheless, the water close to the town is already contaminated by inhabitants living around the valleys, having neither tap-water nor canalization. Therefore the aim of the measurements with Slingram and other techniques was to find water reservoirs at the highest points of the valleys where the karstic water is not yet contaminated.

The karstic water reservoirs in the limestone have lower electric resistivity than the undisturbed (fresh) limestone, therefore they can be followed by geoelectric methods, among them by Slingram and by the radiofrequency methods, too. The Slingram profilings have been carried out through the valleys which were crossed, following the tourist paths.

In Fig. 1a the out-of-phase components along the profile A are shown having two definite decreases in amplitude, corresponding to smooth topographical deepenings (valleys). The anomalies appear first of all at the highest frequencies (7040 and 14080 Hz), referring to the close-to-surface character of the resistivity change. The same tendency was recognized in the inphase components (Fig. 1b). It is to be noted that at the end of the profile there is a strong increase in both components, probably due to artificial sources (water tubes or electric lines) in the soil.

As a result of one-dimensional inversion of this dataset 2 or 3 layers could be determined. The fitting between the measured and the modelled data is illustrated in Figs 2a and b, showing a good (Fig. 2a) and a relatively bad fitting (Fig. 2b). The fitting is generally much better in case of the out-of-phase components.

The layer sequence along the profile A is given in Fig. 3. The great increase

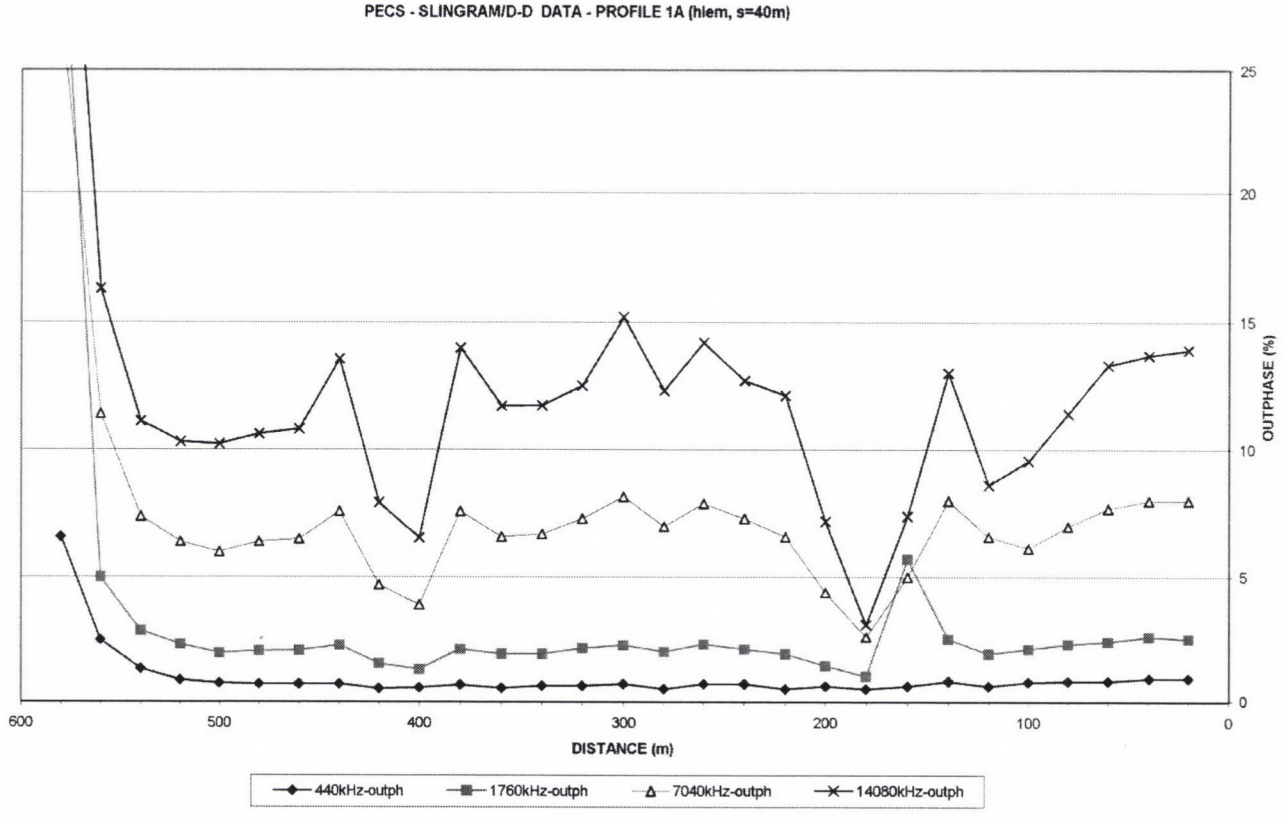


Fig. 1a. Slingram profiles along profile A. Out-of-phase data

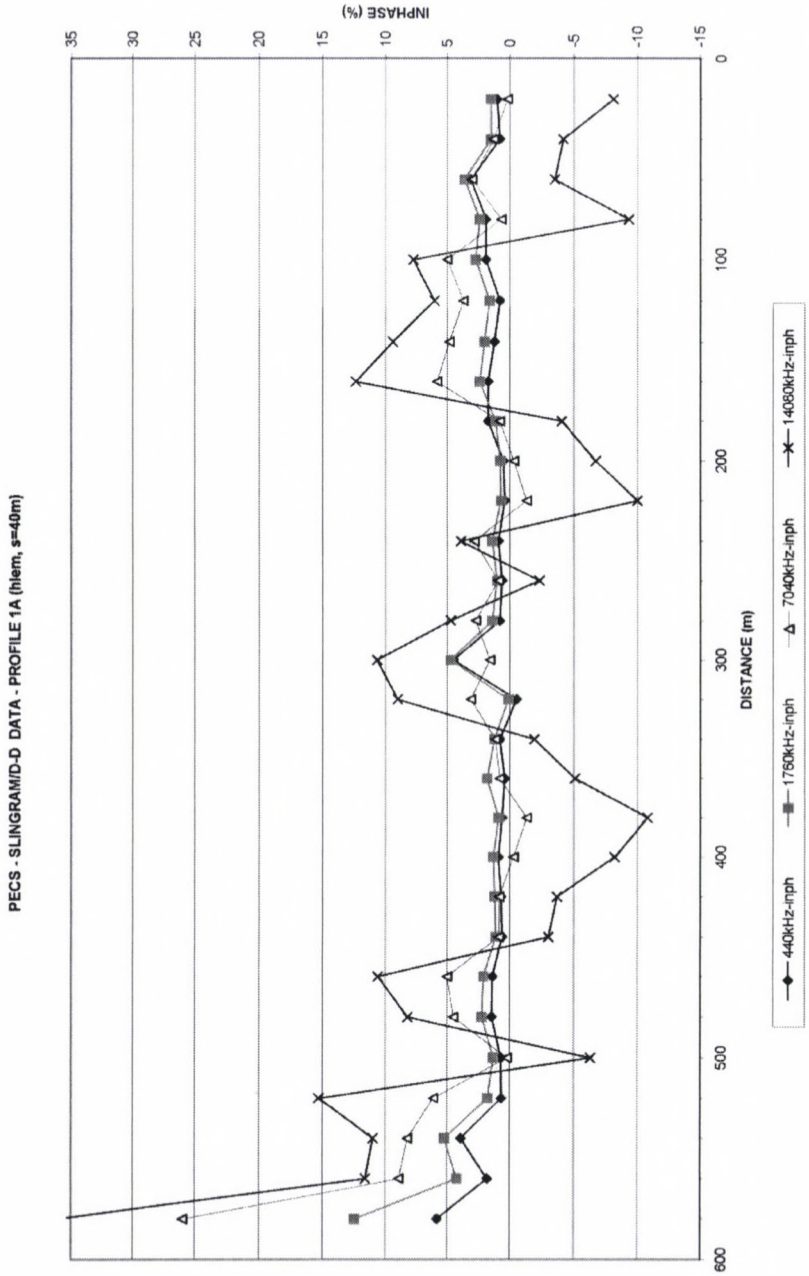


Fig. 1b. Slingram profiles along profile A. Inphase data

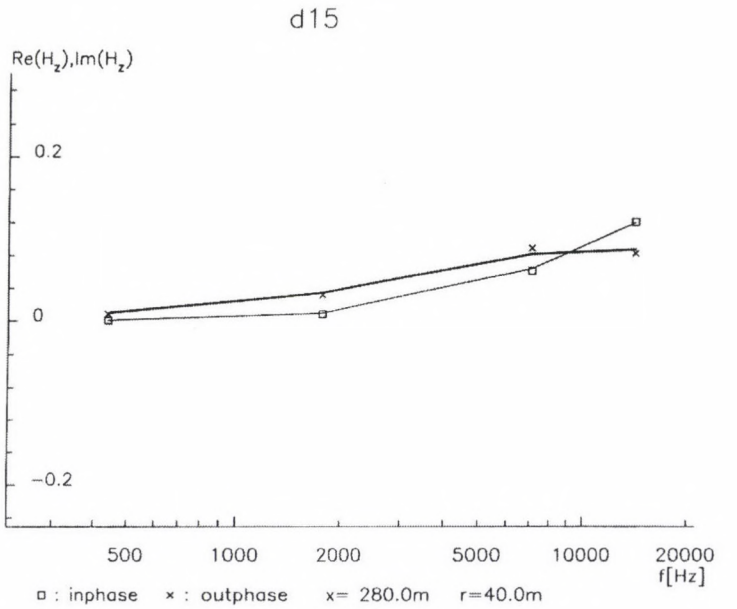


Fig. 2a. Characteristic misfits between the observed and inverted data. An example for a good fitting

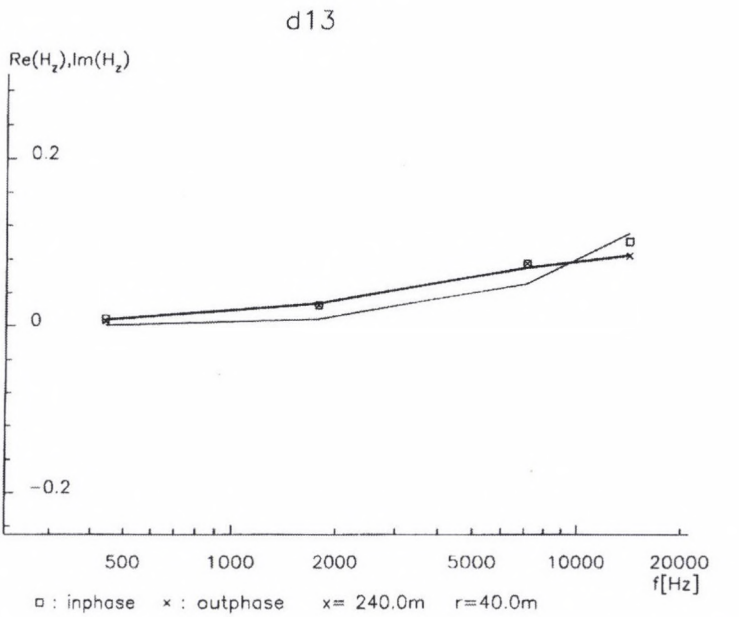


Fig. 2b. Characteristic misfits between the observed and inverted data. An example for a bad fitting

in the out-of-phase component between x = 180 and 260 corresponds to a strong conductive zone with thickness greater than about 50 m under a resistive cover. Along the whole profile an about 10 m thick conductive layer could be determined at a depth of about 10–15 m, which thickens between x = 450 and 500 m. These conductive zones could be perspective for water supply.

Case history B

Along another profile having zero misorientation effect due to a favourable topography very curious Slingram and other electromagnetic data were obtained. This profile crosses very different rock ensembles with water spring at their contact. As it is shown in Fig. 4a, the out-of-phase components have positive and negative peaks. The most interesting anomaly appears before the so called "black rock spring" with a very high negative peak, which — due to its negative sign — cannot be transformed into resistivity values.

The negative peak in the out-of-phase components is followed by a positive increase, which corresponds to a resistivity decrease. This is certainly due to water saturated rocks as indicated by the water outflow at the black rock spring. There are a few consecutive similar phenomena along this profile, hinting at the very complex geological structure of the area. The character of the inphase components is quite the same (as shown in Fig. 4b), having peaks at the same sites, where the out-of-phase components do. The amplitudes in both components reach here also their greatest values at higher frequencies hinting again at the close-to-surface character of the subsurface structure A. 1D inversion of these data resulted in the layer profile shown in Fig. 5. These zones have been found around sites x = 100, 225 m and at the end of the profile near the black rock spring (x = 400 m). Their thickness can reach 40–50 m.

Figure 6 shows a comparison between Sligram out-of-phase values and the RF-EM (216 kHz) out-of-phase data. Although the frequency is much higher in the latter case, the character of the indications is very similar. The only exception is observed around the black-rock spring from where the RF-EM curve becomes very flat. Its relative amplitudes are almost the same as those of the Slingram measurements. The RMT data are plotted together with RF-EM out-of-phase values in Fig. 7. The aforementioned flat character of the RF-EM out-of-phases at the black rock spring similarly appears in the RMT resistivity data, nevertheless, after their long and continuous decrease, starting before the "RÁBAI-fa". This means that the RMT resistivity also reaches its lowest value at the black rock spring corresponding to the possible high water saturation of the rocks.

Conclusion

The Slingram profiling/sounding can indicate those near-surface structural changes which are in connection with water reservoir where the rocks are saturated by water. Therefore this method can be applied to explore water reservoirs in karstic area of the Mecsek Mts. in Triassic limestone, too.

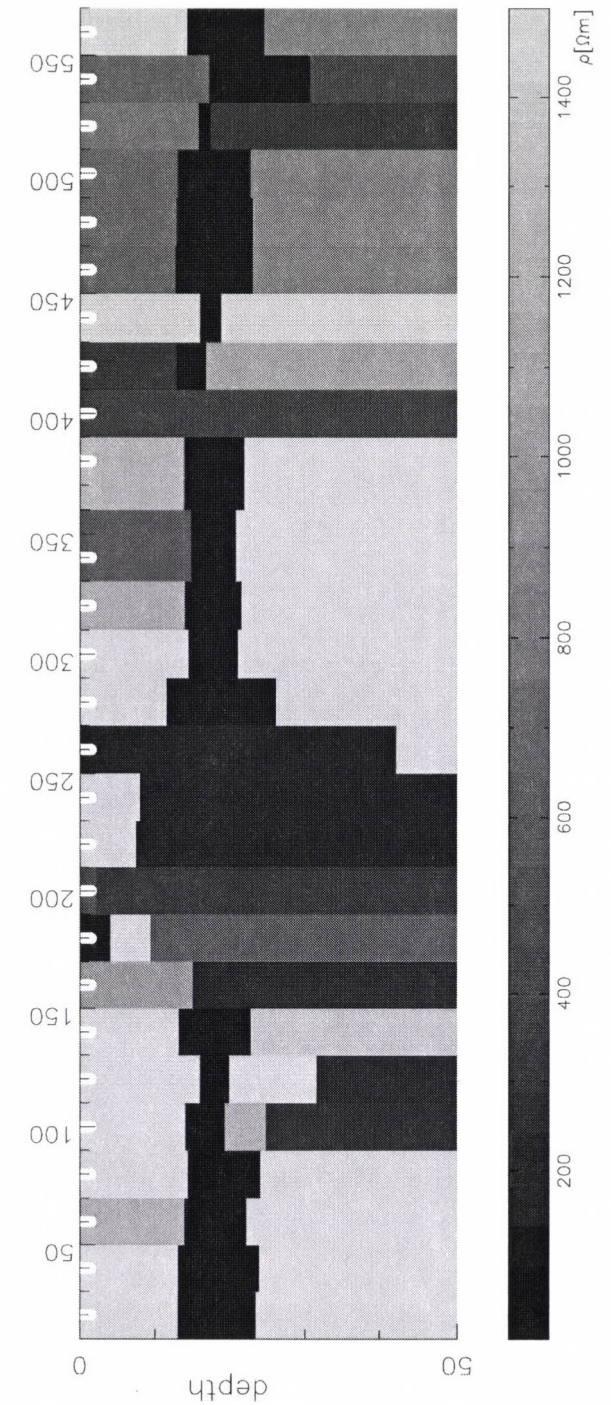


Fig. 3. Inverted resistivity section along profile A as obtained from one-dimensional inversion

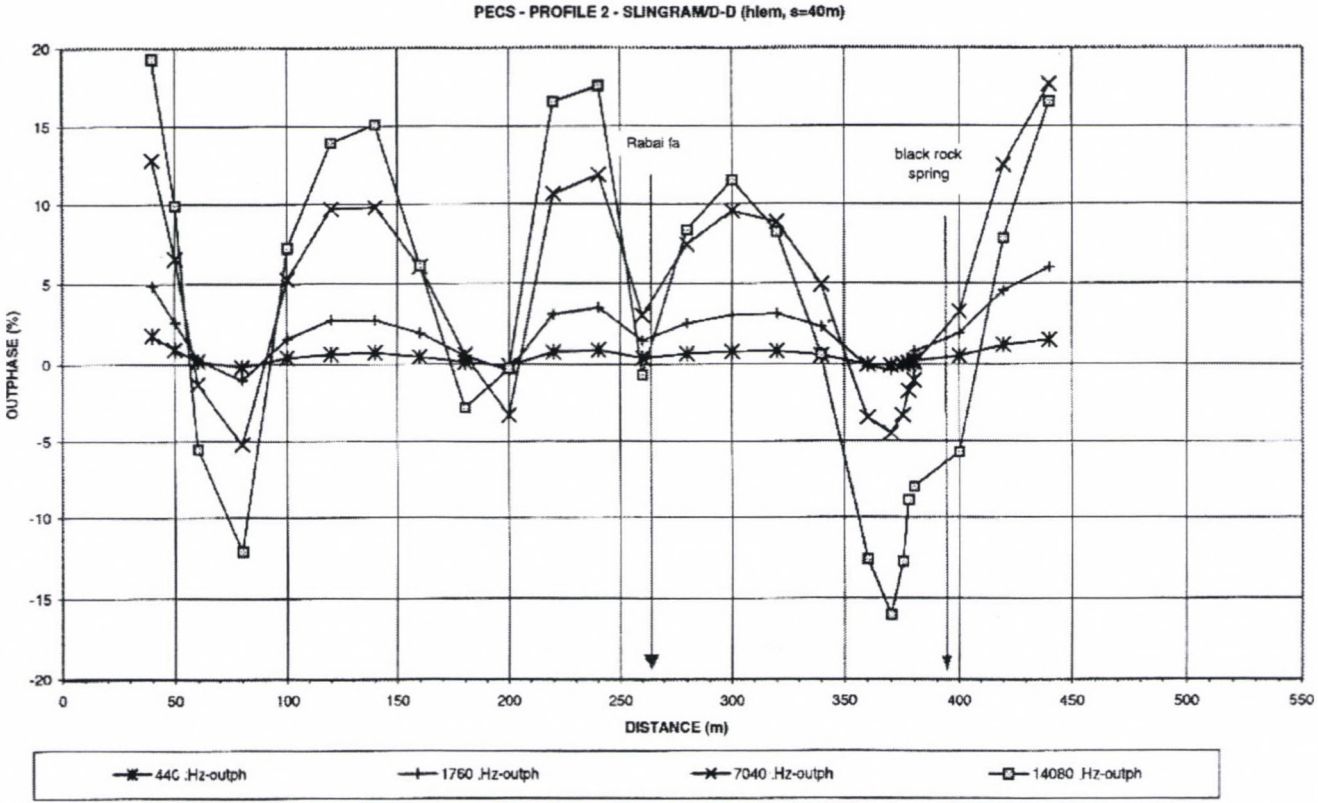


Fig. 4a. Slingram profiles along profile B (at the Rabai fa). Out-of-phase data

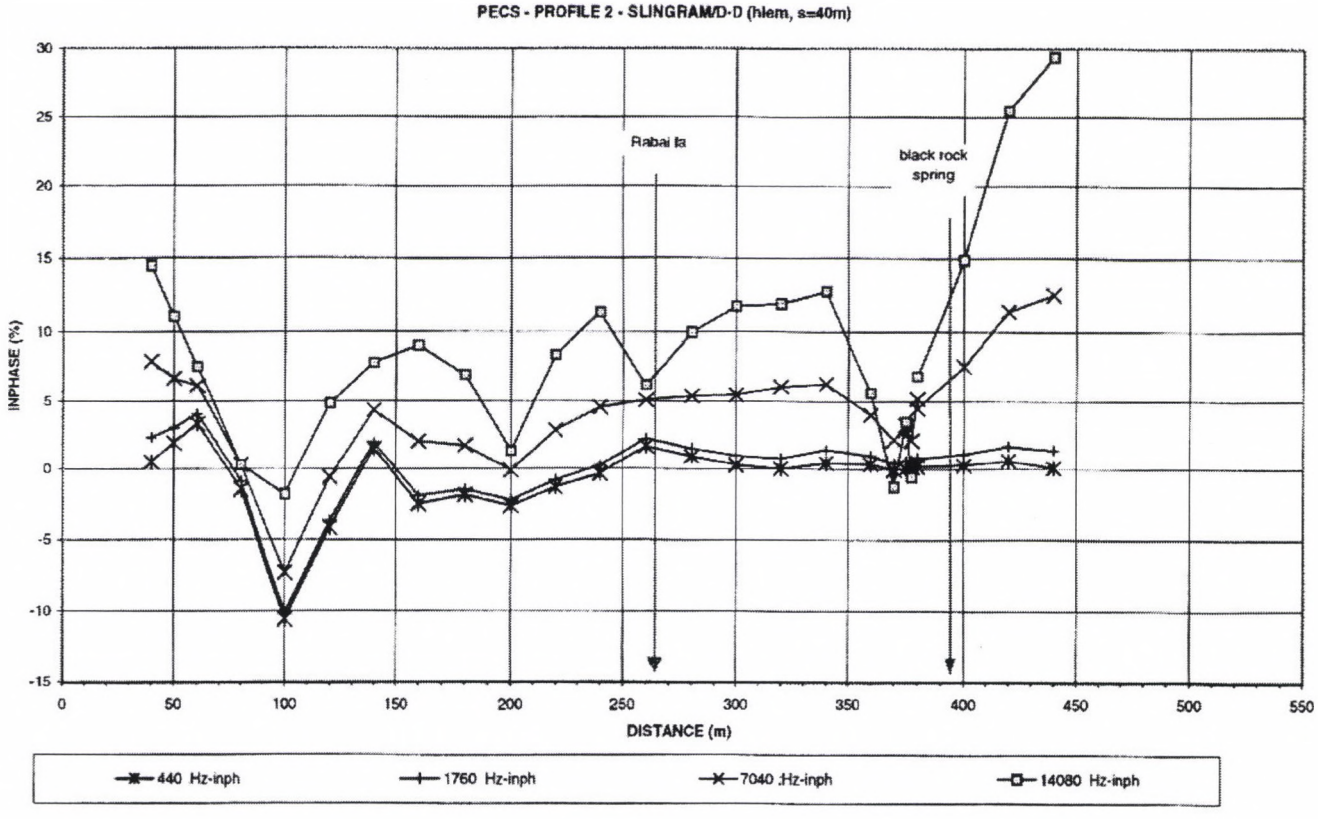


Fig. 4b. Slingram profiles along profile B (at the Rabai fa). Inphase data

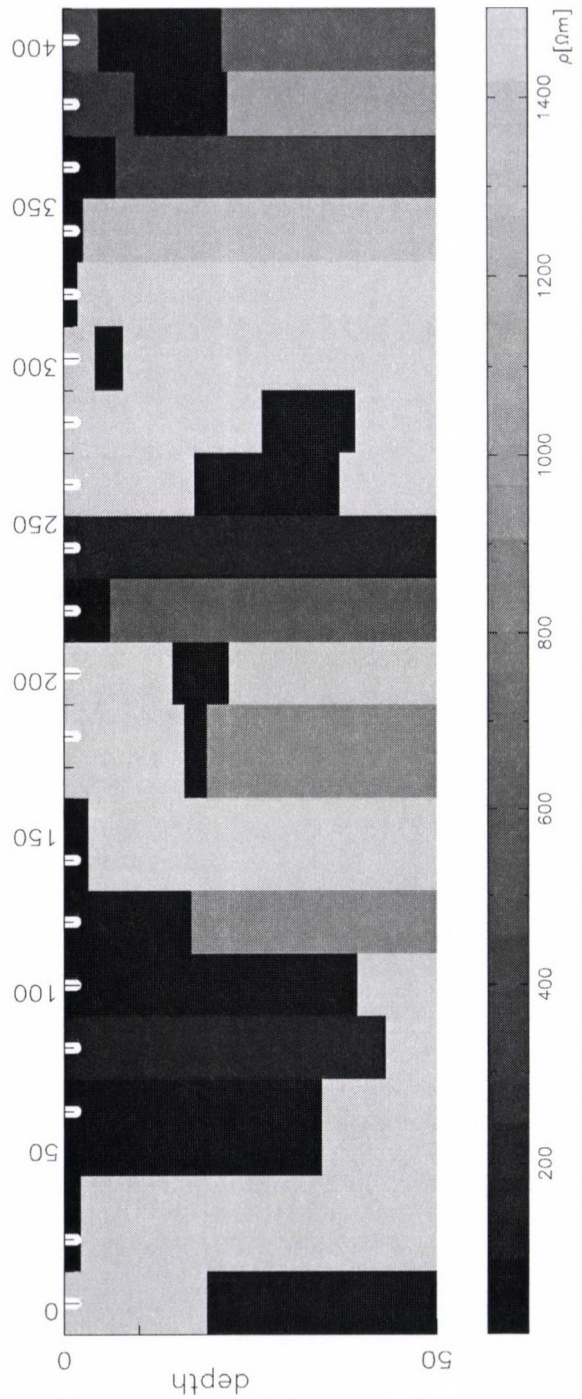


Fig. 5. Inverted resistivity section along profile B as obtained from one-dimensional inversion

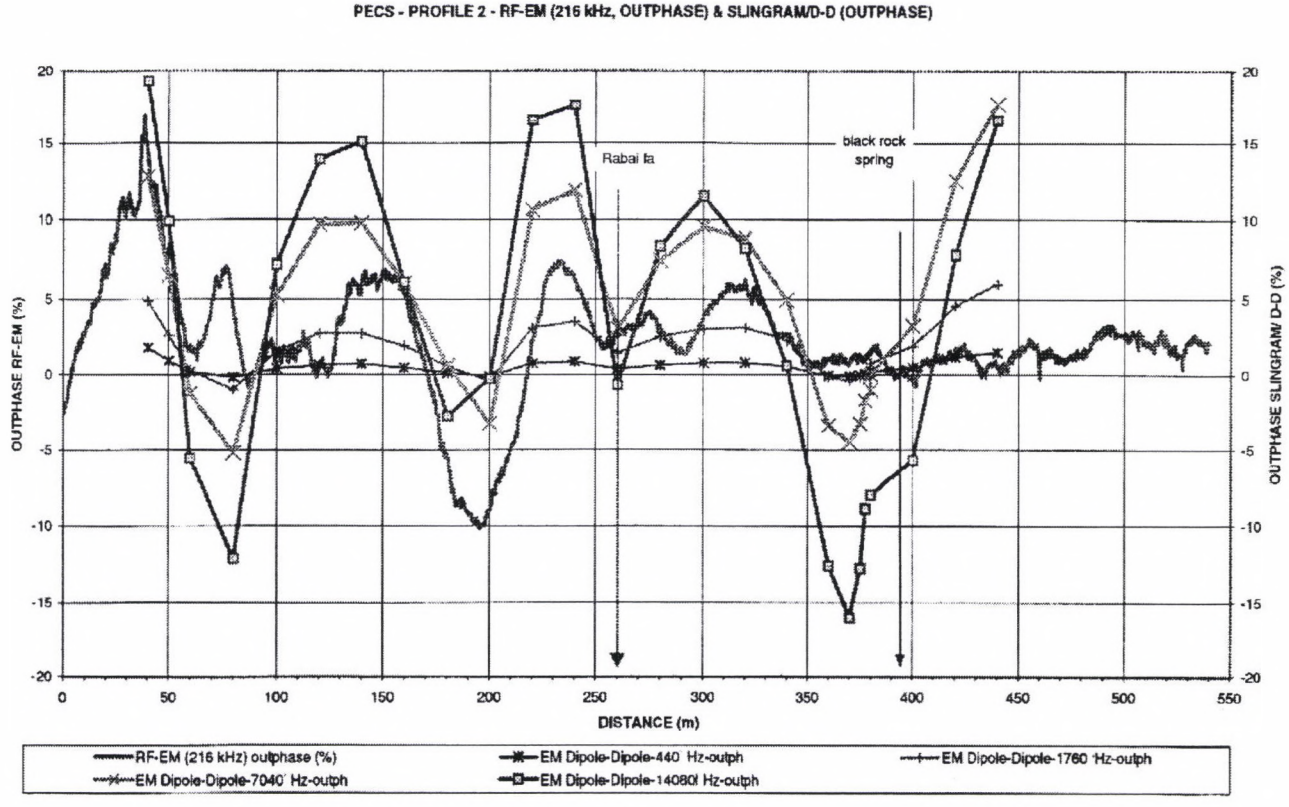


Fig. 6. RF-EM out-of-phase data along profile B at 216 kHz, as compared to Slingarm out-of-phase data at different frequencies

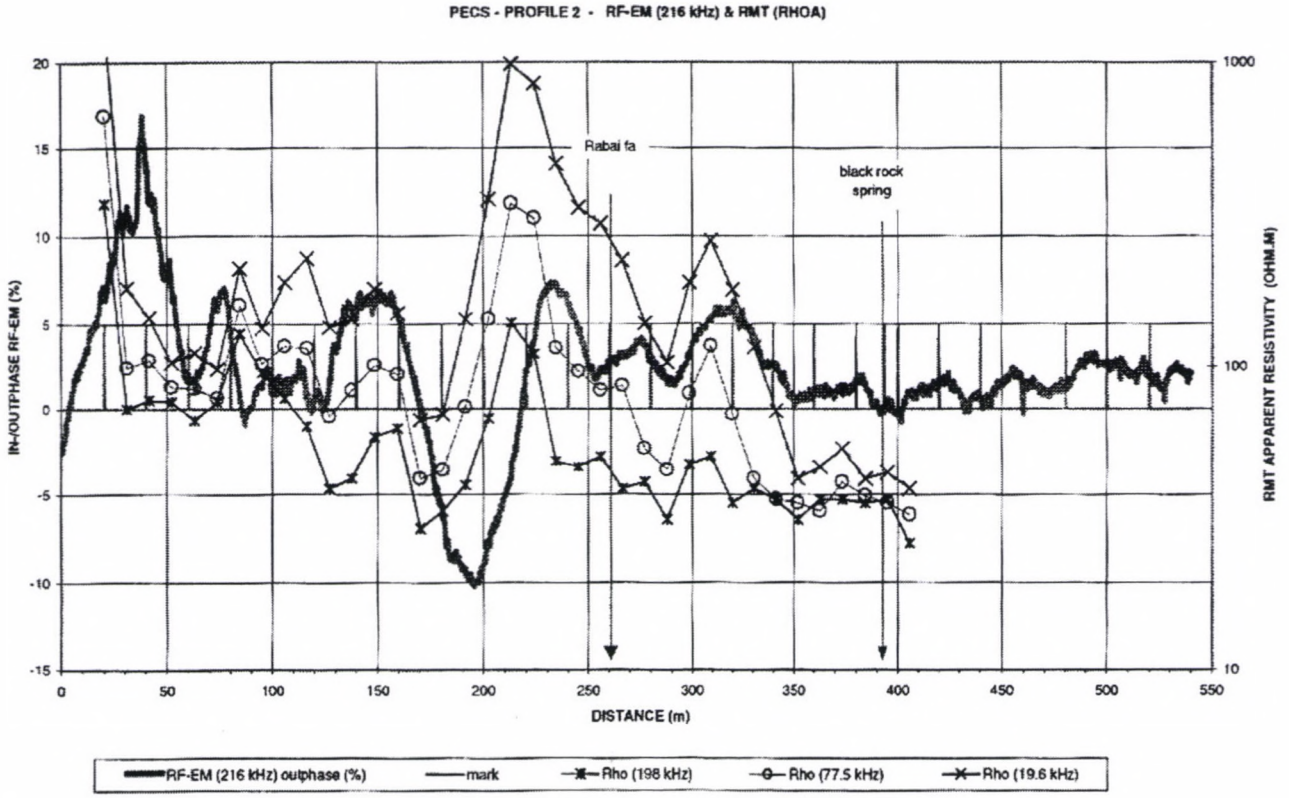


Fig. 7. RF-EM out-of-phase data along profile B at 216 kHz, as compared to radio-magnetotelluric (RMT) apparent resistivities

Acknowledgements

This test measurement was made in frame of a bilateral co-operation between the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences and the Neuchâtel University, supported by the Swiss National Research Foundation under title: Developments in electromagnetic geophysics (1995–1998).

Part of the work was supported by the Hungarian Research Foundation (OTKA), grant No. T0029443.

References

Jackson D D 1992: *Geophys. J. R. astr. Soc.*, 28, 97–109.

Knödel K, Krummel H, Lange G 1997: Geophysik. Handbuch zur Erkundung des Untergrundes von Deponien und Altlasten, Band 3, Springer

Zhdanov M S, Keller G V 1994: The geoelectrical methods in geophysical exploration. Elsevier, Amsterdam, London, New-York, Tokyo

Acta Geod. Geoph. Hung., Vol. 35(4), pp. 415–424 (2000)

HYPOTHESIS OF THE EARTH'S BODY EXPANSION AND GLOBAL PLATE MOTIONS FROM THE POINT OF VIEW OF CONTEMPORARY GEODETIC REFERENCE FRAMES

J KOSTELECKÝ[1,2] and A ZEMAN[2]

[Manuscript received June 3, 1999, revised December 17, 1999]

Based on contemporary observation technologies of space geodesy, annually renewed geocentric reference frames are created. These frames are defined by the co-ordinates of the sites and their time (annual) changes. Using the field of time changes the hypothesis of the Earth's body expansion is checked. Also the model of global tectonic plate movements NUVEL1 is modified and the residual changes at the region of European continent are discussed.

Keywords: Earth's body expansion; geodetic reference frames; plate motion

Introduction

The geocentric terrestrial co-ordinate system will be defined and also the realization of its reference frame will be described. Then repeated determination of the reference frame will be used for consideration on possible global changes of the Earth's body (expanding or contracting). From the results of these considerations the attempt to correct model NUVEL1 will follow and also interpretations of possible relative motions of the European continent will be checked from the point of view of previous considerations. Some authors are interested in present time in definition of a new plate motion model, based on observations see e.g. Argus and Gordon (1991), DeMets et al. (1990), Drewes (1990, 1997a). Drewes (1997b) compares plate rotation vectors of plate kinematic model APKIM from space geodetic observations (realised to 1996) with NNR-NUVEL-1A with very good agreement.

1. Geocentric terrestrial co-ordinate system

For solution of practical tasks of navigation, geodesy, geodynamics, geophysics and other disciplines it is necessary to deal with co-ordinate systems which are in some way "firmly" connected with the Earth's body. So it is necessary to define the terrestrial geocentric system which would enable solution of different types of the tasks (even for the highest level of precision) in global extent.

[1]Research Institute of Geodesy, Topography and Cartography, CZ 250 66 Zdiby 98, email: kost@fsv.cvut.cz

[2]Department of Advanced Geodesy, Faculty of Civil Engineering, CTU Prague, Thakurova 7, CZ 16629 Prague, email: zeman@fsv.cvut.cz

1217-8977/2000/$ 5.00

1.1 Conventional terrestrial co-ordinate system

According to the recommendation of the International Association of Geodesy (IAG) the Conventional Terrestrial Reference System (CTRS) is defined by following way:

1. CTRS is a local co-ordinate system in relativistic sense and is valid for the Earth and its vicinity on the level of relative accuracy 10^{-10}.

2. CTRS is a quasicartesian rotating system derived from the geocentric non-rotating system by space rotation.

3. Coordinate time is geocentric time (TCG). Metric tensor defining the four-dimensional timespace has the components:

$$g_{00} = 1 - 2\frac{V}{c^2} + (4), \quad g_{ij} = \left(1 + 2\frac{V}{c^2}\right)\delta_{ij} + (4), \quad g_{03} = (3)\,, \tag{1}$$

   where δ_{ij} is Kronecker's delta, $i, j = 0, \ldots 3$, c is the velocity of the light and V is gravitational potential of the Earth including the atmosphere, n in brackets indicates the terms of the n-th order.

4. Time, corresponding to the changes of co-ordinates, is identical with TCG.

5. Origin of the system is geocenter, i.e. center of gravity of the Earth including oceans and atmosphere.

6. System has no global residual rotations related to the Earth crust. Cartesian co-ordinate system is connected with IERS (International Earth Rotation Service) reference pole and meridian.

CTRS is realized by

— system of algorithms and constants linked to geometric and physical attributes of the co-ordinate system (see e.g. McCarthy 1992, 1996),

— terrestrial reference frame TRF.

TRF is realized by a finite number of points on the Earth surface with co-ordinates $X(t)$; these co-ordinates are functions of time.

The practical realization depends on definitely used observational techniques and methodology at processing centers. It is necessary to take into consideration that TRF can be — due to its global character — exclusively realized by the techniques of space geodesy. In present there are following techniques:

a) VLBI – Very Long Baseline Interferometry, serves primarily to realize Celestial Reference Frame (relative error of the distances of two stations is 1 ppb),

b) SLR – Satellite Laser Ranging, using pulse laser for observations of the distances between terrestrial station and satellite (rms error of the determination of satellite orbit and also the co-ordinates of the station is 2–3 cm),

Table I. Genesis of the ITRF-XX

Name	Origin	Scale	Orientation
ITRF-0	BTS87	BTS87	BTS87
ITRF-88	ITRF-0	ITRF-0	ITRF-0
ITRF-89	SLR(SSC(CSR))	SLR(SSC(CSR))	NNR to ITRF-88
ITRF-90	SLR(SSC(CSR))	SLR(SSC(CSR))	NNR to ITRF-89
ITRF-91	SLR(SSC(CSR))	SLR(SSC(CSR))	NNR to ITRF-90
ITRF-92	SLR(SSC(CSR))	SLR(SSC(CSR))	NNR to ITRF-91
ITRF-93	special solution		
ITRF-94	SLR + GPS	SLR + GPS + VLBI	NNR to ITRF-92
ITRF-96	SLR + GPS	SLR + GPS + VLBI	NNR to ITRF-94

c) GPS – Global Positioning System, permanent observations of the satellites of the NAVSTAR system using receivers located on the stations connected to the International GPS Service for Geodynamics (IGS) or at regional permanent networks (rms errors of the determination of satellite orbit 5 cm, and relative co-ordinates 1–2 cm),

d) DORIS, based on Doppler observations of the changes of the radial distance between terrestrial station and satellite (rms error of the determination of the distance change is 0.4 mm/s).

ITRS then consists of:

— a set of constants and algorithms — identical with above mentioned CTRS, published at IERS Technical Note (see for instance McCarthy 1992 and 1996),

— a set of the co-ordinates of the stations and their time changes — ITRF (International Terrestrial Reference Frame), also published at IERS Technical Note. In Table I the genesis of co-ordinate frames ITRF are shown, from their first definitions till 1998 when frame ITRF96 was constructed. Abbreviation SLR(SSC(CSR)) means set of co-ordinates, determined from laser observations of the geodynamic satellites at Analytic Centre of Space Research in Texas University, NNR means "no-net-rotation" — see Boucher et al. (1996) and (1998).

1.2 Realization of the reference frame ITRF96

Coordinate frame ITRF96 of the system ITRS96 was created on the base of combination of station co-ordinates by different processing centers at different time epochs. Combination was rigorously carried out according to the least squares method using full variance-covariance matrix. Individual input sets were weighted on the base of mutual comparison (for determination of the weights the residual deviations after Helmert's transformation were used).

ITRF96 is based on combination of:

— selected sets of co-ordinates of individual solutions from individual techniques of space geodesy, delivered to the Central Bureau of the International Earth Rotation Service (IERS) in 1997,

— some older data, published in SINEX format (containing consequently full variance-covariance matrix),

— current data from the GPS and DORIS techniques, specially asked for densification of station network and for better estimations of the time changes of the co-ordinates.

For definition of the ITRF96 4 individual solutions from VLBI, 2 solutions from SLR, 8 solutions coming from processing GPS observations and 3 individual solutions of DORIS were used. Global solution ITRF96 has following characteristics:

— based on complete adjustment of 17 selected geodetic solutions (SSC sets, SSC is Set of Station Coordinates) and of 70 individual solutions from local networks (SINEX files); adjustment was carried out at IERS,

— reference system (origin, scale, orientation, development in time) is identical with ITRF94,

— determined time changes of the co-ordinates are equal for all observational techniques placed on one station.

2. Determination of global motions from geodetic observations

In accordance with present geophysical hypotheses it is possible to divide global movements into horizontal and vertical components. Horizontal component can be expressed by help of the vector of global rotation $\omega(\omega_x, \omega_y, \omega_z)$ for individual tectonic plates. The geophysical model NNR-NUVEL1, which supposes *a priori* zero vertical movements can serve as a starting model. For the components of velocities from model we have simple relations

$$v'_x = \omega_y z - \omega_z y, \quad v'_y = \omega_y x - \omega_x z, \quad v'_z = \omega_x y - \omega_y x\,, \tag{2}$$

where v'_x, v'_y, v'_z are velocities in directions of co-ordinate axes, in our case fulfilling the conditions

$$\begin{vmatrix} v'_x \\ v'_y \\ v'_z \end{vmatrix} = \mathbf{T}'_{x,y,z} \begin{vmatrix} v_S \\ v_E \\ 0 \end{vmatrix} \tag{3}$$

where v_S, v_E are horizontal components in South and East directions respectively. $\mathbf{T}'$ is transposed matrix to the matrix $\mathbf{T}$, matrix $\mathbf{T}$ is transformation matrix between orthogonal components x, y, z and local orthogonal components S, E, R (S is positive to the South, E is positive to the East and R is positive in the direction of outer space). It holds

$$\begin{vmatrix} v_S \\ v_E \\ v_R \end{vmatrix} = \mathbf{T}_{x,y,z} \begin{vmatrix} v_x \\ v_y \\ v_z \end{vmatrix}, \quad \mathbf{T}_{x,y,z} = \begin{vmatrix} \sin\varphi\cos\lambda & \sin\varphi\sin\lambda & -\cos\varphi \\ -\sin\lambda & \cos\lambda & 0 \\ \cos\varphi\cos\lambda & \cos\varphi\sin\lambda & \sin\varphi \end{vmatrix}, \tag{4}$$

where φ, λ are ellipsoidal co-ordinates of the point with orthogonal co-ordinates x, y, z and v_x, v_y, v_z are observed components of velocities.

Reference frame ITRF96 gives us observed components v_x, v_y, v_z, velocity in the radial direction v_R, which is important in connection with the hypothesis of the Earth's body expansion, we shall compute using (4), for the study of horizontal components we use Eqs (4) and (3), respectively.

For further advance it is necessary to take into account following limitations: hypothesis of the Earth's body expansion supposes the phenomenon is "partly continuous" which can be expressed by the function

$$v_R = \frac{1}{l} \sum_{k=1}^{l} \frac{1}{S_k} \int\limits_{dS_k} v_{R,k} dS_k \,, \tag{5}$$

where k is the k-th tectonic plate, l is number of tectonic plates, S_k is the area of that plate, $v_{R,k}$ is the velocity of vertical movement of the element dS_k. In our case we limit ourselves to the point representation, therefore the integral in (5) will be replaced by summation

$$v_R = \frac{1}{l} \sum_{k=1}^{l} \frac{\sum\limits_{i=1}^{n_k} p_i v_{R,i}}{\sum\limits_{i=1}^{n_k} p_i} \tag{6}$$

where n_k is number of the points on the k-th plate, and p_i is weight of the i-th velocity, developed from the real observational data.

Similarly, for global horizontal movement of the k-th tectonic plate, described by the velocity vector ω_k we can write

$$\omega_k = \frac{1}{S_k} \int\limits_{dS_k} (\mathbf{r} \times \mathbf{v})_k dS_k \,, \tag{7}$$

where $\mathbf{r}$ and $\mathbf{v}$ are radiusvector and horizontal velocity of the element dS_k. In our case we know again only the values at the individual points and we have to replace integration by summation, so that for an individual station

$$\omega_k|_i = (\mathbf{r}_i \times \mathbf{v}_i)_k \tag{8}$$

is valid, where i means i-th station on the k-th tectonic plate and the resulting ω_k we obtain by application of the least squares method, considering the accuracy of the i-th velocity, expressed by means of the weight p_i.

Table II. Vertical movements based on ITRF96

Tectonic plate	vertical velocity mm/year	rms error of vertical velocity mm/year	Number of used stations
GLOBAL	**−2.2**	**10.5**	**496**
PCFC (Pacific)	−1.2	14.9	73
AFRC (African)	1.4	14.2	31
ANTA (Antarctic)	5.8	6.2	15
AUST (Australian)	−1.8	10.3	21
CARB (Carib.)	0.3	1.1	2
EURA (Euroasian)	−1.1	6.3	157
INDI (Indian)	2.0	6.9	4
NAZC (Nazca)	1.8	7.1	4
NOAM (North Am.)	−3.3	11.4	162
SOAM (South Am.)	0.7	6.8	24
PHIL (Philip.)	0.6	17.7	2

3. Interpretation of the results obtained by help of the reference frame ITRF96

From above mentioned characteristic of the reference frame ITRF96 it is apparent that this frame is the first where the velocities were determined from real observations. Some stations, however, showed significant dispersion in their results. Therefore the stations which showed rms of the displacements greater than 5 cm/year or if the magnitude of their vector was greater than 10 cm/year, were eliminated from the analysis. These criteria led to the elimination of 30 stations.

3.1 Vertical movements

Computed vertical changes for individual tectonic plates-velocities are in Table II. It is apparent from this Table that statistically significant vertical change was detected for no tectonic plate (we can not reject the zero hypothesis); the same is valid for the whole Earth in global extent. Due to accuracy, characterized by the rms error 10.5 mm for global solution, we can state that if some secular vertical changes (secular expansion/contraction) exist, its absolute value must be less than 1 cm/year with a statistical expectation 67%.

3.2 Horizontal movements, corrections to the model NUVEL1, residual motions

While for the vertical changes until now adequate theories are not supported by observations, in the case of horizontal changes the situation is completely different. Geophysical model responses in some cases fit very well to the motions verified by observations.

We tried to confront observed values with the theoretical ones and carry out prospective corrections to existing model NUVEL1, taking into account the limitations, mentioned in Section 2. For determination of the corrections to the parameters of this model by the least squares method Eq. (8) was used, decomposed to the components. NUVEL1 is a horizontal model and therefore the three decomposed

Table III. Corrections to NUVEL1 based on ITRF96

Plate name PCFC				
NUVEL1 + Corrections = New values and RMS error (*10e-6)				
OMX:	−.001510 +	.000816 =	−.000694	(.000135)
OMY:	.004840 +	.000395 =	.005235	(.000108)
OMZ:	−.009970 +	−.000494 =	−.010464	(.000086)
Plate name AFRC				
NUVEL1 + Corrections = New values and RMS error (*10e-6)				
OMX:	.000891 +	−.000641 =	.000250	(.000082)
OMY:	−.003099 +	.000084 =	−.003015	(.000040)
OMZ:	.003922 +	−.000222 =	.003700	(.000060)
Plate name ANTA				
NUVEL1 + Corrections = New values and RMS error (*10e-6)				
OMX:	−.000821 +	−.000852 =	−.001673	(.000130)
OMY:	−.001701 +	.000086 =	−.001615	(.000149)
OMZ:	.003706 +	−.000544 =	.003162	(.000297)
Plate name AUST				
NUVEL1 + Corrections = New values and RMS error (*10e-6)				
OMX:	.007839 +	−.000805 =	.007034	(.000138)
OMY:	.005124 +	.000611 =	.005735	(.000164)
OMZ:	.006282 +	.000170 =	.006452	(.000123)
Plate name EURA				
NUVEL1 + Corrections = New values and RMS error (*10e-6)				
OMX:	−.000981 +	−.000128 =	−.001109	(.000110)
OMY:	−.002395 +	−.000079 =	−.002474	(.000032)
OMZ:	.003135 +	−.000802 =	.002333	(.000113)
Plate name NOAM				
NUVEL1 + Corrections = New values and RMS error (*10e-6)				
OMX:	.000258 +	−.000039 =	.000219	(.000065)
OMY:	−.003599 +	.000630 =	−.002969	(.000122)
OMZ:	−.000153 +	−.000559 =	−.000712	(.000102)
Plate name SOAM				
NUVEL1 + Corrections = New values and RMS error (*10e-6)				
OMX:	−.001038 +	−.000342 =	−.001380	(.000131)
OMY:	−.001515 +	−.000189 =	−.001704	(.000230)
OMZ:	−.000870 +	.002382 =	.001512	(.000115)

components of the Eq. (8) were completed by the fourth equation (for unknowns $\omega_x, \omega_x, \omega_z$)

$$0 = v_R \quad = \quad \omega_x(y\sin\varphi - z\cos\varphi\sin\lambda) + \omega_y(z\cos\varphi\cos\lambda - x\sin\varphi) + \\ + \omega_z(x\cos\varphi\sin\lambda - y\cos\varphi\cos\lambda)\,. \tag{9}$$

Resulting components of the vector ω of the corrected model NUVEL1 are in Table III.

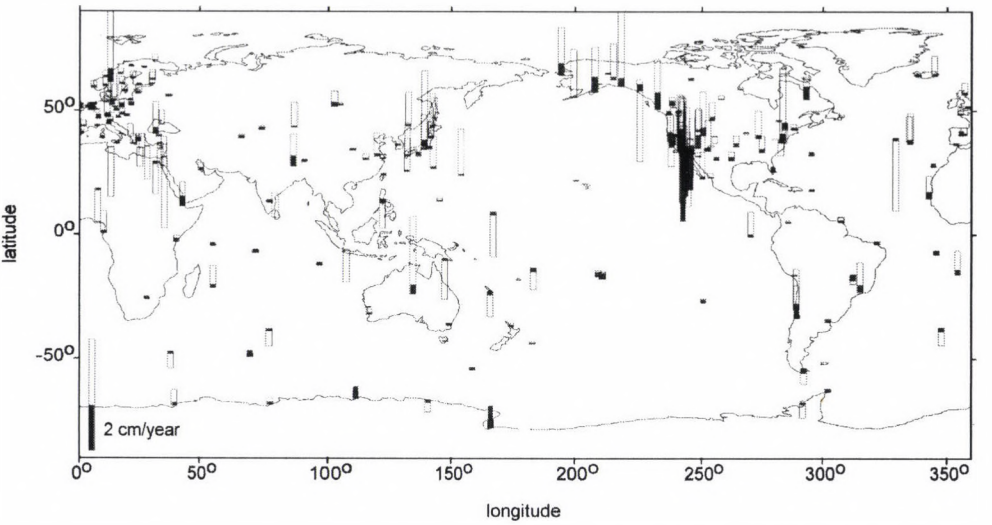


Fig. 1. Residual annual vertical motions (and 2.5 sigma) of the stations, based on ITRF96, after removal of global change

4. Results and conclusions of global solution and the solution limited to European continent

In Fig. 1 the observed global changes of the heights are charted, compared with the hypothesis of spreading body of the Earth. From Fig. 1 it is seen that on the level of 1 cm/year rms the mentioned hypothesis cannot be confirmed. Present accuracy of the model used is not enough to detect even the phenomenon of the Fennoscandian uplift, proved by other methods.

Figure 2 contains the residual horizontal displacements with 2.5 sigma error ellipses for individual tectonic plates. Even in this case it is not possible to detect significantly the residual motion of the tectonic plates from the residual displacements. But there are some individual regions (California, Indonesia) where local motions are statistically detected.

Figure 3 describes residual vertical motions of European stations. No statistically significant motions are detected due to present accuracy of determination.

At Fig. 4 the annual horizontal displacements for the European continent are charted. On the base of the obtained accuracy of the estimation of the changes it is again not enough to find statistically significant residual horizontal displacements.

Acknowledgment

This work was supported by Grant Agency of the Czech Republic, project No. 205/98/1104 and Research Fund No 7 "Innovation of technology in geodesy and cartography" of the Czech Technical University in Prague, Faculty of Civil Engineering.

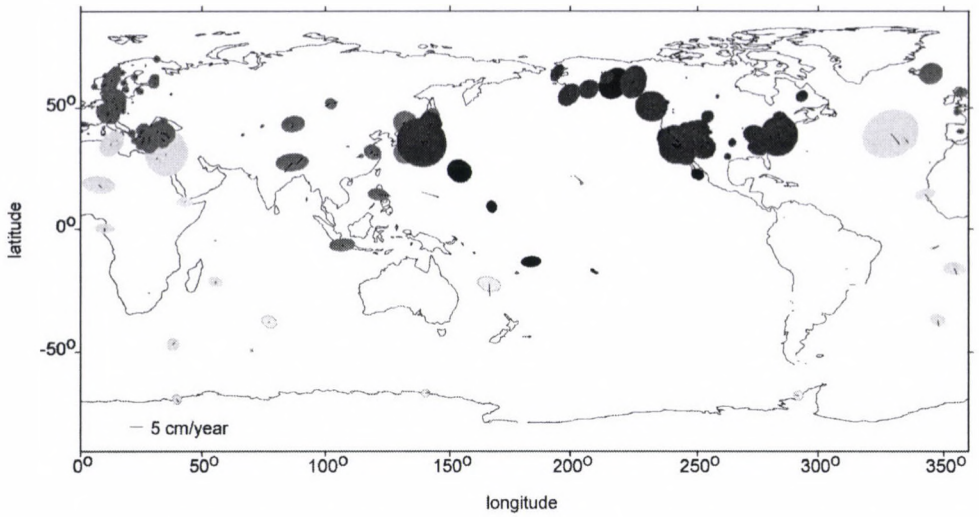


Fig. 2. Residual annual horizontal motions for the stations on individual tectonic plates based on ITRF96 model and 2.5 sigma error ellipses

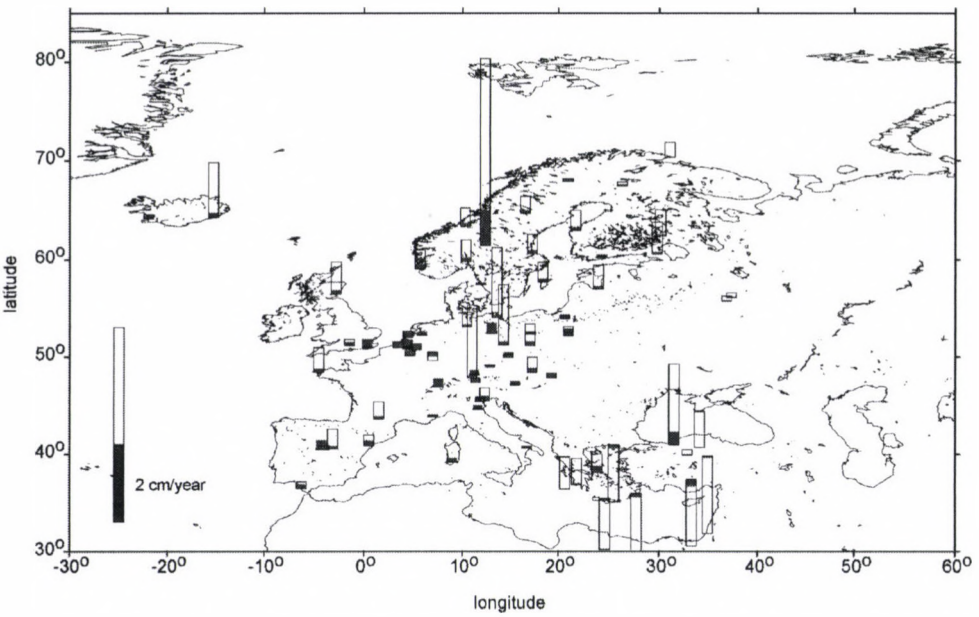


Fig. 3. Residual annual vertical motions (and 2.5 sigma) for European stations based on ITRF96 model

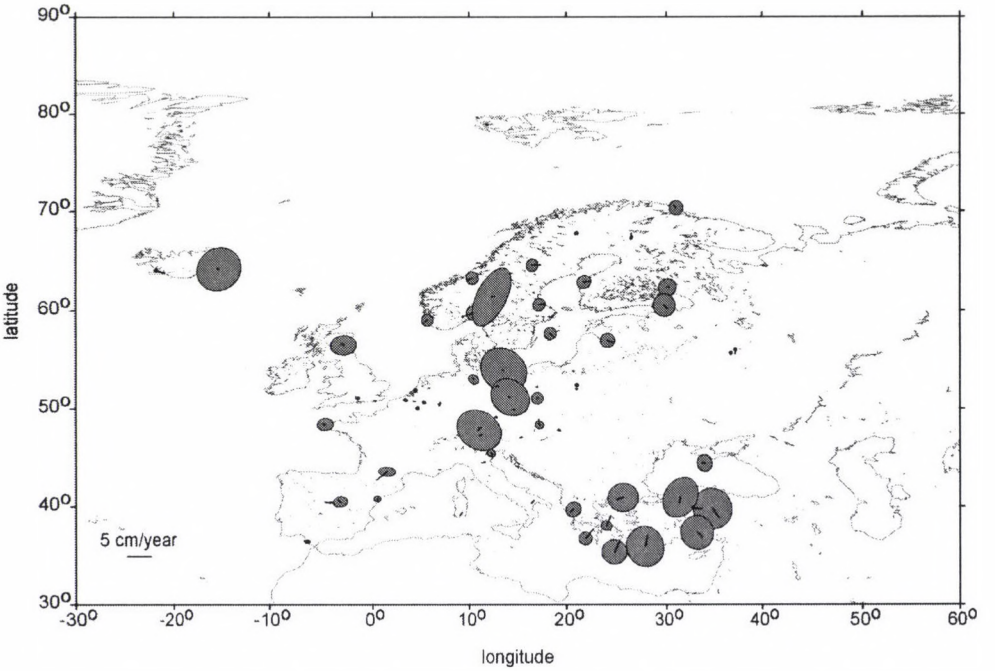


Fig. 4. Residual horizontal annual motions of European stations, based on ITRF96, model after global shift removing and 2.5 sigma error ellipses

References

Argus D F, Gordon R G 1991: *Geoph. Res. Lett.*, 18, 2039–2042.

Boucher C, Altamimi Z 1998: Report on ITRF96 data analysis. IERS 1998 (http://lareg.ensg.ign.fr/ITRF/ITRF96- rep.htm).

Boucher C, Altamimi Z, Feissel M, Sillard P 1996: IERS Technical Note, 20, Observatoire de Paris

DeMets C, Gordon R G, Argus D F, Stein S 1990: *Geoph. J. Int.*, 101, 425–478.

Drewes H 1990: Global plate motion parameters derived from actual space geodetic observations. Springer, IAG Symposia 101, 30–37.

Drewes H 1997a: *Terra Nostra*, 4/97, 4 pp.

Drewes H 1997b: In: Geodesy on the Move, Gravity, Geoid, Geodynamics, and Antarctica. R Forsberg, M Feissel, R Dietrich eds, IAG Symposia, Springer, 119, 377–382.

McCarthy D ed. 1992: IERS Standards 1992, IERS Technical Note 13, Observatoire de Paris

McCarthy D ed. 1996: IERS Conventions 1996, IERS Technical Note 21, Observatoire de Paris

Acta Geod. Geoph. Hung., Vol. 35(4), pp. 425–431 (2000)

INVESTIGATIONS CONCERNING SEMI-INTERQUANTILE RANGES

F STEINER[1] and B HAJAGOS[2]

[Manuscript received January 05, 2000]

The paper shows that the general use of the semi-intersextile range (Q) as error characteristics is more advantageous as that of the semi-interquartile range q. (Denoting the general interquantile range by $q(p)$, $Q = q(1/6)$ and $q = q(1/4)$.) In Section 3 it is shown that three differing characteristics: Q, the A asymptotic scatter of the standard most frequent value determinations based on samples and finally the U uncertainty (the minimum value of the P-norm) approximately coincide for $f_a(x)$-probability distributions of the Cauchy-Gaussian error-type interval.

Keywords: asymptotic scatter; error-type; P-norm; relative asymptotic scatter; semi-interquantile range; semi-intersextile range; standard most frequent value

1. Introduction

$F(x)$ is the distribution function of the actual error distribution and F^{-1} its inverse. The upper p-quantile $q_u(p)$ is the value $F'(1-p)$ and the lower p-quantile $Q_l(p)$ is to be calculated as $F^{-1}(p)$. The half of their differences, i.e.,

$$q(p) = \frac{1}{2}[q_u(p) - q_l(p)] \qquad (0 < p < 0.5) \tag{1}$$

is called semi-interquantile range concerning to p and obviously characterizes in a global sense the error by a single value (the whole characterization can be made by the functions F or F^{-1}).

Two special p-values are used in the practice of statistics: if $p = 0.25$, the simple notation

$$q = q(1/4) \tag{2}$$

is commonly used and this is named as semi-interquartile range. (As equally 50–50% of the data lie inside and outside of the whole interquartile interval, q was named by Bessel properly as "probable error". — It should be also mentioned that the error of radioactive measurements is frequently characterized by q.)

The second, often used choice of p is 1/6; in this case the common simple notation is

$$Q = q(1/6) \tag{3}$$

and this error-characteristics is named semi-intersextile range. In case of Gaussian type of errors (this is supposed nearly always in the classical statistics) $Q \approx \sigma$ holds where σ means the scatter; the difference amounts only 3.3 percent.

[1]Geophysical Department, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary
[2]Institute of Mathematics, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary

1217-8977/2000/$ 5.00

In case of error-distributions symmetric to the origin q and Q can be defined directly by using the F^{-1}-function:

$$q = F^{-1}(1 - 1/4) \tag{4}$$

and

$$Q = F^{-1}(1 - 1/6)\,; \tag{5}$$

obviously the simpler definition

$$q(p) = F^{-1}(1 - p) \tag{6}$$

is in this case also generally usable.

Cramér (1945) gave the formula of the relative asymptotic scatter $A_{q(p)}$ for arbitrary p and for the case if $q(p)$ is determined empirically, i.e., based on a sample:

$$\frac{A_{q(p)}}{q(p)} = \frac{\sqrt{p \cdot (0.5 - p)}}{q(p) \cdot f[q(p)]}\,. \tag{7}$$

In geostatistics (and also in other disciplines using statistical methods) the types of the supermodel $f_a(x)$ can be properly used:

$$f_a(x) = \Gamma(a/2) \cdot \pi^{-1/2} \cdot \Gamma^{-1}[(a-1)/2] \cdot (1 + x^2)^{-a/2}\,. \tag{8}$$

(Equation (8) belongs obviously to the standard cases, i.e., if the location parameter T equals zero and the parameter of scale S is the unity). For six type-parameter values $\underline{a}$ in Fig. 1 are given the $A_{q(p)}/q(p)$-curves versus the p-value ($0 < p < 0.5$).

In some cases the quantity $t = 1/(a-1)$ can be advantageously used instead of the type parameter $\underline{a}$, e.g., the density function of $\underline{t}$ can be written for earth sciences (and also for other disciplines) simply as

$$f(t) = 16 \cdot t \cdot e^{-4t} \tag{9}$$

(see the $f_D(t)$ curve in Fig. 7.4 on page 148 by Steiner 1997).

2. Which semi-interquantile range as error characteristics should be preferred: q or Q?

It is not questionable that the choice must be made according the criterion: the empirical value of q or Q should have minimum statistical fluctuation.

It is sure that if we would *a priori* know the type of the error, according to the curves of Fig. 1 (or similar curves constructed on ground of Eqs (7) and (8) for arbitrary $\underline{a}$ type parameters) the answer is very easy: e.g., in case of the Cauchy-type of error distribution the quantity $\underline{q}$ (the semi-interquartile range) has minimum relative asymptotic scatter (see Fig. 1 or Table II to $t = 1$) and therefore in this concrete case the use of q is preferable instead of Q.

The occurrence probability density of the Cauchy-type ($a = 2$) is, however, much less than that e.g. of the Jeffreys-type ($a = 9$) and even less than the probability

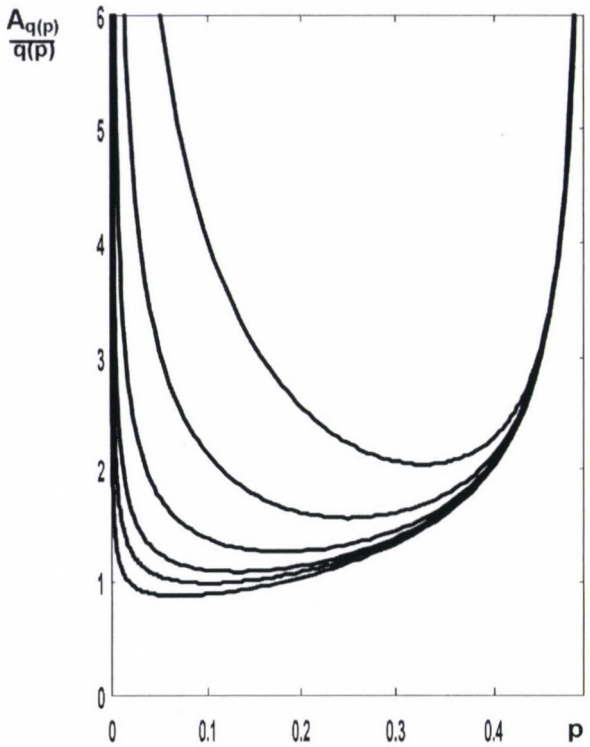


Fig. 1. Relative asymptotic scatters of empirically determined $q(p)$ semi-interquantile ranges vs. $\underline{p}$ for $f_a(x)$ error-distribution types in cases $a = 1.5$; 2; 3; 5; 9 and $a \to \infty$ (i.e., for the Gaussian type); the curves lie in such a way that the lower curve belongs to greater value of $\underline{a}$

density value of the (geo)statistical type ($a = 5$): at this error-distribution type, i.e., at $t = 0.25$ the $f(t)$-function given in Eq. (9) reaches its maximum. But the main problem is that we never (or rarely) exactly know *a priori* the type of the actual error distribution and therefore in our choice all types must play their role but with the weight of their occurrence probability density given in Eq. (9). — This can be carried out in two ways.

2.1 Choice according to the first alternative

We consider as first alternative the weighted integral of the relative asymptotic scatters in function of $\underline{p}$. To avoid complicate notations, the $A_{q(p)}/q(p)$ quotient should be denoted by $\varphi(p,t)$ for the $f_a(x)$-supermodel (see Eq. (8)); as it was defined earlier, $t = 1/(a-1)$.

With this notation our question can be formulated as follows: whether the minimum of the weighted means of the $\varphi(p,t)$-functions for the whole range of $\underline{t}$-s, i.e., the minimum place of the integrals

$$\varphi_w(p) = \int_0^\infty \varphi(p,t) \cdot f(t)dt \tag{10}$$

is nearer to $p = 1/4 = 0.25$ or to $p = 1/6 = 0.1667$? — Table I (and Fig. 2) give the answer: the minimum place is at $p = 0.1875$, consequently it is much nearer to $p = 1/6$ than to $p = 1/4$, the use of Q is therefore more advantageous than that of q, in respect of the "weighted averages" of the relative asymptotic scatters.

2.2 A more sophisticated way of choice as second alternative

For all demonstrated cases figuring in Fig. 1 (and also for arbitrary $\underline{a}$), the relative asymptotic scatters for all $\underline{a}$ type parameters have a single minimum place, denoted in the followings by p_o. After converting $\underline{a}$ to $\underline{t}$ (by using the already many times cited equation $t = 1/(a-1)$), we have a $p_o(t)$ function of the minimum places (calculated according to Eqs (7) and (8)). The numerical values are given in Table II and demonstrated as curves on Figs 3a and 3b. As the optimum $\underline{p}$-value (p_{opt}) can obviously be accepted the $f(t)$-weighted mean of the p_o minimum places, and the result of this single integral is

$$p_{opt} = \int_0^\infty p_o(t) \cdot f(t)dt = 0.1707\,. \tag{11}$$

This value is near to 0.1667 which latter $\underline{p}$-value defines the $\underline{Q}$ semi-intersextile range. The value of p_{opt} is far from 0.25 which defines the $\underline{q}$ semi-interquartile range. The choice of Q as generally usable semi-interquantile range is evident.

Table I

p	$\varphi_w(p)$
0.14	1.3338
0.1428	1.3299
0.15	1.3206
0.16	1.3103
0.1667	1.3054
0.17	1.3035
0.18	1.2998
0.185	1.29904
0.1875	1.29893
0.19	1.29901
0.195	1.29967
0.2	1.3010
0.21	1.3057
0.22	1.3128
0.23	1.3228
0.24	1.3354
0.25	1.3504

Table II

t	p_o
0.00001	0.06915
0.0001	0.06918
0.001	0.06942
0.01	0.07177
0.02	0.07437
0.04	0.07955
0.1	0.09479
0.15	0.10706
0.2	0.11889
0.25	0.13024
0.3	0.14110
0.4	0.16137
0.5	0.17981
0.6	0.19656
0.7	0.21180
0.8	0.22569
0.9	0.23838
1	0.25000
1.2	0.27051
1.5	0.29581
1.7	0.30986
2	0.32769
2.5	0.35105
3	0.36887
4	0.39424
5	0.41140
6	0.42378
10	0.45113
15	0.46627
20	0.47424
100	0.49462
1000	0.49946

3. Approximate but interesting coincidences of three totally different quantities: Q, A and U for the Cauchy-Gaussian error-type interval

The reader should regard this Subsection only as a preliminary comment. Deeper investigation of this curious approximate coincidence lies outside of the scope of this paper but the authors feel as their obligation to mention these interesting interconnections between Q (from that just turned out in Subsection 2.2 to have optimum properties) — and quite other two characteristics: the asymptotic scatter A of the most frequent value M (calculated according to Eqs (I-20) and (I-24) on pages 30 and 31 by Steiner 1997 using $k = 2$), and the U uncertainty, i.e., the minimum value of the P-norm of the deviations X_i (see in the fifth row of the Table P-2 on page 20 by Steiner 1997).

Table III shows the Q, A and U values for some error-models of the Cauchy-Gaussian type-interval, based upon Table 5.3 on page 186 (A-column for $k = 2$) by

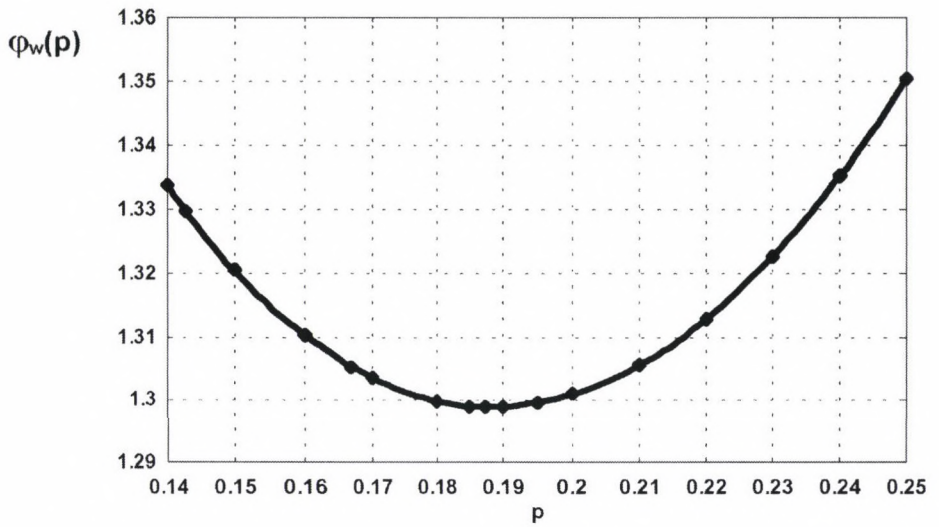


Fig. 2. The function $\varphi_w(p)$ vs. p (see Eq. (10) in Subsection 2.1). The numerical values are given in Table I

Table III

a	Q	$A(k=2)$	U	A/Q	U/Q
2	1.7320	1.5000	1.5000	0.8660	0.8660
3	0.9236	0.8944	0.8605	0.9684	0.9621
4	0.6642	0.7080	0.6628	1.0659	0.9979
5	0.5496	0.5917	0.5585	1.0766	1.0161
6	0.4787	0.5173	0.4916	1.0806	1.0269
10	0.3407	0.3694	0.3564	1.0842	1.0459
40	0.1568	0.1699	0.1669	1.0835	1.0641
∞ (Gaussian)	0.9674	1.0466	1.0327	1.0819	1.0675

Steiner (1990) and upon Table 3.4 on page 106 in the same book (see the U and Q values). The last two columns of Table III show that the quotients A/Q and U/Q equally are near to the unity: the greatest departures are round 13% in the negative direction at the Cauchy-type for both quotients A/Q and U/Q, in positive direction the greatest departures are round 8% in case of A/Q and 7% in case of U/Q, both values concern to the Gaussian type. It seems to be more comfortable (in this type-interval) as an acceptable approximation to determine the simple and obvious Q-value instead of the more complicated A or U.

Acknowledgement

The present research work was supported by the Hungarian Ministry of Education (projekt number 0914/1997).

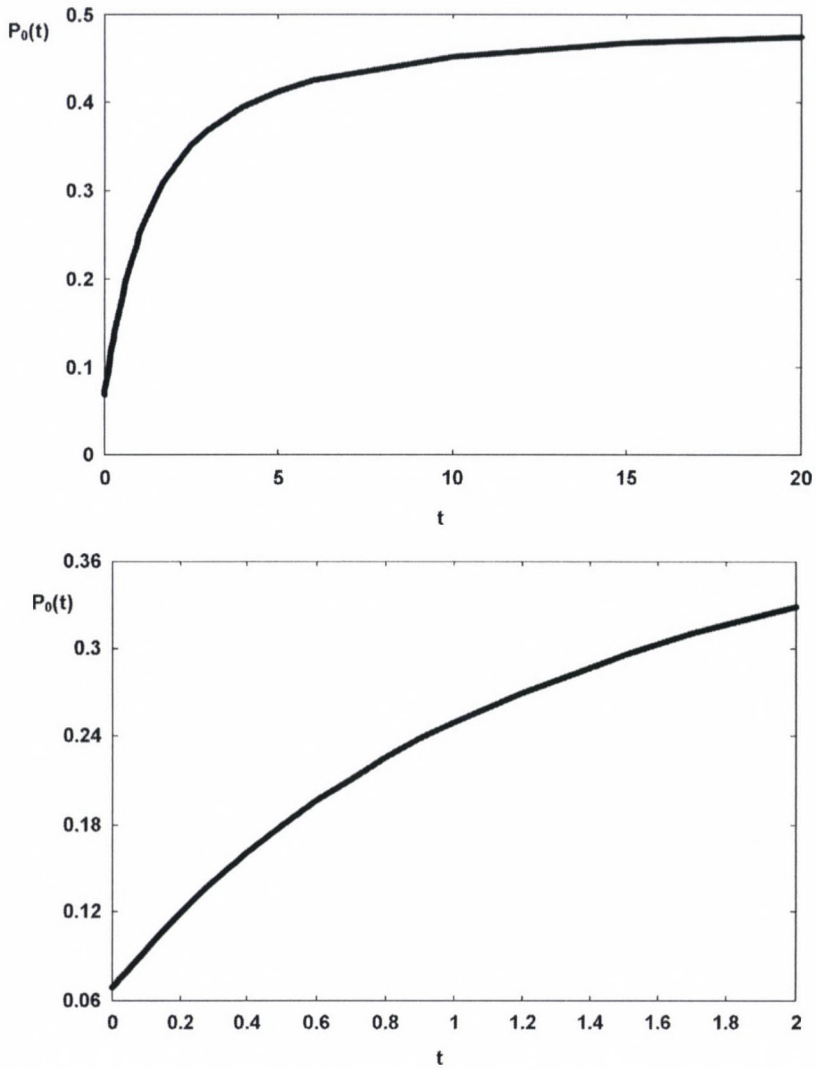


Fig. 3. The $p_o(t)$ function of minimum places (see Fig. 1) vs.the type parameter t. The numerical values are given in Table II

References

Cramér H 1945: Mathematical Methods of Statistics. Almqvist and Wiksells, Uppsala
Steiner F ed. 1997: Optimum Methods in Statistics. Akadémiai Kiadó, Budapest
Steiner F 1990: Introduction to Geostatistics (in Hungarian). Tankönyvkiadó, Budapest

Acta Geod. Geoph. Hung., Vol. 35(4), pp. 433–451 (2000)

ACCURACY OF GPS PHASE AND CODE OBSERVATIONS IN PRACTICE

P Bona[1]

[Manuscript received April 22, 1999]

In order to examine the accuracy of GPS observations, both precision and biases have to be taken into account. The precision of a GPS observation reflects its stochastic variability and is described by the standard deviation. Biases are non-random effects in the GPS observations (like for instance multipath).

In this article these two factors will be examined for seven GPS receivers. In order to examine precision and multipath, an experiment with both zero and short baselines was carried out, which will be briefly discussed. As results, time series of residuals will be presented, on the one hand to give a picture of the precision capabilities of GPS receivers at present, and on the other hand to examine the impact of non-modeled effects like multipath on GPS observations in practice.

Keywords: GPS code observations; GPS phase observations; noise; precision

1. Introduction

Least squares estimation is the common processing method in GPS positioning applications. In order to get a best linear unbiased estimation (BLUE) of the parameters, we need to set up both the functional model and the stochastic model properly.

The precision reflects the noise of the observations and is taken into account in the *a priori* variance-covariance matrix of the observations. In commonly used stochastic models the phase observations have a standard deviation of some millimeters and the code observations of some decimeters.

Biases are those unpredictable effects in the observations, which are not captured by the present model. Neglecting these biases leads to false adjustment results. Among other errors, such as receiver and satellite clock error, orbit error and atmospheric delays, multipath plays a crucial role in high precision applications. Since there is generally no explicit functional model for multipath, it is of high importance to study the effects of it.

Multipath is the phenomenon when the signal is not only coming directly from the satellite but also reflected from objects around the antenna. The receiver then records a mix of the direct and the reflected signals. This reflected signal is usually weaker than the direct signal. Multipath can be significant in those areas where the measurements are performed near reflecting objects such as in cities and industrial areas.

[1]Department of Mathematical Geodesy and Positioning, Delft University of Technology, The Netherlands, e-mail: p.bona@geo.tudelft.nl

1217-8977/2000/$ 5.00

Multipath affects both code and phase measurements. According to (Langley 1998 and Braasch 1996) the theoretical maximum on a P-code observation is 15 meter and 150 meter for the C/A-code. Multipath on code observations usually shows a sinusoidal oscillation of periods of typically 6 to 10 minutes. Practical examinations in [ibid] show somewhat better results: some meters for the P-code and larger for the C/A-code. Theoretically, the multipath effect can amount to about 5 cm on phase observations. Since receiving a reflection of the signal arriving from a low satellite is more likely than for a high satellite, the effect of multipath is usually larger for low elevation satellites.

Because of its complexity, there is no general and explicit model for multipath delays. Instead, several methods are used to avoid multipath or to minimize its effects:

1. Measuring during a considerably long time. By averaging, the effect of multipath can be reduced.

2. For high precision static applications, if the measurements are performed over more than 24 hours, the satellites return to the same positions and multipath will be likely to repeat as well. By differencing the observations over 24 hours time, the multipath will be largely eliminated.

3. Careful selection of the site, although in practice this is not always possible.

4. Properly designed antennas (choke-ring) to eliminate the signals arriving at negative elevation angles.

5. Receivers can apply sophisticated signal processing to reduce the multipath. In the following, the experiment will be described that was carried out for examining the measurement noise and the effect of multipath in practice.

2. The experiment

Two measurement configurations were applied: the zero baseline for examining the pure receiver's noise and the short baseline for analyzing the precision in practice and studying the effect of multipath (and other possible disturbances) on the observations. In the zero baseline configuration, two receivers of the same type are connected to one antenna. In the single difference model for this configuration with fixed co-ordinates and ambiguities, common error sources such as the atmosphere (ionosphere, troposphere), satellite orbits and clocks (satellite position and clock bias) and multipath, are absent.

In the short baseline configuration the receivers are each connected to one antenna with a short distance between them. While the atmospheric errors and satellite-dependent errors are practically absent, multipath will be present.

In the experiment seven GPS receiver pairs were used. The first part was on April 9th and 10th (day of year: 99 and 100; GPS week: 1004), and the second part on July 23rd and 24th (day of year: 204 and 205; GPS week: 1019), 1999. The measurements took place some kilometers south of Delft, in a flat meadow

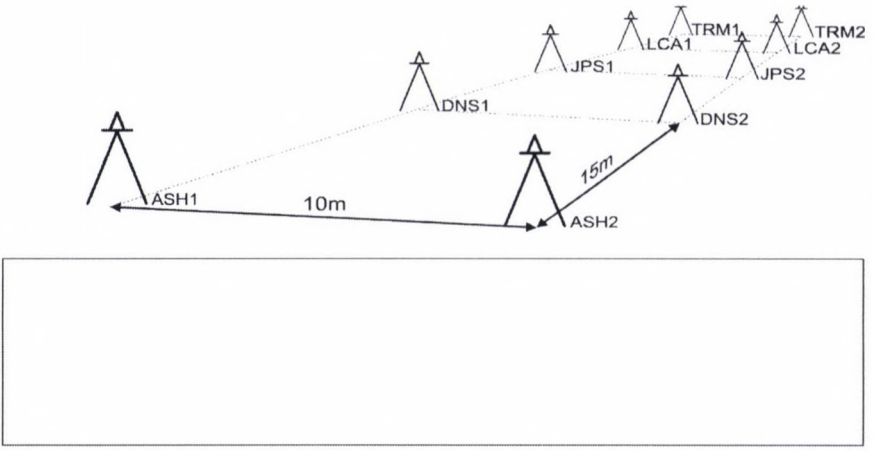


Fig. 1. The short baseline configuration, baselines of 10 meter, separated by 15 meters

Table I. The used receivers and antennas

Receiver	Antenna	Observables (Rinex notation)
Ashtech Z-XII	Ashtech chokering	L1, L2, C1, P1, P2
Dassault-Sercel NP Scorpio 6000	Dassault-Sercel	L1, L2, C1, P1, P2, D1, D2
Javad Positioning Systems Legacy	RegAnt-2 JPS dual-depth chokering	L1, L2, C1, P1, P2, D1, D2
Leica system 500 (SR 530)	Leica AT502	L1, L2, C1, P2
Trimble 4000 SSi	Trimble Compact L1/L2 antenna	L1, L2, C1, P2
Leica CRS 1000	Leica AT504 chokering	L1, L2, C1, P2
Trimble 4700/MS 750	Trimble geodetic L1/L2	L1, L2, C1, P2, D1

with no obstructions or features within several hundred meters. The time of the measurement was chosen so as to observe the same satellite configuration and as many satellites as possible. In April this was between 0800–0900 UTC, in July it was between 0100–0200 UTC.

The short baseline configuration in April is shown in Fig. 1. The receivers were placed in a grid, with 15 meter between the different receiver-pairs and 10 meter between the same receivers. In Fig. 2 the visible satellites during the measurement period are displayed. Ten satellites were available for a one-hour period above 10 degrees elevation. The used receivers' list can be seen in Table I. There were no two Trimble MS750 receivers available, so one Trimble MS750 and one Trimble 4700 were used as a pair. As stated by the Trimble representative, they are similar receivers.

During the zero baseline measurement, the antennas were on the ground, to avoid any effect of multipath. We used WR Inc. splitters for all antennas. The JPS receivers did not function properly with this splitter: for low satellites many

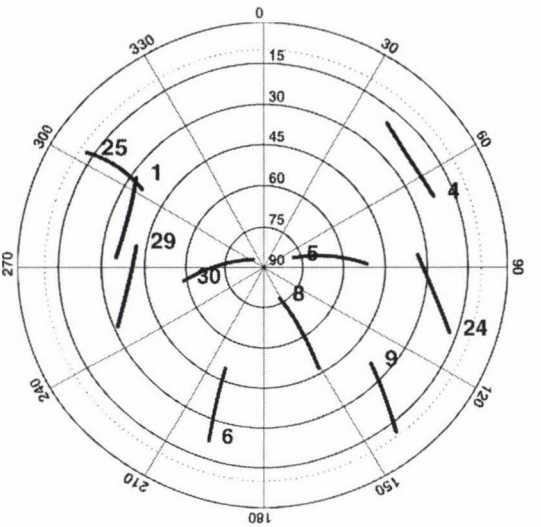


Fig. 2. Satellite visibility during one hour session

observations were missing. By eliminating the low satellites from the processing we were able to obtain meaningful residuals. For the short baseline measurements, the antennas were some 0.70-meter above the ground.

Although some receivers were able to collect other types of observables (Doppler phase) and observe GLONASS satellites as well, we studied only the usual GPS observation types based on Rinex files (Gurtner 1994): the phase measurements on both frequencies (L1, L2 in Rinex notation) and the code measurements on both frequencies (C1, P1 and P2), collected under Anti-Spoofing.

We set the elevation cut-off of the receivers to 10 degrees and the sampling interval was one second so we aimed to observe 3601 epochs in the one-hour measurement period.

3. The processing method

In the single-difference relative positioning model over short distances the atmospheric effects, satellite orbit and clock errors have no effect on the baseline vector between the two antennas. In the short baseline case multipath can be present. If we constrain the ambiguities and the co-ordinates, we get the following functional model:

$$E\left\{\begin{array}{c}\underline{P}_{12}^{\prime 1}\\ \underline{P}_{12}^{\prime 2}\\ \vdots\\ \underline{P}_{12}^{\prime m}\end{array}\right\}=\left(\begin{array}{c}1\\1\\\vdots\\1\end{array}\right)(c\delta_{12}(t))\ , \tag{1}$$

where $P_{12}^{\prime i}$ is the single difference in meters, with known baseline coordinates and resolved phase ambiguities, for one epoch, one observation type and m satellites. For

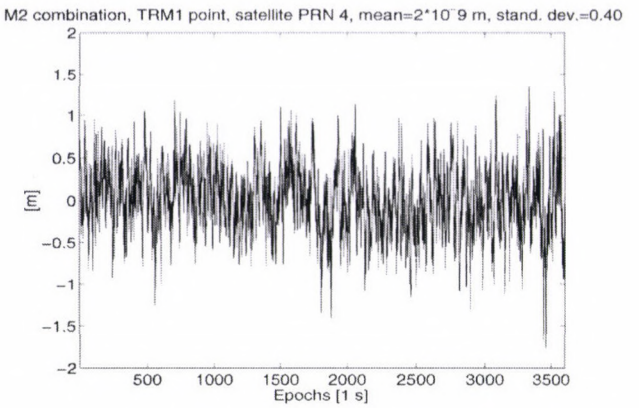


Fig. 3. The M2 combination on TRM1

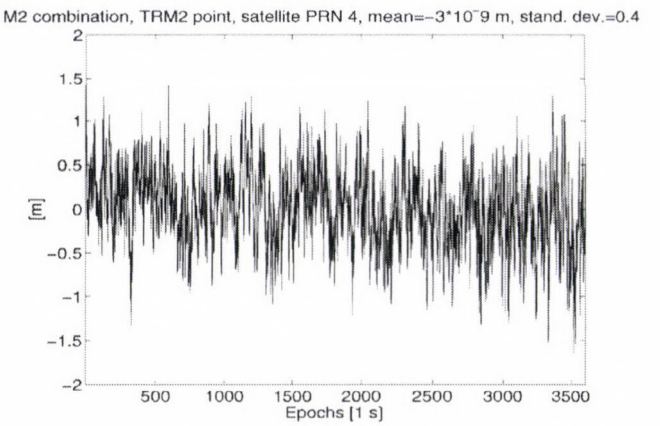


Fig. 4. The M2 combination on TRM2

each epoch there is one parameter to solve for: $\delta_{12}(t)$ the receiver clock difference, c is the speed of light. For the processing of the baselines, the GPSveQ program was used (de Jonge 1998) which carries out the adjustment in an equivalent undifferenced implementation.

4. Results

Since the effect of multipath is usually larger on observations for a low satellite than for a high satellite, in the following the residuals for PRN 4 (elevation angle: 20°) will be presented both for the zero baseline and for the short baseline. Because two receivers of the same type (likely with similar but independent stochastic properties) and in similar surroundings (so with similar but not identical multipath effects) were used in a pair, the undifferenced residuals can be presented.

To demonstrate that the receiver noise and possibly multipath is similar at both ends of the short baseline (in our experiment), the M2 combination of the observations is presented in Fig. 3 and in Fig. 4 for the Trimble 4000 SSi. This

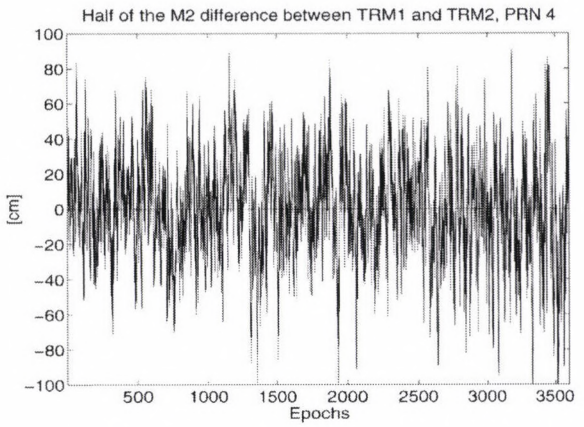


Fig. 5. The half of the M2 differences

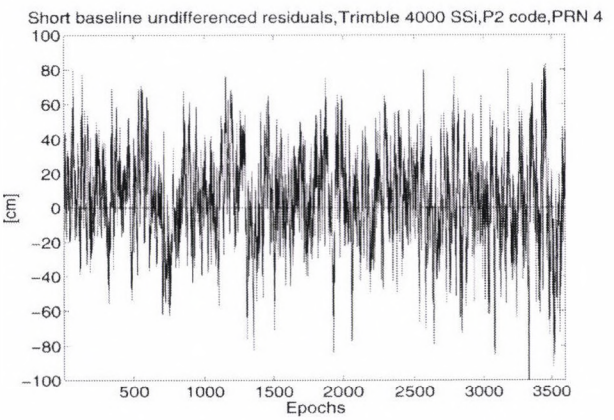


Fig. 6. Undifferenced P2 residuals

combination is based on the observations of a single receiver for a single satellite and it reads:

$$M2 = P2 - 5.092 \cdot L1 + 4.092 \cdot L2\,. \tag{2}$$

This combination reflects next to the P2 code noise the effect of multipath on the P2 observation. The two time series are clearly not identical but have similar characteristics (the mean is about zero and they have similar standard deviations). The difference (divided by two) of the two M2-time series is given in Fig. 5. It is worth comparing the half of the difference of the M2 combination with the corresponding undifferenced residuals on the short baseline, Fig. 6, since they reflect the same phenomenon, but are computed along different ways. The two figures are almost identical which indicates that the two receivers obviously contribute equally to the residuals of model (1). The residuals computed from the single difference solution for the short baseline include the (single station) noise and multipath effects at both ends.

4.1 Time series

In the following the time series of the residuals will be presented for satellite PRN 4 (Figs 7–36). Because of the lack of place the residuals of Dassault-Sercel and JPS receivers will not be presented. These residuals represent the noise of the receiver and possibly multipath expressed in terms of undifferenced observations. The difference between the zero and short baseline time series can be caused by the noise of the antenna next to the receiver noise and multipath which is usually absent on a zero baseline. The time series of L2 phase residuals are very similar to those of L1 phase residuals, so the L2 will not be presented. For similar reasons the P1 code observations also will not be shown next to the P2. The residuals for the zero baseline will be presented on the left side, for the short baseline it will be on the right side. The scale for phase observations is –4 mm to 4 mm, for code it is –100 cm to 100 cm. Where the scale is different, it will be indicated.

Ashtech

From the zero baseline results we see that the noise on L1 is less than 1 mm. For the short baseline the effects of multipath are clearly visible and the residuals increased to several millimeters. The noise of C1 code is some centimeters, while that of P2 code is some decimeters. The multipath has increased the residuals in both cases to about 60-cm.

Leica 500

The scale for L1 is –8 mm to 8 mm and for C1 and P2 code the scale is –20 cm to 20 cm. The noise of L1 phase is less than 1 mm. The residuals for the short baseline are some millimeters and a time varying bias can be clearly seen. The residuals of C1 and P2 code for the zero baseline are exceptionally small and less than 5 cm but a systematic part can be seen in the time series. These systematic parts are probably caused by an internal receiver filtering for reducing the noise and possibly the effect of multipath on the code observations. The effects of this filter can be seen in the time series of the residuals for the short baseline. The noise is very small but the residuals clearly do not have a zero mean.

Trimble 4000 SSi

The noise of L1 phase is less than 1 mm. The residuals of the short baseline measurement have increased to a few millimeters. The noise of C1 code is about 15 cm and probably because of multipath the residuals increased to about 25 cm. The noise on P2 code is about 20 cm and in the short baseline measurement the residuals have grown larger than 60 cm.

Leica CRS 1000

Because of the incidentally missing epochs only 25 minutes of data for the zero baseline and only 45 minutes for the short baseline were available for processing.

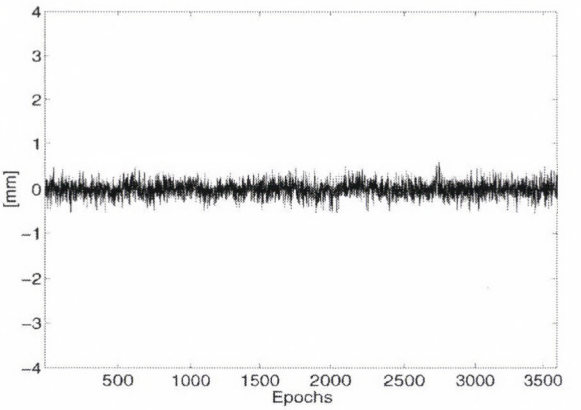


Fig. 7. Zero baseline undifferenced residuals for Ashtech, L1 phase, PRN 4

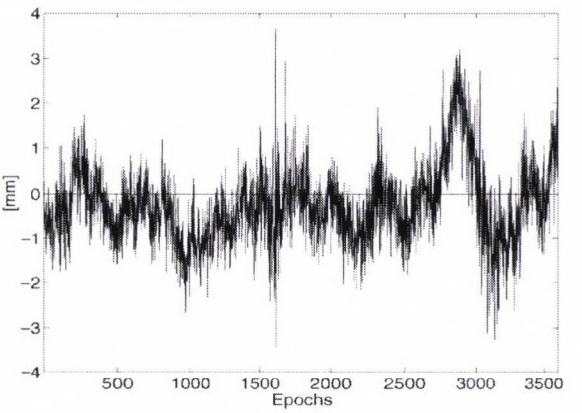


Fig. 8. Short baseline undifferenced residuals for Ashtech, L1 phase, PRN 4

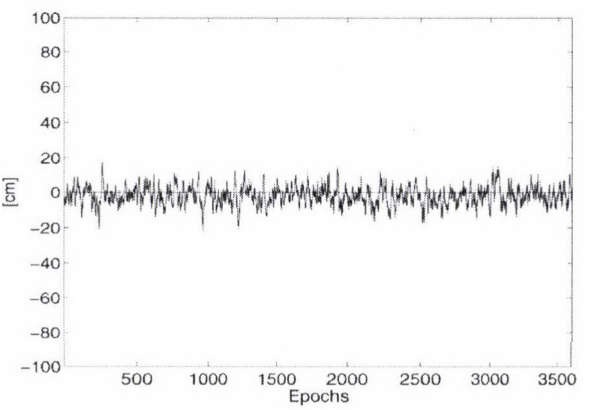


Fig. 9. Zero baseline undifferenced residuals for Ashtech, C1 code, PRN 4

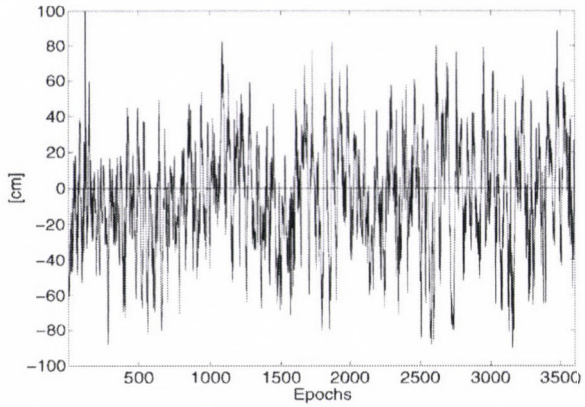


Fig. 10. Short baseline undifferenced residuals for Ashtech, C1 code, PRN 4

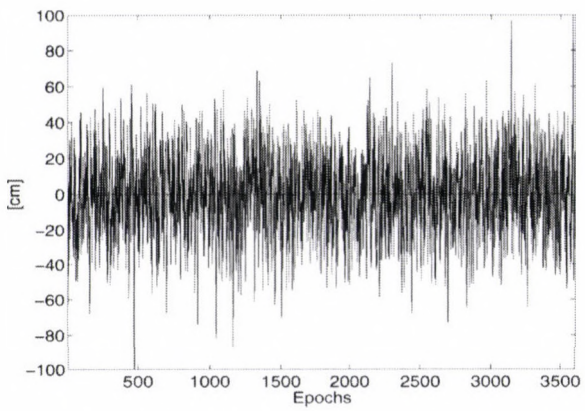


Fig. 11. Zero baseline undifferenced residuals for Ashtech, P2 code, PRN 4

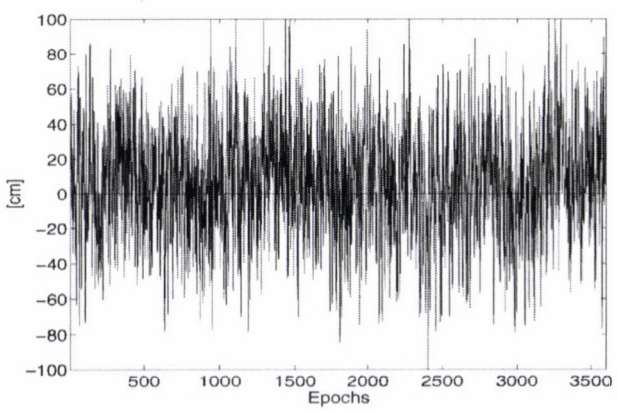


Fig. 12. Short baseline undifferenced residuals for Ashtech, P2 code, PRN 4

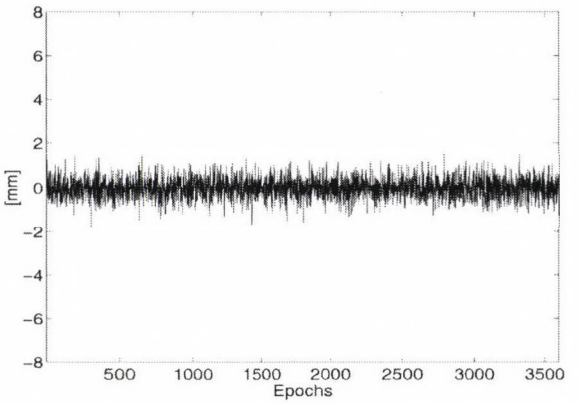


Fig. 13. Zero baseline undifferenced residuals for Leica 500, L1 phase, PRN 4

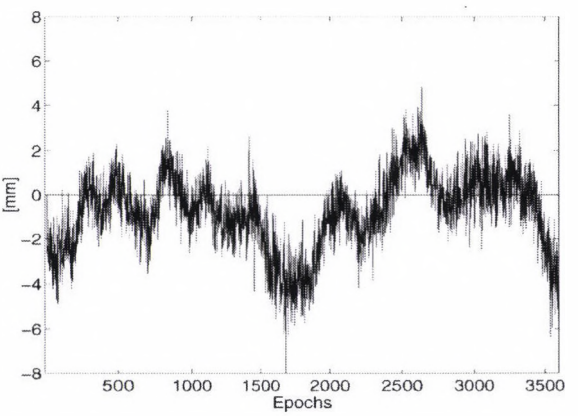


Fig. 14. Short baseline undifferenced residuals for Leica 500, L1 phase, PRN 4

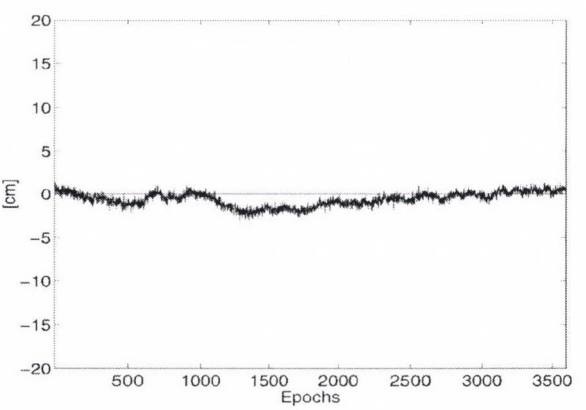


Fig. 15. Zero baseline undifferenced residuals for Leica 500, C1 code, PRN 4

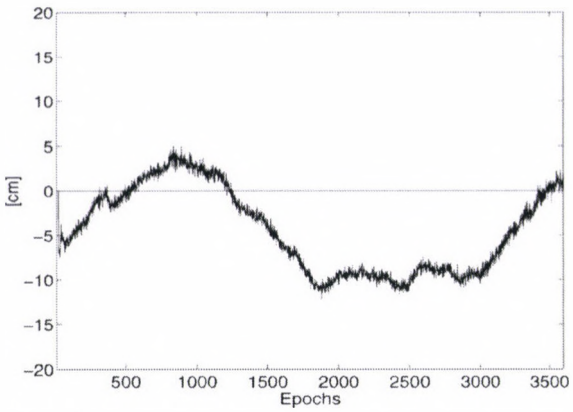


Fig. 16. Short baseline undifferenced residuals for Leica 500, C1 code, PRN 4

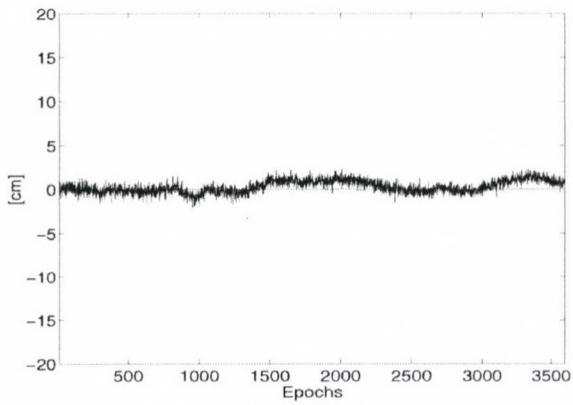


Fig. 17. Zero baseline undifferenced residuals for Leica 500, P2 code, PRN 4

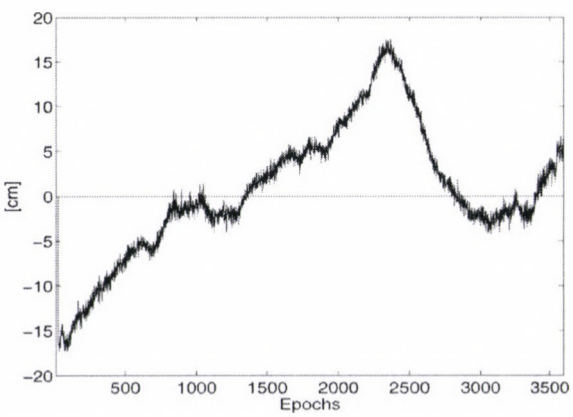


Fig. 18. Short baseline undifferenced residuals for Leica 500, P2 code, PRN 4

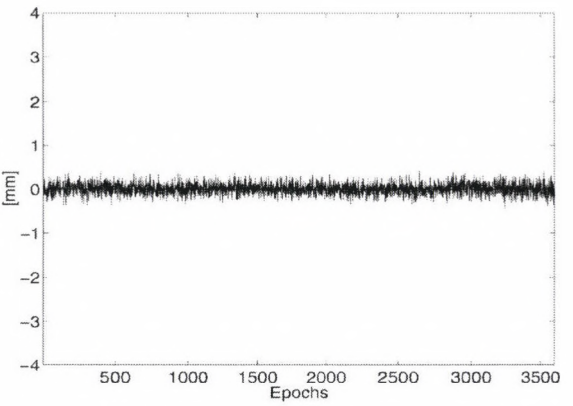


Fig. 19. Zero baseline undifferenced residuals for Trimble 4000 SSi, L1 phase, PRN 4

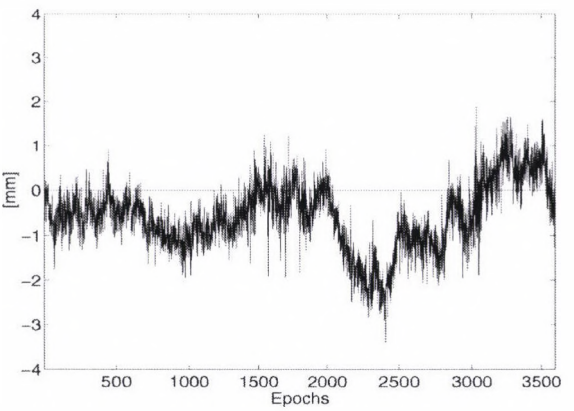


Fig. 20. Short baseline undifferenced residuals for Trimble 4000 SSi, L1 phase, PRN 4

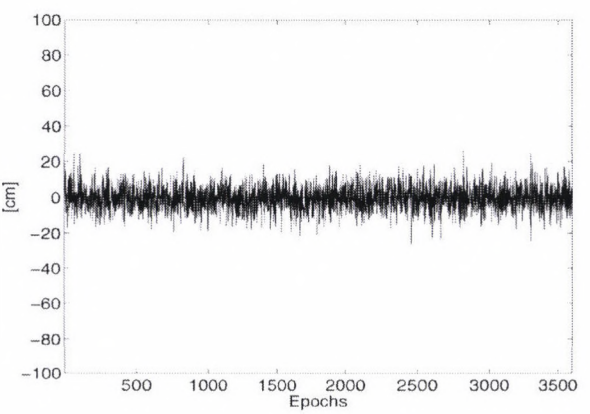


Fig. 21. Zero baseline undifferenced residuals for Trimble 4000 SSi, C1 code, PRN 4

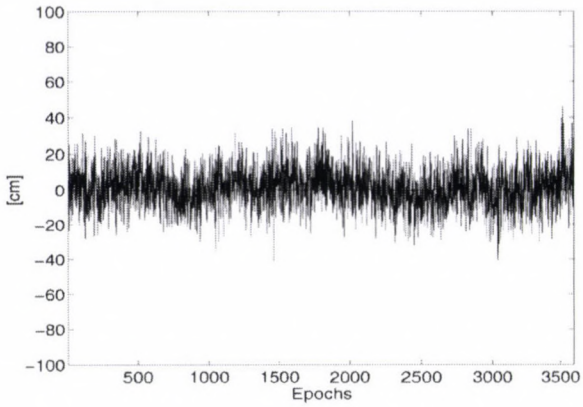


Fig. 22. Short baseline undifferenced residuals for Trimble 4000 SSi, C1 code, PRN 4

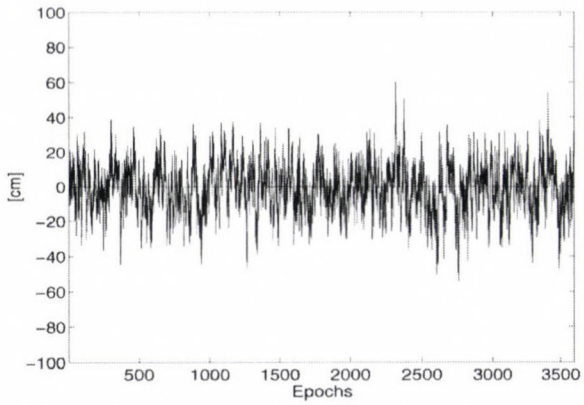


Fig. 23. Zero baseline undifferenced residuals for Trimble 4000 SSi, P2 code, PRN 4

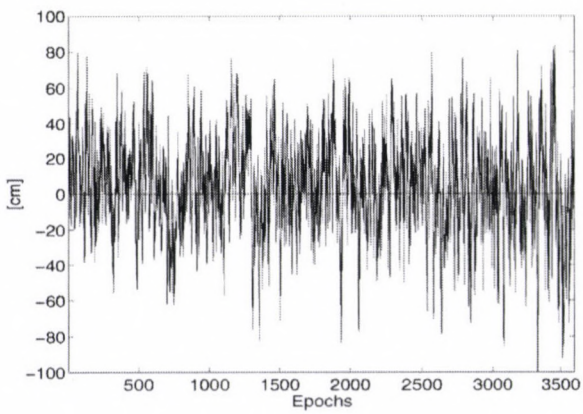


Fig. 24. Short baseline undifferenced residuals for Trimble 4000 SSi, P2 code, PRN 4

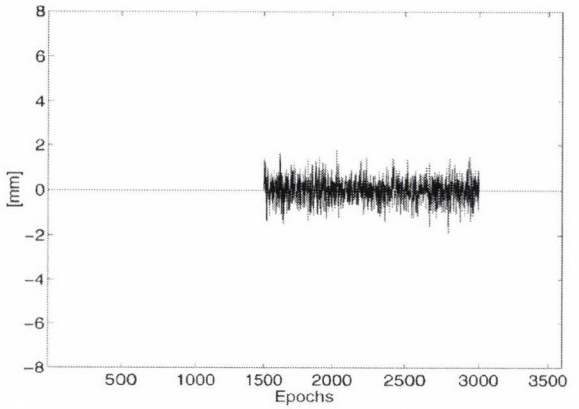


Fig. 25. Zero baseline undifferenced residuals for Leica CRS 1000, L1 phase, PRN 4

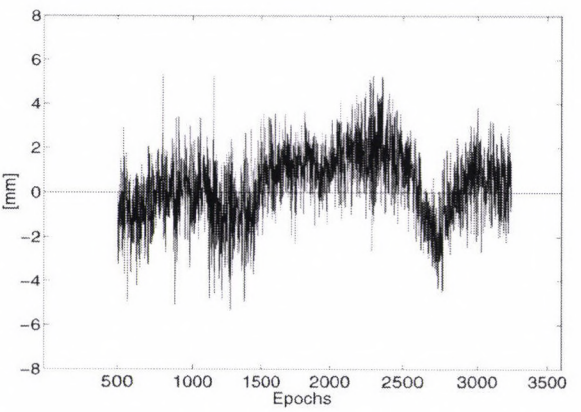


Fig. 26. Short baseline undifferenced residuals for Leica CRS 1000, L1 phase, PRN 4

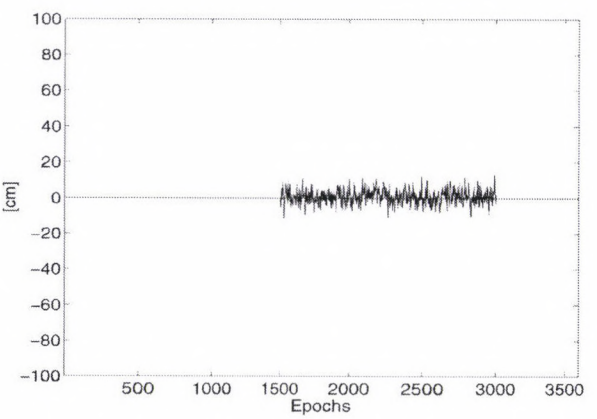


Fig. 27. Zero baseline undifferenced residuals for Leica CRS 1000, C1 code, PRN 4

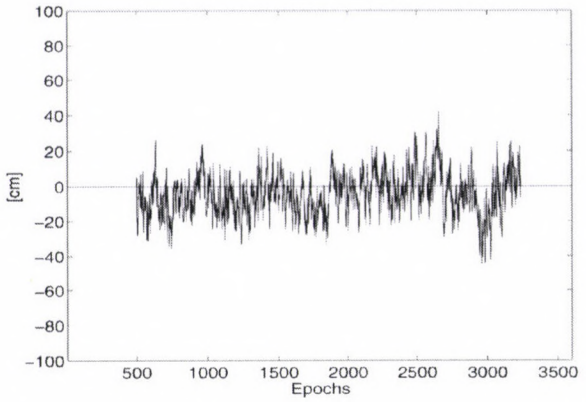


Fig. 28. Short baseline undifferenced residuals for Leica CRS 1000, C1 code, PRN 4

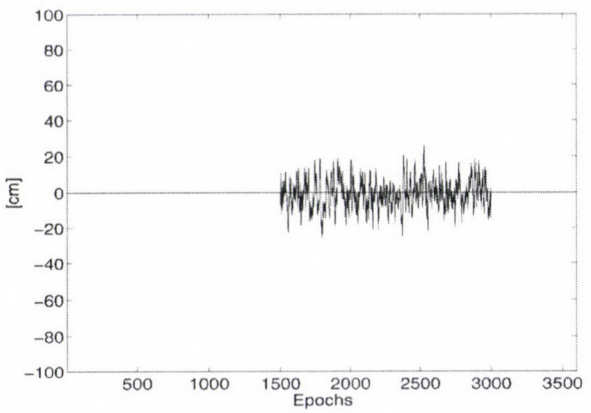


Fig. 29. Zero baseline undifferenced residuals for Leica CRS 1000, P2 code, PRN 4

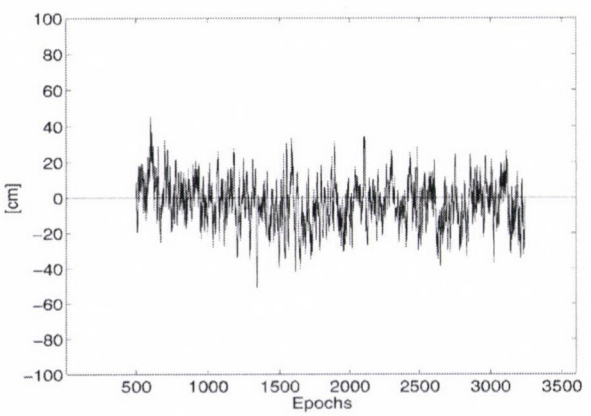


Fig. 30. Short baseline undifferenced residuals for Leica CRS 1000, P2 code, PRN 4

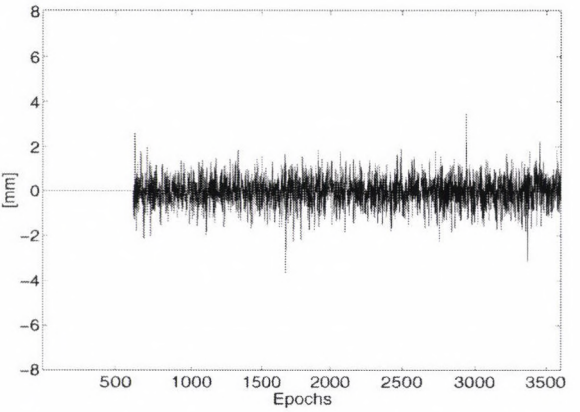


Fig. 31. Zero baseline undifferenced residuals for Trimble 4700/MS 750, L1 phase, PRN 4

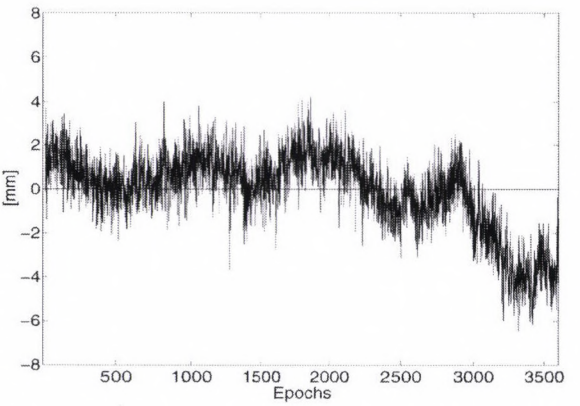


Fig. 32. Short baseline undifferenced residuals for Trimble 4700/MS 750, L1 phase, PRN 4

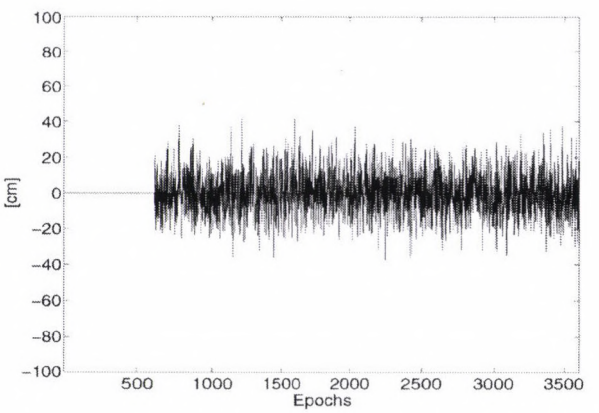


Fig. 33. Zero baseline undifferenced residuals for Trimble 4700/MS 750, C1 code, PRN 4

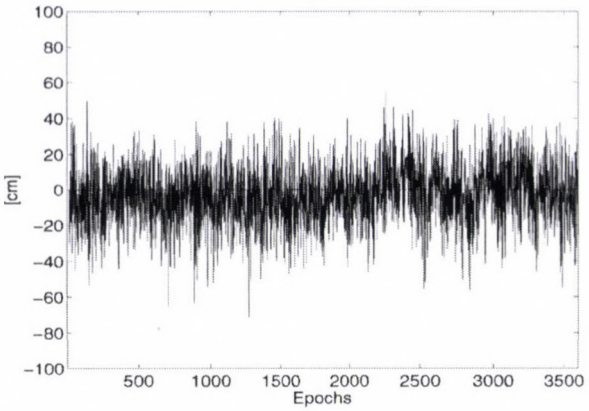


Fig. 34. Short baseline undifferenced residuals for Trimble 4700/MS 750, C1 code, PRN 4

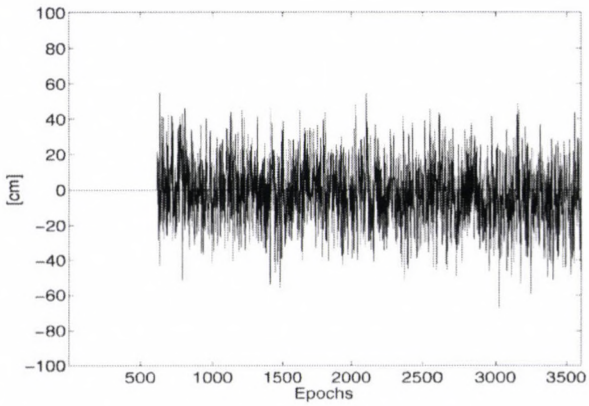


Fig. 35. Zero baseline undifferenced residuals for Trimble 4700/MS 750, P2 code, PRN 4

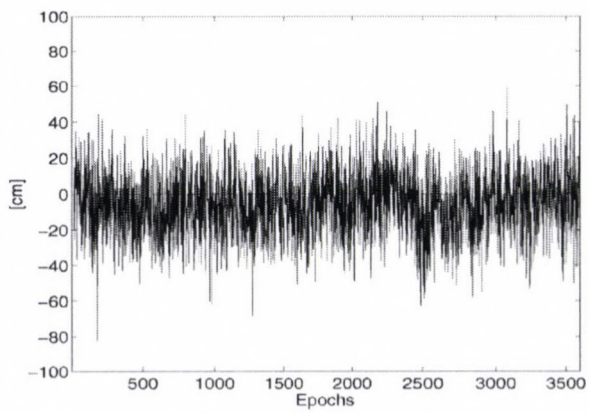


Fig. 36. Short baseline undifferenced residuals for Trimble 4700/MS 750, P2 code, PRN 4

The scale for L1 phase is –8 mm to 8 mm. The noise for L1 is about 1 mm. The residuals for the short baseline are some millimeters. For C1 and P2 code the noise is about 10 cm and because of multipath they increased to slightly over 20 cm.

Trimble 4700/MS 750

The scale for L1 phase is –8 mm to 8 mm. Because of the incidentally missing epochs, 50 minutes of continuous data were available for processing the zero baseline measurement. The noise for L1 is about 1 mm and the residuals are some millimeters for the short baseline. The noise for C1 and P2 code is about 30 cm and only slightly larger for the short baseline.

5. Conclusions

Table II summarizes the results obtained for satellite PRN 4 in the zero baseline measurement. These numbers were obtained from the zero baseline time series of the previous chapter and they represent the 'worst case' standard deviation since they refer to a low elevation satellite. Every receiver achieved millimeter or better precision. The precision of code observations is some decimeter.

Table II. The noise of the observations of PRN 4 for the zero baseline expressed as standard deviation

	L1 mm	C1 cm	P2 cm
Ashtech	0.3	9	50
Leica 500	0.7	3	3
Trimble 4000 SSi	0.3	14	24
Leica CRS 1000	1.0	7	22
Trimble 4700/MS 750	1.1	12	28

The supposed multipath effect can be clearly seen in most of the time series of the short baseline residuals. Table III shows the largest values from these time series (by visual judgement); these numbers give an impression of what we can expect for the multipath bias in the observations under these measurement conditions (no obstructions near the receivers, just the ground). The largest values of the phase residuals, probably due to multipath, are some millimeters for all receivers. On the C1 and P2 observables the effect is several decimeters, largest for the Trimble 4000 SSi receiver (80 cm), and smallest for the Leica 500 (10–15 cm). The internal filtering of the Leica 500 caused a systematic bias on the time series of the code residuals, but this filtering seems to decrease effectively the effect of multipath on the code residuals of the short baseline measurement.

The nominal noise of the various receivers differs significantly, the multipath equalizes the noise of the observations which is some millimeter (3–5) for phase and some decimeter (1–10) for code observations.

Table III. The maximum values of the residuals of PRN 4 for the short baseline

	L1 mm	C1 cm	P2 cm
Ashtech	3	60	60
Leica 500	4	10	15
Trimble 4000 SSi	3	20	80
Leica CRS 1000	4	20	20
Trimble 4700/MS 750	5	40	40

References

Braasch M S 1996: Global Positioning System: Theory and applications. In: B Parkinson and J Spilker eds, Vol. I, Chapter 14, American Institute of Aeronautics and Astronautics, Washington, DC

De Jonge P 1998: A processing strategy for the application of the GPS in networks. PhD thesis, TU Delft

Gurtner W 1994: RINEX: Receiver Independent Exchange Format. GPS World, July, Innovation, 48–52.

Langley R B 1998: In: GPS for geodesy. P Teunissen and A Kleusberg eds, Chapter 3, Second Edition, Springer Verlag, Heidelberg

Acta Geod. Geoph. Hung., Vol. 35(4), pp. 453–463 (2000)

THE FULFILMENT OF THE LAW OF LARGE NUMBERS FOR ARITHMETIC MEANS IN CASE OF INFINITE ASYMPTOTIC SCATTER

B HAJAGOS[1] and F STEINER[2]

[Manuscript received May 25, 2000]

It is quantitatively presented for a large error-type domain (from the Gaussian type to the Cauchy-distribution) how the accuracy of the arithmetic means increases (as an estimate of the location parameter) if the n sample size gets larger and larger. An even theoretically interesting result is that the law of large numbers can be fulfilled if the asymptotic scatter is infinite.

Keywords: asymptotic scatter; error-type; law of large numbers; L_2-norm; P-norm

1. Introduction

The principal question of mathematical statistics is properly formulated by Van der Waerden (1965) p.1: "How far can lie from the corresponding ideal values the quantities calculated from the samples", i.e., from the estimates? ("*Die Kernfrage der mathematischen Statistik ist ...: Wie weit können sich die aus der Stichprobe berechneten Größen von den entsprechenden idealen Werten entfernen?*") It is evidently a justified demand that the distances between estimates and the true value must diminish if the number of elements in the sample (n) increases, i.e., if the "law of large numbers" fulfils. (It should be mentioned that this is the formulation of the practitioners; mathematicians mean summarized and simply a set of theorems (with exactly given premissae) as "law of large numbers".) If n is really large, often decrease the deviations according to the rule

$$\sigma_{est} = \frac{A_{asympt}}{\sqrt{n}} \tag{1}$$

or

$$Q_{est} = \frac{Q_{asympt}}{\sqrt{n}}\,, \tag{2}$$

where σ_{est} and Q_{est} mean the scatter and semi-intersextile range of the estimates as global characteristics of the deviations; A_{asympt} and Q_{asympt} are called as asymptotic scatter and asymptotic semi-intersextile range. Estimates of the location parameter, especially the $\overline{x}$ arithmetic mean have often Gaussian probability distribution type (see e.g. Cramér 1945, p. 575) but in this case

$$Q = 0.9674 \cdot \sigma \tag{3}$$

[1]Institute of Mathematics, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary
[2]Geophysical Department, University of Miskolc, H–3515 Miskolc, Egyetemváros, Hungary

1217-8977/2000/$ 5.00

holds and therefore Eqs (1) and (2) hardly differ from another. Curiously enough, the law of large numbers is fulfilled in some cases if Eqs (1) and (2) are not valid (being A_{asympt} and Q_{asympt} infinite); this circumstance is treated quantitatively in Section 3 of the present paper.

In the already cited excellent book by van der Waerden (1965) unfortunately only the conventional (L_2-norm based) statistics is treated. No wonder: Huber's famous paper (1964) was published approximately simultaneously; the latter article is customarily accepted as the first step made to develop the so-called robust statistics (see e.g. Hampel et al. 1986).

The classical aspect can be illustrated by the followings (see van der Waerden p. 24) which is principally important: "Events of very small probability are customarily accepted as they would be almost impossible". (The complete sentence is in German: "*Ereignisse aber, die eine sehr kleine Wahrscheinlichkeit haben, pflegt man als fast unmöglich zu betrachten; man rechnet nicht mit ihrem Eintreffen bei einer einmaligen Realisierung der Bedingungen, unter denen sie theoretisch möglich sind. Auf diesem Prinzip beruht Überhaupt jede praktische Anwendung der Wahrscheinlichkeitsrechnung.*")

In contradiction to the above citate, practicioners often observe outliers, i.e., sample element values which are characterized by nearly vanishing probability densities. Applying classical (L_2-based) statistics, the results can be significantly distorted by outliers or even catastrophically. This circumstance is formulated as follows: the classical statistical methods are not resistant against outliers; therefore practitioners (and also mathematicians) published a lot of papers (even books) how can be eliminated this effect? (One of the mostly citated publications of such kind is Anscombe's (1960). According to the Abstract of the latter article "Numerous criteria for the rejection of outliers have been proposed and discussed during the past 100 years.")

Modern statistical procedures are *ab ovo* resistant. For example, if we derive procedures based on another norm instead of L_2, e.g. upon the P-norm (a menu of nine such norms are given in Table P.2 on page 20 by Steiner 1997), outliers cause only a small or even negligible ΔT modification of the most probable T value, where T means the estimate for the location parameter. The ΔT-value can be calculated also quantitatively by using the so-called "influence function" denoted by IC (Influence Curve, introduced by Hampel 1974) in the following simple way:

$$\Delta T = \frac{IC(F, T; X)}{n}, \tag{4}$$

where the distribution function of the parent distribution is denoted by F, T means here the T-estimation method itself, x is the value of one single outlier and n is supposed to be large enough. If the calculation of T is carried out by using one of the so-called M-estimations of Huber characterized by the function $\psi(.)$, the influence curve can be calculated for standardized probability distributions as

$$IC(F, T; x) = \frac{1}{\int\limits_{-\infty}^{\infty} \psi'(y) f(y) dy} \psi(x) \tag{5}$$

(see e.g. Huber 1981, p. 45) where x is an arbitrary value. (It should be mentioned that all customarily used $f(x)$ probability density models of error-distributions are unimodal and symmetrical to the origin but for the whole open interval $(-\infty, +\infty)$ $f(x) \neq 0$ holds.) If T is estimated by the arithmetic mean of the sample elements (denoted by $\overline{x}$), for arbitrary to the origin symmetrical $f(x)$

$$IC(F, \overline{x}; x) = x \tag{6}$$

holds and this results in the vanishing resistance of this classical method of T-determination. In the contrary, if the estimate of location parameter is calculated as the most frequent value (see e.g. Eqs (I-20) and (I-24) on pages 30–31 for $k = 2$ by Steiner 1997), for large values of $|x|$ the absolute values of the influence function (i.e., of $|IC(F, M; X)|$) decrease if $|x|$ increases (see e.g. the curve for $k = 2$ in Fig. 5.5 on page 157 by Steiner (1990) where the density function of the parent distribution is also indicated by dashed curve). The larger is $|x|$ the less is the IC-value, and therefore (see Eq. 4) a great degree of resistance is achieved calculating most frequent values (or otherwise expressed: minimizing the P-norm).

Hampel et al. (1986) demonstrated that the whole theory of robust statistics can be based on the influence function, this means that the IC-function is usable not only for quantitative characterization of the resistance. For example, the asymptotic variance can be calculated as

$$A^2_{asympt} = \int_{-\infty}^{\infty} IC^2(F, T; x) f(x) dx \,. \tag{7}$$

For the special case investigated in the present paper, i.e., for $T = \overline{x}$ (and for to the origin symmetrical $f(x)$) we get (see Eq. 6)

$$A_{asympt} = \sqrt{\int_{-\infty}^{\infty} x^2 f(x) dx} = \sigma \,, \tag{8}$$

i.e., the asymptotical scatter equals to the scatter of the parent distribution. This value can be very simply calculated as

$$A_{asympt} = \frac{1}{\sqrt{a-3}} \tag{9}$$

for the error-types of the $f_a(x)$ supermodel defined by

$$f_a(x) = \frac{\Gamma(a/2)}{\sqrt{\pi} \cdot \Gamma[(a-1)/2]} (1 + x^2)^{-a/2} \qquad (a > 1) \tag{10}$$

(see e.g. Steiner 1997 on page 366). Evidently Eq. (9) gives finite A_{asympt}-values only if $a > 3$; for $a = 3$ is A_{asympt} infinite. It will be shown, however, in Section 3 that the law of large numbers is fulfilled even if for the type-parameter a the relation $3 \geq a > 2$ holds.

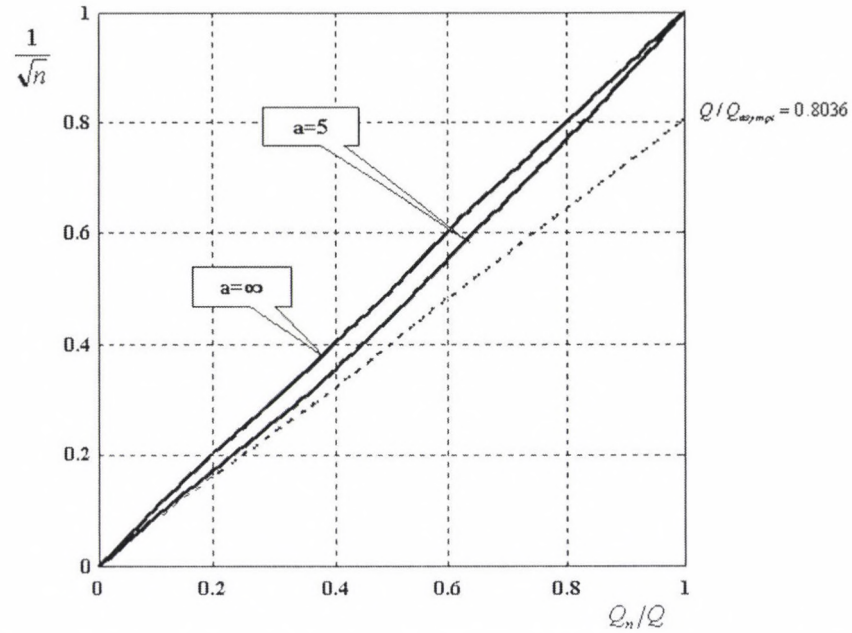


Fig. 1. Semi-intersextile range of the parent distribution (Q) divided by the semi-intersextile range of the arithmetic means (Q_n) calculated on ground of samples of n-elements coming from the (geo)statistical type characterized by $a = 5$. The tangent in the origin of this curve is also illustrated; the straight line defined by the points (0;0) and (1;1) corresponds to the Gaussian distribution

2. The case of finite asymptotic scatters

According to Eqs (2) and (3) it can be written

$$Q_{asympt} = 0.9674 \cdot A_{asympt} \,. \tag{11}$$

Let be given an $f_a(x)$ error distribution type characterized by a type parameter $a > 3$. If Monte Carlo investigations are made calculating the classical estimation for the location parameter, i.e., the arithmetic mean $\overline{x}$ for different sample sizes n (repeated N times), and (as error-characteristics of the $\overline{x}_n$-s) the empirical semi-intersextile range Q_n is determined on ground of the N (generally different) $\overline{x}_n$ values: 2/3 of these values lie in the interval $(-Q_n, +Q_n)$, $N/6$ values are greater than Q_n and $N/6$ values are less than $-Q_n$. To be independent from the parameter of scale, the abscissa on Figs 1 and 2 (belonging to $a = 5$ and $a = 9$) is the quotient Q_n/Q where the equation

$$\int_0^Q f_a(x)dx = 1/3 \tag{12}$$

defines the theoretical value of the semi-intersextile range Q of the parent distribution itself.

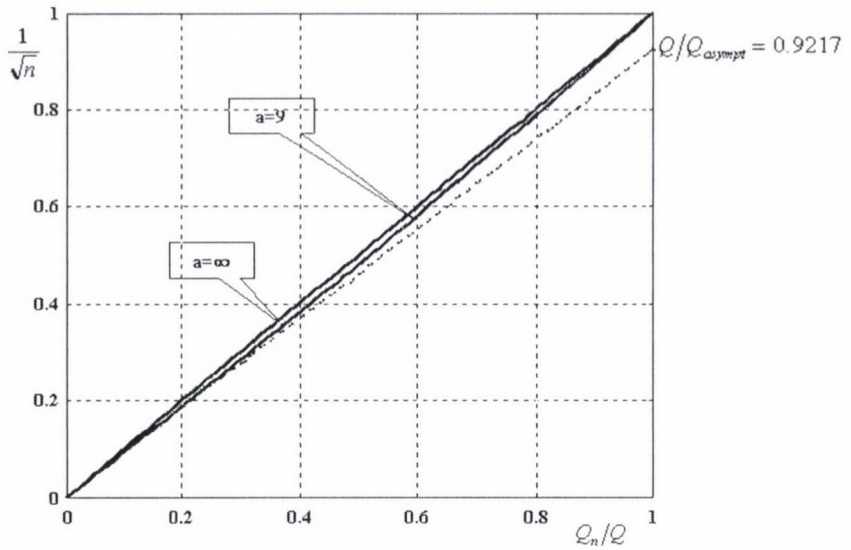


Fig. 2. Semi-intersextile range of the parent distribution (Q) divided by the semi-intersextile range of the arithmetic means (Q_n) calculated on ground of samples of n-elements coming from the Jeffreys type characterized by $a = 9$. The tangent in the origin of this curve is also illustrated; the straight line defined by the points (0;0) and (1;1) corresponds to the Gaussian distribution

If $1/\sqrt{n}$ is chosen as ordinate, Q/Q_{asympt} is the tangent in the origin of the curve got as result of the Monte Carlo calculations; these tangents are indicated by dashed straight lines in Figs 1 and 2. (What the numerical value belongs, Table I gives more accurate values for Q/Q_{asympt}.)

The curves follow accurately enough the following analytical form, if for simplification the notation $x = Q_n/Q$ is introduced:

$$\frac{1}{\sqrt{n}} = \frac{Q \cdot x}{Q_{asympt}} + \left[1 - \frac{Q}{Q_{asympt}}\right](x^u - w \cdot x^v \cdot \ln x)\,. \tag{13}$$

Table I

a	$t = 1/(a-1)$	Q/Q_{asympt}	u	v	w
9.0	0.1250	0.921677	2.1	1.8	0.1
5.0	0.2500	0.803558	2.3	1.6	0.2
4.0	0.3333	0.686611	2.4	1.57	0.3
3.2	0.4545	0.383977	2.6	1.45	0.5
3.1	0.4762	0.281401	2.62	1.48	0.55
3.0	0.5000	0	2.84	1.5	0.78
2.7	0.5882	0	3.36	1.7	0.9
2.4	0.7143	0	4.6	2.3	1.3
2.2	0.8333	0	7.0	3.4	2.0
2.1	0.9091	0	9.5	5.3	3.3

All four parameters (Q/Q_{asympt}, u, w and v) are contained in Table I for the type-parameters $a = 9$; 5; 4; 3.2 and 3.1.

It should be stressed that the tangents coincide with the curves on different $(0; 1/\sqrt{n})$ intervals: graphically determined, for $a = 5$ is this interval about (0; 0.1) and for $a = 9$ (0; 0.25), see Figs 1 and 2. (For the Gaussian type, i.e., for $a = \infty$ this interval is (0;1) — see the straight lines between the origin and the point (1; 1) on both figures, — as in this case well known for arbitrary n $Q_n = Q_{asympt}/\sqrt{n}$ holds.) In non-Gaussian cases a rather obscure formulation is customary: the $1/\sqrt{n}$ law is valid if n is large enough. The above mentioned intervals give concrete answers which n_0 is already "large enough" for the parent distributions in question: in case of $a = 5$ the relation $n \geq 100$ ($n_0 = 100$), in case of $a = 9$ only the relation $n \geq 16$ must be hold to apply the asymptotic rule. (On ground of the third, fourth and fifth row of Table I one can calculate also for $a = 4$; 3.2 and 3.1 more and more larger n_0 sample size values for these parent distribution types.) Naturally, on ground of Eq. (13) and using the already cited values of Table I, we are not any more forced to use only the asymptotic rule if we search after that n sample size which realizes the demanded Q_n/Q accuracy.

3. The case of infinite asymptotic scatter

If the quantity $t = (a - 1)^{-1}$ is introduced, it can be stated, on one hand, that even the theoretical expectation does not exist for $t = 1$ (i.e., for $a = 2$) being this type equal to the Cauchy distribution (see e.g. p. 432 by Rényi 1954). On the other hand, if $t \geq 0.5$ ($a \leq 3$), the asymptotic scatter of the arithmetic means are infinite (see Eq. 9). Consequently, in this Section the type-domain $0.5 \leq t < 1$ ($3 \geq a > 2$) is investigated. As Q_{asympt} is infinite in this type-region, consequently Eq. (13) simplifies to

$$\frac{1}{\sqrt{n}} = x^u - w \cdot x^v \cdot \ln x\,, \tag{14}$$

(where x means the quotient Q_n/Q). Carrying out the Monte Carlo calculations, the u, w and v-values can be determined; for $a = 3$; 2.7; 2.4; 2.2 and 2.1 these values are also given in Table I.

To give a guess about the accuracy of the Monte Carlo determination of the parameters u, v and w, Figs 3b, 3c and 3d show the curves of these parameters vs. $t = (a - 1)^{-1}$; for completeness, in Fig. 3a is also the curve Q/Q_{asympt} vs. t shown (naturally only for the type-domain $0 \leq t < 0.5$). The approximate formulas belonging to this curves can be given as follows:

$$Q/Q_{asympt} = \sqrt{0.998369 - 0.821812 \cdot t - 2.33797 \cdot t^2} \qquad (0 \leq t < 0.5); \tag{15a}$$

$$u = \frac{2 - t}{1 - t} \qquad (0 \leq t < 0.9); \tag{15b}$$

$$v = 2.25 - \sqrt{t} + 8 \cdot t^{7.5} \qquad (0 \leq t < 0.9); \tag{15c}$$

$$w = t + 4 \cdot t^{5.5} \qquad (0 \leq t < 0.9)\,. \tag{15d}$$

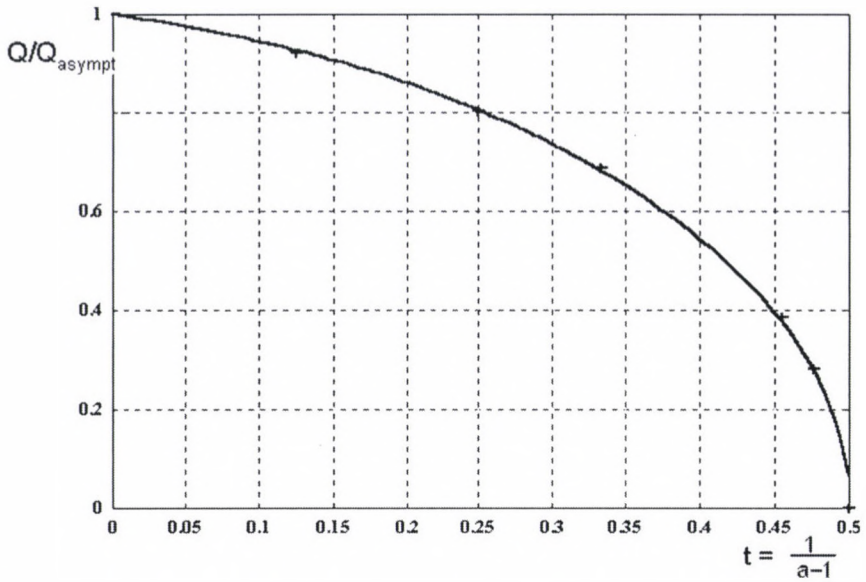


Fig. 3a. The $[Q/Q_{asympt}; t]$ points in the range $0 \leq t \leq 0.5$ lay on an ellipse. As Eq. (13) shows, this value can play an important part by the analytical description of the curves of $1/\sqrt{n}$ vs. Q_n/Q (see also Figs 1 and 2)

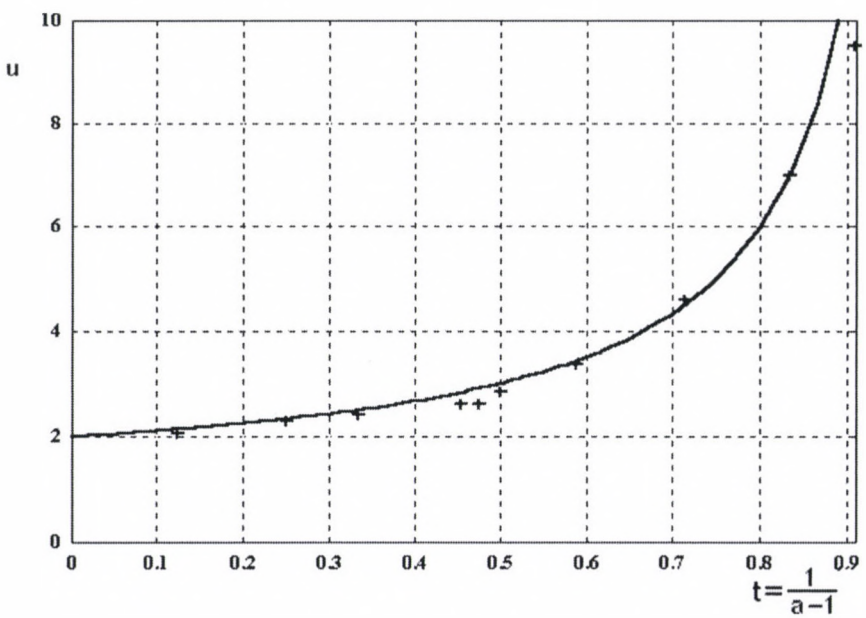


Fig. 3b. The exponent u figuring in Eqs (13) and (14) vs. the type parameter $t = (a-1)^{-1}$

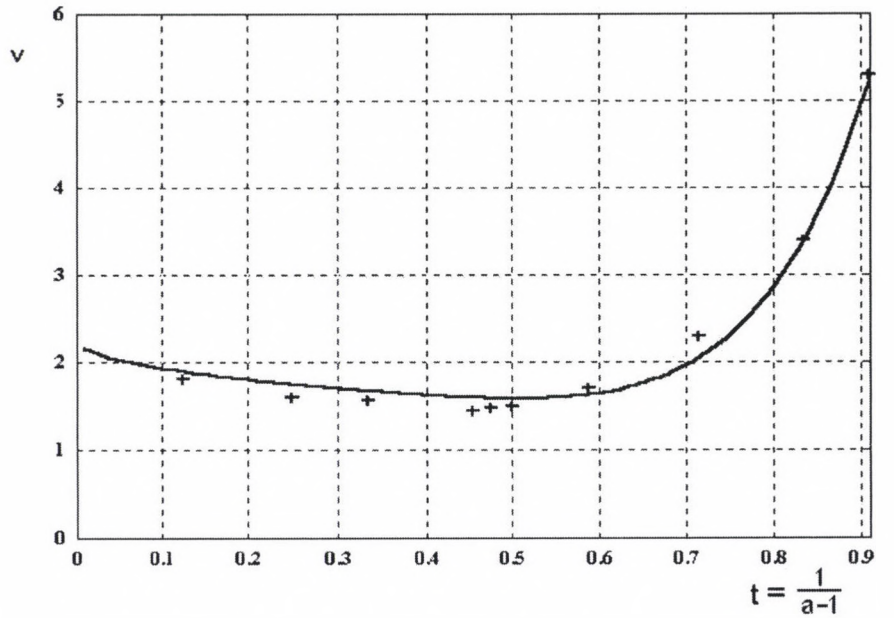


Fig. 3c. The exponent v figuring in Eqs (13) and (14) vs. the type parameter $t = (a-1)^{-1}$

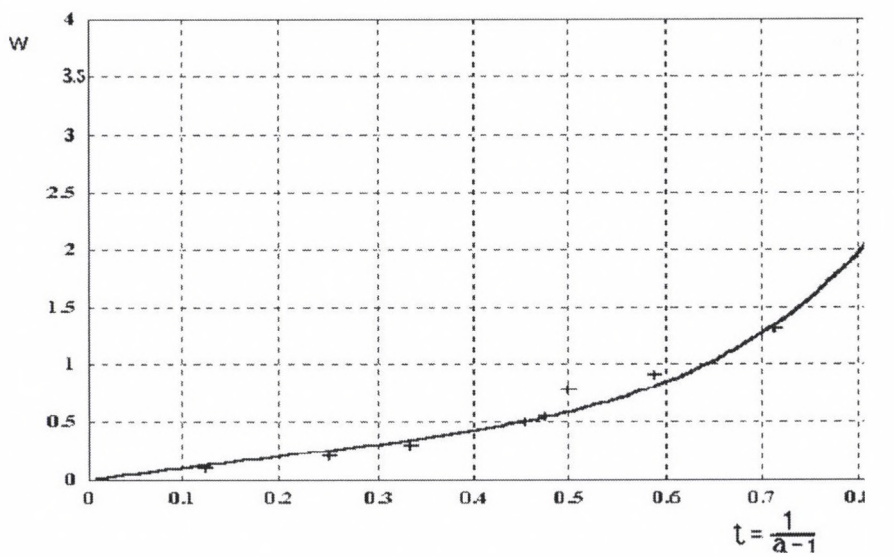


Fig. 3d. The factor w figuring in Eqs (13) and (14) vs. the type parameter $t = (a-1)^{-1}$

Figures 4 and 5 are very instructive: how can the practically formulated law of large numbers fulfilled (i.e., how decreases the Q_n/Q error of the $\overline{x}_n$ arithmetic means if the sample size increases) even if the asymptotic characteristics (A_{asympt}

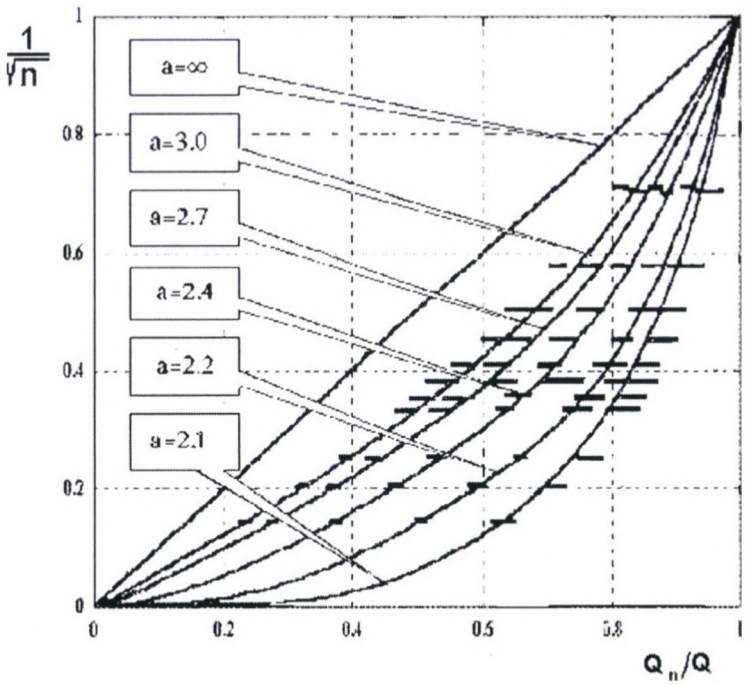


Fig. 4. The $1/\sqrt{n}$-curves vs. Q_n/Q according to Eq. (14) and to the four parameters contained in Table I for the parent distribution types $a = 2.1$; 2.2; 2.4; 2.7 and 3.0; the straight line defined by the points (0;0) and (1;1) characterizes the Gaussian distribution ($a = \infty$). The horizontal sections for $n = 2$; 3; 4; 5; 6; 7; 8; 9; 16; 25 and 49 correspond to the total range of the Monte Carlo results

and Q_{asympt}) are infinite. All $\left(\frac{1}{\sqrt{n}};\, Q_n/Q\right)$ curves start horizontally but the Q_n/Q-domain which practically coincides with the abscissa tends to the unity if $a \to 2$, on the one hand, and on the other hand, the length of this domain tends to zero if $a \to 3$. In Fig. 5, where the ordinate is enlarged, even for $a = 2.7$ (i.e., near to $a = 3$) is hardly to be seen this short, approximately horizontal interval.

At last, the most important conclusion must be formulated on ground of Figs 4 and 5: although the law of large numbers is really fulfilled also in the type-domain ($3 \geq a > 2$), this fulfilment is less and less economical if a decreases: to $Q_n/Q = 0.5$ we need a sample size of $n = 278$ if $a = 2.1$ (being $1/\sqrt{n} = 0.06$, see Fig. 5), and a sample size of only $n = 9$ if $a = 3$ (being $1/\sqrt{n}$ equal about to 0.34, see Fig. 4); in case of the Gaussian type to reach $Q_n/Q = 0.5$ is well known $n = 4$ enough. If we demand greater accuracy than $Q_n/Q = 0.5$, say, be $Q_n/Q = 0.28$, at the parent distribution type $a = 2.1$ we get the practically unrealizable sample size $n = 40000$ (see Fig. 5), while the type $a = 3$ needs to the same accuracy only $n = 31$ data, see Fig. 4, not to say anything from the Gaussian case: to reach $Q_n/Q = 0.28$, already $n = 13$ is enough.

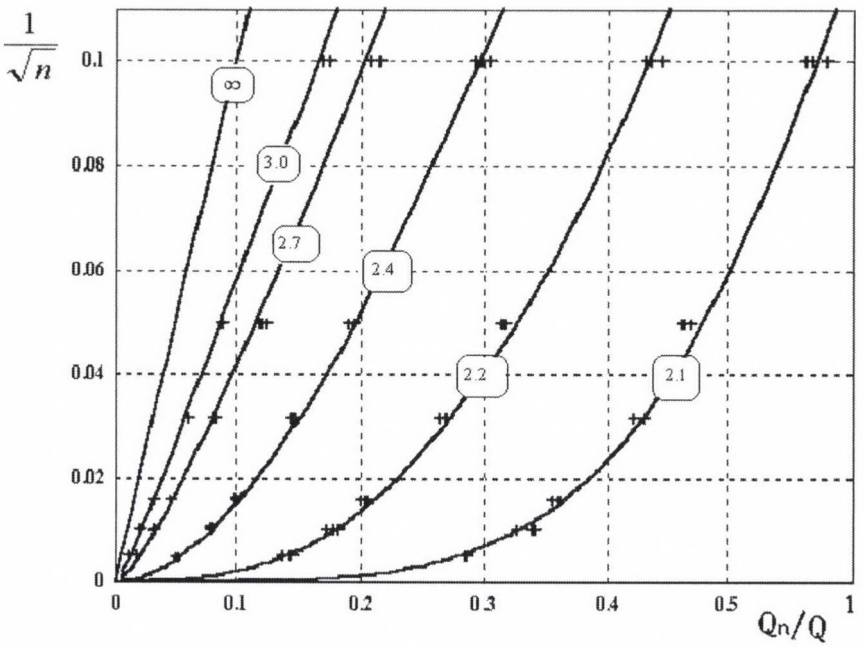


Fig. 5. The $1/\sqrt{n}$-curves vs. Q_n/Q according to Eq. (14) and to the four parameters contained in Table I for the parent distribution types $a = 2.1$; 2.2; 2.4; 2.7 and 3.0; the straight line defined by the points (0;0) and (0.1;0.1) characterizes the Gaussian distribution ($a = \infty$). The + signs correspond to Monte Carlo results for $n = 100$; 400; 1000; 4000; 10000 and 40000

4. Concluding remarks

In the present paper the semi-intersextile range was consequently used as error-characteristics. Some (not all!) favourable behaviours of Q is treated by Steiner and Hajagos (2000b); the authors propose therefore also to the readers to apply instead of the scatter the semi-intersextile range as error-characteristics. (For the Gaussian distribution $Q \approx \sigma$ holds, see Eq. 3.)

In Sections 2 and 3 of the present paper (analytically and on figures, too) the error of the arithmetic mean $\overline{x}$ is given not only for the asymptotic case (as usual) but for the whole interval ($1 \leq n < \infty$) of the sample sizes.

The thorough investigation of the $\overline{x}$-s means not in the least that the authors prefer or propose the general use of the L_2-norm based statistical algorithms. Not at all, although Steiner and Hajagos (2000a) demonstrated that for a great type-domain of the generalized Gaussian supermodel

$$f_p(x) = \frac{p^{1-1/p}}{2 \cdot \Gamma(1/p)} \cdot e^{-|x|^p/p} \tag{16}$$

the asymptotic efficiency of the L_2-norm is significantly greater than that of the L_1-norm. On the other hand, however, L_1 procedures are resistant against outliers, in the contrary, L_2-norm based procedures are extremely outlier-sensitive.

Using P-norm based statistical algorithms, the resistance can be even greater than that of the L_1-methods (see e.g. Fig. 8.15 and Fig. 10.13 by Steiner 1997), and what the asymptotic efficiency (e) belongs, Steiner and Hajagos (2000a) have proven that the relation

$$e(P) > e(L_1) \tag{17}$$

holds for all types of the $f_a(x)$ supermodel (see Eq. 10).

References

Anscombe F J 1960: *Technometrics*, 2, 123–147.
Cramér H 1945: Mathematical Methods of Statistics. Almqvist and Wiksells, Uppsala
Hampel F R 1974: *J. Amer. Statist. Ass.*, 62, 1179–1186.
Hampel F R, Ronchetti E M, Rousseeuw P J, Stahel W A 1986: Robust Statistics. The Approach Based on Influence Functions. Wiley, New York
Huber P J 1964: *Ann. Math. Statist.*, 35, 73–101.
Huber P J 1981: Robust Statistics. Wiley, New York
Rényi A 1954: Calculation of Probabilities (in Hungarian). Tankönyvkiadó, Budapest
Steiner F 1990: Introduction to Geostatistics (in Hungarian). Tankönyvkiadó, Budapest
Steiner F ed. 1997: Optimum Methods in Statistics. Akadémiai Kiadó, Budapest
Steiner F, Hajagos B 2000a: *Acta Geod. Geoph. Hung.*, 35, 305–312.
Steiner F, Hajagos B 2000b: *Acta Geod. Geoph. Hung.*, 35 (present issue)
van der Waerden B L 1965: Mathematische Statistik. Springer, Berlin-Heidelberg-New York

Acta Geod. Geoph. Hung., Vol. 35(4), pp. 465–473 (2000)

ATTENUATION OF CODA WAVES IN HUNGARY

M KISZELY[1]

[Manuscript received January 13, 2000]

Microearthquake codas from 19 shallow depth local events have been analysed in Hungary. We estimate coda attenuation Q_c in the frequency range 6 to 24 Hz used Sato's (1977) single scattering model for a coda window of 12.5±2.5 s beginning at twice the S-wave travel time. Travel times were between 2–23 sec. The stations showed similar values of Q_c which were about 400 at 6 Hz and about 1400 at 24 Hz. The Q_c values in Hungary indicate moderate tectonic characteristic.

Keywords: attenuation; coda waves; seismic quality factor

1. Introduction

Attenuation of seismic waves in the lithosphere at high frequencies (1–20 Hz) is an important property for the study of the Earth's structure. One of the most useful parameters in describing this attenuation is the seismic quality factor Q, which contains meaningful information even at short distances (Singh et al. 1982). In general, the behaviour of the seismic coda is described by the average decay of the envelope of a seismogram, rather than by the amplitude of a particular arrival. Bisztricsány (1970) calculated for Hungarian earthquakes codas but he worked with surface waves at low (0.01–0.1 Hz) frequencies.

The Q value — determined from coda waves as well as from direct shear waves using spectral methods — has been demonstrated to be a function of frequency. The results of several investigations (Aki 1980, Aki and Chouet 1975) confirm that the envelope of the coda can be adequately explained by the scattering of primary elastic waves in a random medium. Coda wave attenuation is caused by a combination of two effects: scattering and anelastic attenuation. Since the amplitude decay with distance (r) for both of these processes generates the same mathematical form αe^r, they cannot be easily separated. Dainty (1981) has suggested that the frequency dependence of coda Q, in the frequency range 1 to 20 Hz, is primarily due to scattering while anelastic attenuation is relatively frequency independent. Hoshiba (1993) found that coda Q_c estimates could be a good approximation to intrinsic attenuation.

There is a strong correlation between the dependence of Q on frequency and the tectonic complexity of a region (Aki 1980a,b, and Pulli 1984). Areas of strong tectonic heterogeneity have strong frequency dependence of coda Q as compared to that found in stable shield areas.

[1]Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H–1112 Budapest, Meredek u. 18, Hungary

1217-8977/2000/$ 5.00

We used Sato's formulation (1977) to estimate Q_c through measurements of the amplitude decay of coda waves with time. Assuming discrete randomly distributed heterogeneities and single isotropic scattering, the ratio of the mean energy density of the scattered (E_{sc}) waves to the primary S-wave energy (E_s) from a point source can be approximated at the frequency ω by the square of the amplitude ratio:

$$[A_c(t)/A_s]^2 \qquad \text{as}$$

$$E_{sc}(r,t\backslash\omega)/E_s(r\backslash\omega) = [A_c(t)/A_s]^2 = n\sigma u\beta e^{\omega(ts-t)/Q_c}K(a)\,,$$

where r is the distance between the source and the receiver, A_s is the maximum amplitude of S waves, $A_c(t)$ is the amplitude of the coda around the time t, σn is the effective scattering coefficient , β is the S-wave velocity, t is the travel time of the coda wave, t_S is the travel time of the S-wave, both measured from the earthquake origin time, $K(a) = 1/a\ln[(a+1)/(a-1)]$ and $a = t/t_s$. Taking the logarithm of both sides and rearranging terms, we obtain

$$\log_{10}[(A_c(t)/A_s)^2K^{-1}(a)] = C - b(t-t_s)\,,$$

where $b = 2\pi f(\log_{10}e)/Q_c$ and $C = \log_{10}(n\sigma\beta u)$. This is a linear equation, thus for each frequency, Q_c can be determined from the slope b.

2. Data

Out of 138 events detected in Hungary from January 1995 to October 1999 only 19 records allowed to calculate coda. Only these had good signal to noise ratios, and were enough long to include two times the travel time of S-wave. Data have arrived from the Microseismic Monitoring Network established by the Paks Nuclear Power Plant Ltd. operated by GeoRisk Ltd. All these stations use Lennartz three directional 1 Hz geophones with sampling rates of 125 per second. In the determination of the parameters of earthquakes we used the seismological stations data of the Geodetic and Geophysical Research Institute as well. For coda analysis horizontal components of microearthquakes were used. For locating the events HYPO71 (Lee and Lahr 1975) was used.

The velocity model is given in Table I. We used 1.78 vp/vs Poisson ratio to calculate S-wave velocity. Table II shows detailed information of events. (Tóth et al. 1996, 1997, 1998, 1999). The locations of stations and events are displayed in Fig. 1. The travel times were between 2–23 sec. Among these events there were four explosions too. The data that derived from these explosions were calculated separately. The depths of earthquakes were about 10 km. The explosions occurred near the surface.

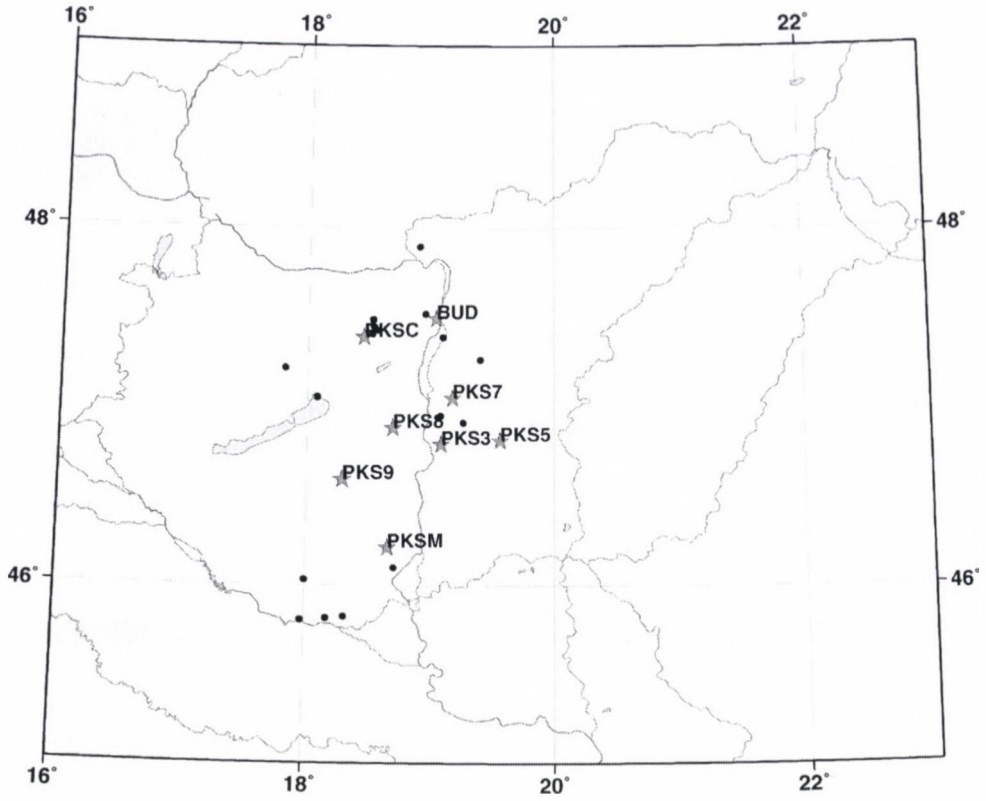


Fig. 1. Map of Hungary with stations (stars), and epicentres (points)

Table I. Velocity Model

Velocity (vp) km/s	Depth km	Thickness km
5.60	0.0	20.0
6.57	20.0	11.0
8.02	31.0	∞

3. The process of calculation

— We bandpass filtered each record for 5 frequencies between 3 and 24 Hz with an eight-pole Butterworth filter and bandwidths of two-thirds of the centre frequency used (Table III).

— Root mean square amplitudes A_c were determined for sliding windows of 3 sec width and 0.5 sec of advance. This time window was chosen to smooth out irregularities in the amplitude.

— We plotted $F(t) = \log_{10}[(Ac/As)^2/K(a)]$ versus $(t - t_s)$.

t_s is the elapsed time of the S-wave

t is the elapsed time of the coda wave, both measured from the earthquake origin time.

$K(a) = 1/a \ln[(a+1)/(a-1)]$ and $a = t/t_s$.

— We fitted a straight line by the least squares method to linear portion of $F(t)$ to determine the slope from which Q_c was obtained. The linear portion of $F(t)$ was visually chosen for each seismogram. The time window started always after twice the S-wave travel time, when the amplitude appeared to decay monotonically. The width of windows was between 10 and 12 sec.

The procedure is shown in Fig. 2. Each time was calculated from origin time. Q_c was calculated from the slope of the fitted line. The interval over which the slope

Table II. Located events

ID	Date	Time	Lat.	Long.	ML	Locality	Stations
1.	1995.09.18	08:26:10.8	47.89	18.87	3.7	Börzsöny mt.	BUD
2.	1996.03.28	06:31:22.2	46.91	19.25	3.0	Szabadsz.	PKS 7,3,5
3.	1996.10.21	01:23:25.5	47.38	19.07	2.5	Dunaharaszti	PKS 7
4.	1997.11.27	10:40:56.1	47.26	19.38	2.5	Nyáregyháza	PKS 8,C
5.	1998.01.12	04:15:45.8	47.05	18.07	2.0	Balatonfűzfő	PKS C
6.	1998.03.11	10:55:45.5	47.47	18.51	1.8	Csákvár (expl.)	PKS C
7.	1998.05.08	04:06:54.0	47.51	18.93	2.0	Budakeszi	PKS C
8.	1998.05.13	10:27:32.0	47.44	18.52	2.2	Vértesboglár	PKS C
9.	1998.06.01	13:07:46.6	45.80	17.93	2.7	Vejti	PKS M
10.	1998.07.03	15:35:21.1	46.93	19.04	2.1	Szalksz.	PKS M
11.	1998.12.06	14:35:22.0	47.20	17.82	2.6	Hárskút	PKS M
12.	1999.02.08	03:32:26.2	46.95	19.06	1.2	Szalksz.	PKS 7
13.	1999.06.04	02:09:58.9	45.82	18.30	3.0	Drávaszab.	PKS M
14.	1999.06.20	14:28:31.3	47.40	18.50	1.8	Vértes (expl.)	PKS C
15.	1999.09.07	20:30:00.0	46.80	17.95	1.4	Bálványos (exp)	PKS 8, C
16.	1999.09.09	06:15:58.7	45.81	18.17	1.9	Diósviszló	PKS M
17.	1999.09.09	20:59:59.8	46.10	18.69	1.7	Dunaszek (exp.)	PKS M
18.	1999.10.09	10:51:34.0	46.03	18.00	2.0	Szabadszentk.	PKS M
19.	1999.10.13	18:53:29.6	47.42	18.85	1.9	Budaörs	BUD

Table III. Frequency bands of filtered data

Centre Frequency Hz	Width Hz
3	2
6	4
10	7
16	10
24	16

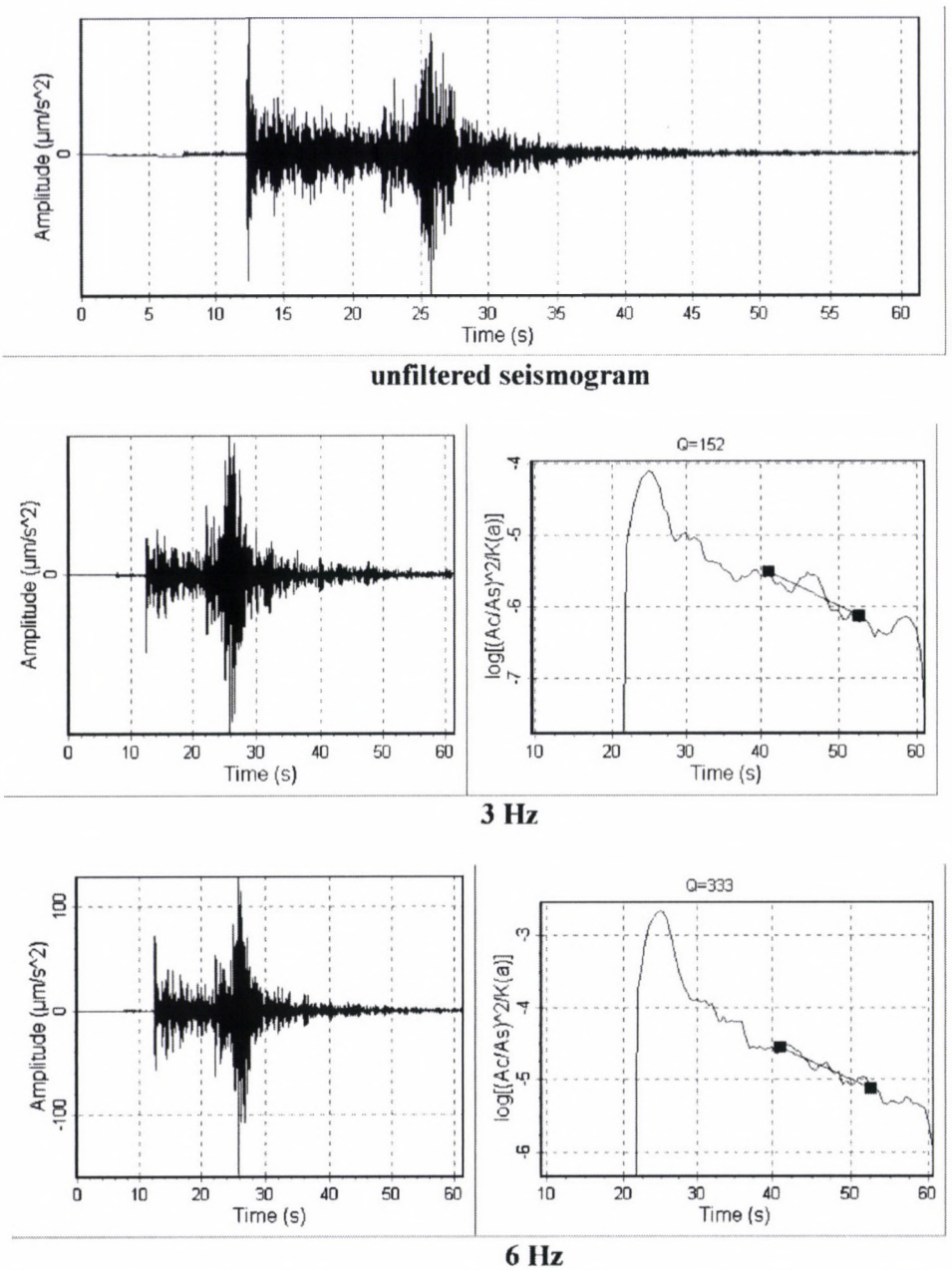


Fig. 2a. Procedure for determining coda Q values. The unfiltered seismogram is plotted at the top. The central frequency of filtering is printed on the bottom of figures. $F(t)$ versus lapse time are plotted on the right side. Q_c was calculated from the slope of the fitted line

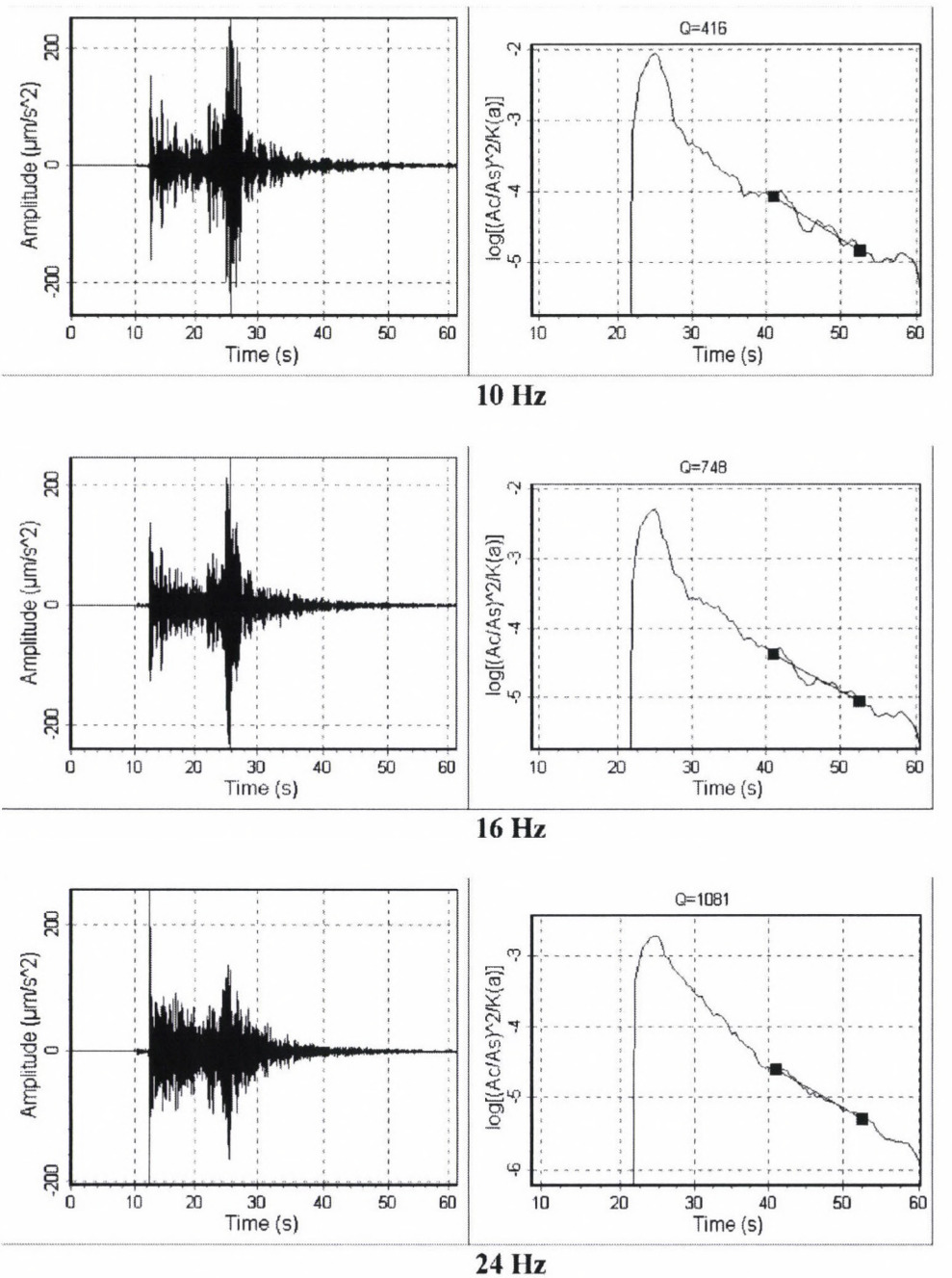


Fig. 2b. Procedure for determining coda Q values. The central frequency of filtering is printed on the bottom of figures. $F(t)$ versus lapse time are plotted on the right side. Q_c was calculated from the slope of the fitted line

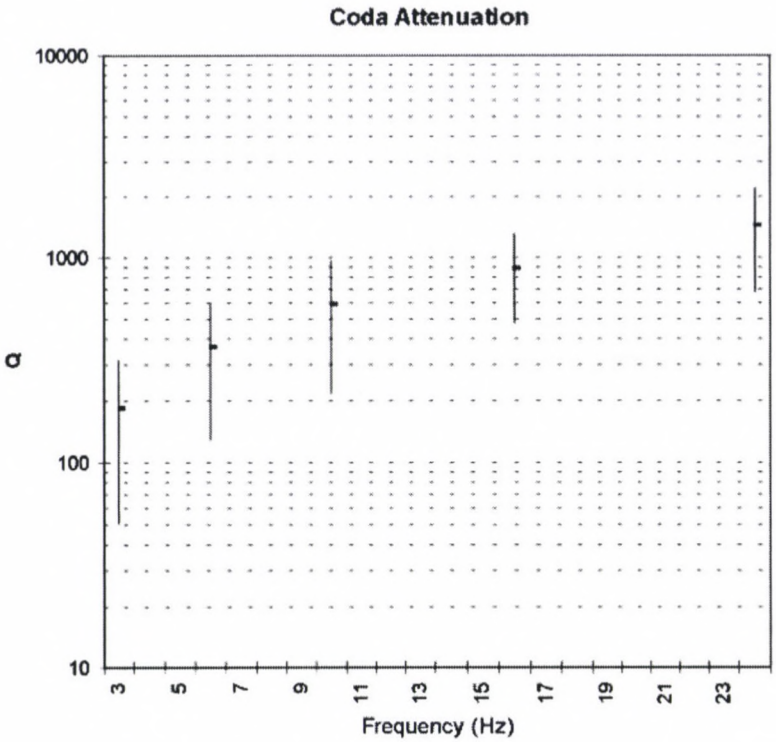


Fig. 3. Coda Q values obtained as a function of frequency. Plotted values were averaged values using lapse time windows between 10 and 12 sec after 2 times the S-wave arrival. Vertical bars indicate ± scattering

Table IV. Coda Q attenuations at earthquakes, and at explosions averaged for a given frequency at each stations

	3 Hz	6 Hz	10 Hz	16 Hz	24 Hz
Q Earthquake	196	416	662	900	1417
Q Explosion	92	179	330	639	1313

was determined, is indicated by straight lines. The event shown is Vejti 1998.06.01. at the station PKSM.

Q_c was determined using all the events shown in Table II. We averaged the Q_c values obtained at different stations, for a given frequency. The results are shown in Fig. 3. $Q_c(f)$ in this study is an averaged estimation of coda attenuation in Hungary. The values are in Table IV. While Q_c increases with frequency at explosions the values are less then at earthquakes. This result obtained Su (1991) as well. Su presented a new quarry blast — earthquake discrimination method based on the analysis of seismic coda waves.

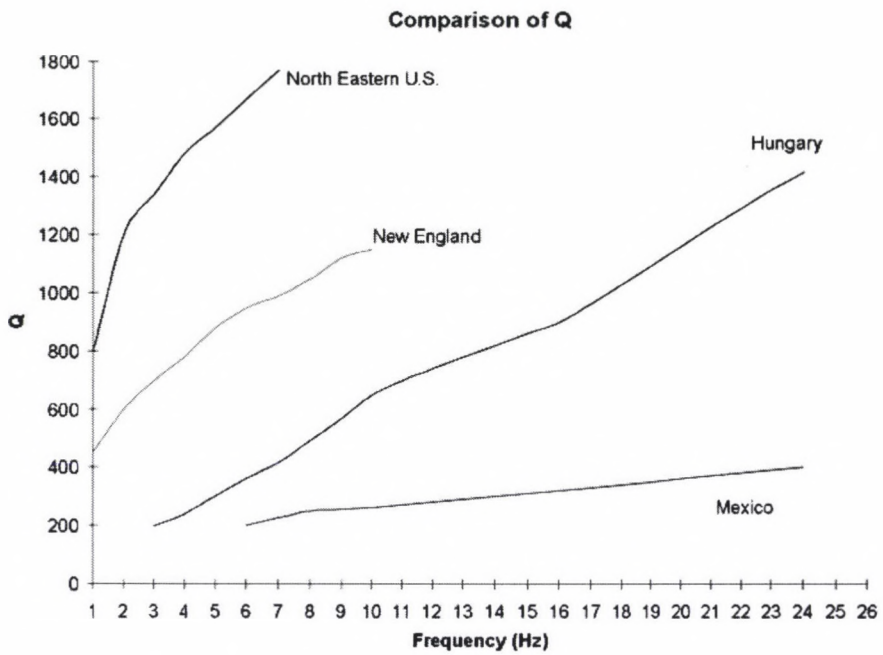


Fig. 4. Comparison of Q obtained at different regions

The number of used events is very low, it would be good to collect more data.

Dominguez et al. (1997) found that Sato's single scattering theory to estimate coda Q at frequencies lower than 6 Hz is unable to estimate coda Q. We have found this property of coda too, and negative Q_c occurred according to Zeng's (1991) model when a media with little intrinsic attenuation would have a negative Q_c.

As many authors have pointed out Q is a tectonic indicator. High values occur at stable regions, while low values are associated with active tectonic region (Singh and Herrmann 1983). The Q_c values in Hungary indicate moderate tectonic characteristic. In Fig. 4. Q-s obtained at different regions were compared (Singh and Herrmann (1983) in the North Eastern U.S., Pulli (1984) in New England, Dominguez et al. (1997), in Baja California, Mexico).

4. Results

$Q_c(f)$ in this study was estimated by Sato's formulation, and these estimations could be a good approximation to the intrinsic attenuation. Within the range of variation, stations showed similar values of Q_c which were about 400 at 6 Hz and about 1400 at 24 Hz. The Q_c values in Hungary indicate moderate tectonic characteristic. While Q_c increases with frequency, the values are less at explosions than at earthquakes.

Acknowledgements

The Microseismic Monitoring Network Project is sponsored by Paks Nuclear Power Plant Ltd. and carried out by GeoRisk Geophysical Research and consulting Ltd. The author is grateful to Tamás Maróti helping in the programming.

References

Aki K 1980a: *J. Phys. Earth.*, 21, 50–60.
Aki K 1980b: *J. Res.*, 85, 6496–6504.
Aki K, Chouet B 1975: *J. Geophys. Res.*, 80, 3322–3342.
Bisztricsány E 1970: *Geofizikai Közl.*, 19, 21–49.
Dainity A 1981: *Geophys. Res. Lett.*, 11, 1126–1128.
Dominguez T, Rebollar C J, Fabriol H 1997: *Bull. Seism. Soc. Am.*, 87, 1368–1374.
Hoshiba M 1993: *J. Geophys. Res.*, 98, 15809–15824.
Lee W H, Lahr J C 1975: Hypo71 (revised): a computer program for determining hypocenter, magnitude and first motion pattern of local earthquakes. U.S. Geol. Surv. Open-File Rept. 75–311.
Pulli J J 1984: *Bull. Seism. Soc. Am.*, 74, 1149–1166.
Sato H 1977: *J. Phys. Earth.*, 22, 27–41.
Singh S, Herrmann R B 1983: *J. Geophys. Res.*, 88, 527–538.
Singh S K, Apsel R J, Fried J, Brune J N 1982: *Bull. Seism. Soc. Am.*, 72, 2003–2016.
Su F, Aki K, Biswas N N 1991: *Bull. Seism. Soc. Am.*, 81, 162–178.
Tóth L, Mónus P, Zsíros T 1996–1999: Hungarian Earthquake Bulletin, GeoRisk, Budapest
Zeng Y 1991: *Bull. Seism. Soc. Am.*, 81, 1022–1029.

Acta Geod. Geoph. Hung., Vol. 35(4), pp. 475–480 (2000)

Professor dr Péter Biró septuagenario

The community of Hungarian geoscientists respectfully congratulate Professor Péter Biró, Ordinary Member of the Hungarian Academy of Sciences and Professor Emeritus of the Budapest University of Technology and Economics on his 70th birthday occurring on August 8, 2000.

Professor Biró was born in Budapest in 1930. He finished his school there, and graduated at the Faculty of Civil Engineering of the Technical University of Budapest (TUB) in 1952. He was awarded the Doctor Degree by the same University in 1961. He was Professor between 1976–2000, currently is Professor Emeritus at TUB.

All of his professional activities has connected him to the higher education, the scientific research and partly to the practice of geodesy and surveying, but mostly to the geodetic science.

His work in professional education has been carried out at the Department of Surveying and since 1961 at that of Geodesy of the Faculty of Civil Engineering, TUB. After some years of assistance he became lecturer (1961), associate professor (1967) in Geophysics, Geodesy, Cosmic Geodesy, National Geodetic Control Nets in Hungarian and in English in different periods. He became Head of Department (1971) and full professor (1976) in Geodesy. He is author, resp. co-author of several lecture notes in the subjets mentioned before. He has frequently visited universities abroad (esp. in Austria, Germany and Great Britain) and has good co-operations with several of them. He was the chairman of the Scientific Committee of the Faculty (1979–1991). As a recognition of his merits in professional education, he was awarded the title "Distinguished Professor of the Faculty" (1984), and

1217-8977/2000/$ 5.00

became the Director of the Institute for Geodesy, Surveying and Photogrammetry of the Faculty of Civil Engineering (1987–1992). As experienced university professor he became Vice-Rector (1991) and was elected to Rector of TUB for the period 1994–1997, Chairman of the Hungarian Rectors' Conference (1994–1995), Chairman of the Budapest University Alliance (1996–1997), Vice-President of the Hungarian Accreditation Commission in Higher Education (1994–1997) and Chairman of the Doctoral Board at the TUB (1993–1994, 1997–2000). He has been the Honorary Member of the Senate of the University of Karlsruhe. He was awarded the designation "European Engineering Educator" (Ing-Paed IGIP) and has been Honorary Member of the International Society for Engineering Education since 1996. He was active in Commission 2 (Professional Education) of FIG for more periods. He was awarded by the Honorary Medal of TUB (1998) and that of "For the Academic Education in Geodetic Sciences" (1999).

His scientific activity started with researching instrumental errors (such as the wobble of the line of collimation of precise levelling instruments caused by moving the focussing lens), the results being published in his first scientific paper (in Hungarian) in 1956. Soonly his interest turned to the astronomic and physical geodesy and geodynamics. His researches led to important results in the following topics:

— some theoretical and practical problems of geodetic gravimetry;

— the practical use of gradiometry in the Hungarian National Control Network;

— the effect of the time variation of the earth's gravity field onto the heights of bench marks and astronomic co-ordinates of geodetic control stations, "relative" and "true" surface movements;

— the common evaluation of terrestric and cosmic geodetic and gravimetric observations;

— some principal problems of the mathematical representation of the earth's gravity field.

His results were published in his books (such as Time variation of height and gravity (in English), 1983), his nearly 60 scientific papers and reports (such as that to the IUGG/IAG General Assemblies, etc.) in English, German and Hungarian (see list of publications attached).

He was member of several special study groups (SSG) of IAG subsequently.

The Hungarian Academy of Sciences (HAS) acknowledged his scientific merits by electing him its corresponding (1985) and full member (1990), respectively. In the recent decades he have had an important role in geodetic sciences as member, later chairman (1981–1990) of the Hungarian Geodetic Commission, member and chairman (1990–1994) of the Hungarian National Committee for IUGG and IAG. He had a role in the management of the Hungarian scientific activity as Vice-President (1990–1996) of the Section Earth Sciences HAS; member(1990–1996) of the Commission of International Affairs of the HAS; elected member (1996–1999) of the Presidium HAS.

An acknowledgement of his activity abroad is the corresponding membership of the German Geodetic Commission at the Bavarian Academy of Sciences in Munich (1987).

Beside his activity in higher education and scientific research Professor Biró spent some years in geodetic surveying practice. He participated in geodetic and astronomic observations of the Hungarian Geodetic Control Network (1962–1963) and acted as Vice-Director (1969–1971) of the National Institute of Geodesy, Surveying and Mapping. He was the chairman (1992–1994) of the Advisory Board of the Hungarian National Survey.

He had an active role in the professional Society for Geodesy, Surveying, Mapping and Geoinformatics as secretary, later Head of Section Geodesy. Later he became co-chairman of the Society, now "Member for Life".

He is member of advisorial and editorial boards of the Journal of "Allgemeine Vermessungs-Nachrichten", "Acta Geodaetica and Geophysica Acad. Sci. Hung." and "Geodézia és Kartográfia".

As appreciations of his merits in scientific research and higher aducation Professor Biró was awarded several honours (beside that of mentioned before). He is holder of the State Széchenyi-Prize, the ministerial prizes L Eötvös and A Fasching resp. and the Honorary Medal "Lázár Deák" of the Geodetic and Cartographic Society.

We wish Professor BIRÓ at his seventieth birthday good health for many years coming, that his activities in the professional, scientific and university public life should not cease. We also wish him further successes and happiness in his personal life, too.

József Ádám

J Ádám

List of publications

A) Books, lecture notes

Általános geodézia (Műszaki értelmező szótár, társszerzőkkel), Szerk.: Dr. Rédey István, Terra, Budapest, 1961

A geodézia korszerű irányai (Szakmérnöki jegyzet, co-author: L Homoródi), Tankönyvkiadó, Budapest, 1966

Geofizika (Egyetemi jegyzet, co-author: L Homoródi), Tankönyvkiadó, Budapest, 1966

Műszaki Lexikon (geodéziai cikkei, co-author: E Hőnyi) Akadémiai Kiadó, Budapest, 1974

Time variations of height and gravity, Wichmann Verlag, Sammlung Wichmann Bd. 14. Karlsruhe, 1983. Akadémiai Kiadó, Budapest, 1983

Felsőgeodézia (Egyetemi jegyzet), Tankönyvkiadó, Budapest, 1985

A nehézségi erőtér időbeli változásának geodéziai hatása. Akadémiai Kiadó, Budapest, 1988

B) Scientific papers

A belső képállítólencse mozgatása által okozott irányvonalingadozások, *Geodézia és Kartográfia*, 8 (1956), 142–149.

A geoidundulációk meghatározása geometriai módszerrel, *ÉKME Tudományos Közleményei*, 7 (1961), 3–45.

A kvázigeoidról, *ÉKME Tudományos Közleményei*, 7 (1961), 105–129.

A normálmagasság gyakorlati meghatározása, *Geodézia és Kartográfia*, 14 (1962), 1–6, 84–89.

A tudományos geodézia újabb irányai, *Geodézia és Kartográfia*, 15 (1963), 42–47.

Eine modifizierte Lösung der Randwertaufgabe der geodätischen Gravimetrie auf die Oberfläche der Erde, *Studia Geophysica et Geodaetica*, 9 (1965), 119–123.

Berechnung der mittleren Koordinatenfehler der Hauptpunkte IV. Ordnung ausgeglichen mittels elektronischer Rechenanlange, *ÉKME Tudományos Közleményei*, 11 (1965), 3–15.

A geodéziai gravimetria időszerű kérdései, *Geodézia és Kartográfia*, 17 (1965), 178–185.

A geodéziai gravimetria alapfeladatának megoldása a fizikai földfelszínre, *ÉKME Tudományos Közleményei*, 12 (1966), 3–30.

Az alapfelület meghatározásának tisztán fizikai módszere, *Geodézia és Kartográfia*, 18 (1966), 256–262.

A mesterséges holdak geodéziai alkalmazásának dinamikai módszerei, *Geodézia és Kartográfia*, 19 (1967), 241–249.

On the accuracy of the deflection of the vertical interpolated by gravimetric methods, *Österreichische Zeitschrift für Vermessungswesen*, Sonderheft 25, (1967), 86–90.

Eötvös Loránd kutatásainak geodéziai jelentősége, *Magyar Geofizika*, 10 (1969), 177–179.

A Föld alakja az újabb kutatások tükrében, *MTA X. Osztály Közleménye*, 3 (1970), 295–309.

A dinamikai szatellitageodézia újabb eredményei. *Felsőfokú Földmérő Technikum Tudományos Közleménye*, 1 (1970), 9–13.

Die vertikalen Erdkrustenbewegungen und Sekularvariationen des Erdschwerefeldes, *Periodica Polytechnica CE*, 16 (1972), 4–9.

A geodéziai alapfelületek, *Geodézia és Kartográfia*, 24 (1972), 401–412.

Der einflußvon zeitlichen Änderungen des Erdschwerefeldes auf die Höhe von Nivellementsfestpunkten. *Mitteilungen aus dem Institut für Theoretische Geodäsie der Universität Bonn*, 12 (1973), 1–12.

A földi nehézségi erőtér tanulmányozása a mesterséges holdak megfigyelése és a földfelszíni gravitációs mérések eredményeinek együttes felhasználásával, *Geodézia és Kartográfia*, 25 (1973), 321–324.

The Earth's Shape as Shown by Modern Researches – Studies on the Material and Energy Flows of the Earth. MTA X. Osztály Kiadványa (1973), 332–335.

Gravimetric Research in Hungary (co-authors: M Polhammer, G Bartha) *Acta Geod. Geoph. Mont. Hung.*, 9 (1974), 177–182.

Über einige Probleme der Höhenbestimmung im zeitlich variablen Schwerefeld, *Periodica Polytechnica CE*, 19 (1975), 25–32.

A geodéziai tudomány helyzete hazánkban (co-authors: F Halmos, I Joó), *Geodézia és Kartográfia*, 28 (1976), 237–248, 334–341.

A felszínmozgások vizsgálata és a Föld dinamikai folyamatai, *Geodézia és Kartográfia*, 29 (1977), 9–18.

Ergänzende Bemerkungen zur zeitlichen Änderungen von Schwere und Höhe. *Zeitschrift für Vermessungswesen*, 102 (1977), 451.

Geodynamische Aspekte der Geodäsie, *Periodica Polytechnica CE*, 21 (1977), 3–15.

Zur Anwendung der Stokes'schen Formel für zeitliche Schwereänderungen, *Zeitschrift für Vermessungswesen*, 106 (1981), 523–531.

Dynamic aspects of repeated geodetic levellings and gravity observations, *Periodica Polytechnica CE*, 24 1-2, (1980), 3–12.

Geodynamic Aspects of Repeated Geodetic Levellings and Gravity Observations, *Acta Geod. Geoph. Mont. Hung.*, 16 (1981), 15–26.

Physikalische Interpretation der Meßergebnisse von wiederholten geodätischen Beobachtungen. *Vermessungstechnik*, 32 (1984), 48–49.

Modelling of secular variations in gravity and geoidal undulations (co-authors: N C Thong, E Weisz), *Periodica Polytechnica CE*, 30 (1986), 22–36.

Time variation in geopotential in spherical harmonics (co-author: N C Thong), *Acta Geod. Geoph. Mont. Hung.*, 21 (1986), 31–40.

Relative and true surface movement (co-author: N C Thong) *Journal of Geodynamics*, 8 (1987), 193–197.

On a new approach into the solution of the three-dimensional geodetic-geodynamic boundary value problem (co-authors: N C Thong, E Weisz), Report to the joint meeting of SSG 3.84, 3.86 and 3.87 IAG, Paris 1985. *Bulletin d'Information Bureau Gravimetrique International*, 57 (1985), 82–96.

A szatellita-geodéziai állomáskoordináták változása és a valódi felszínmozgások (co-author: B Heck), *Geodézia és Kartográfia*, 38 (1986), 238–243.

Physikalische Modelle der Erdmessung, Jahresbericht 1988, Deutsche Geodätische Kommission, München 1989, 242–249.

Zum Begriff "Schwere" und zu den SI Maßeinheiten, *Zeitschrift f. Vermessungswesen*, 114 (1989), 209–218.

A hazai földmérőmérnök képzés helyzete, *Geodézia és Kartográfia*, 41 (1989), 318–323.

Hazay István jelentősége a magyar földmérésben. *Geodézia és Kartográfia*, 43 (1991), 243–244.

On the representation of the earth's gravity field. *Allgemeine Vermessungs-Nachrichten*, International Edition '93, 10 (1993), 14–20.

Dr. Rédey István szerepe a geodézia oktatásában, kutatásában, *Geodézia és Kartográfia*, 51 (1999), Különszám, 12–17.

A földmérőmérnök-képzés 50 éve, *Geodézia és Kartográfia*, 51 (1999), 12, 5–10.

C) Scientific reports

A magassági rendellenességek meghatározása elméletének továbbfejlesztése, ÁFTH pályázat, 1964

On the needed density of gravimetric stations in geodetic nets. Report presented at the XIV-th General Assembly of IUGG, Lucern, 1967

A kéregmozgási szintezési hálózat gravitációs méréseinek egyes gyakorlati kérdései. OFTH kutatási feladat zárójelentése, 1968

Erdkrustenbewegungen und Säkularvariationen des Erdschwerefeldes. Report to the XIV-th General Assembly of IUGG, Moszkva, 1970

Dynamic aspects of repeated geodetic levellings. Report to the XVII-th General Assembly of IUGG, Canberra, 1979

Hungarian IAG Report for the General Assembly of IUGG Canberra, Australia, 1979. Section V: Research in Physical Interpretation, Sopron, 1979

Geodynamic interpretation of repeated gravity observations (co-author: L Völgyesi), Proceedings of the IV. International Symposium Geodesy and Physics of the Earth, Karl-Marx Stadt, 1980

Hungarian IAG Report for the XVIII-th General Assembly of IUGG Hamburg, 1983. Section V. Research in physical interpretations, Sopron, 1983. 30–34.

The role of gravimetry in the geodynamic interpretation of geodetic observations. Report presented at the XVIII-th General Assembly of IUGG, Hamburg, 1983. Proceedings of IAG Symposia, Hamburg, 1983. Ohio State University, Columbus, Ohio, 1984, 54–62.

A contribution to the network design of the International Absolute Gravity Base Station Network (co-authors: N C Thong, E Weisz) Paper for discussion for the SSG 386 IAG, Budapest, 1984

An estimation on the needed number of stations in the IAGBN (co-authors: N C Thong, E Weisz), Report to the joint meeting of SSG 3.84, 3.86 and 3.87 IAG, Paris 1985. Bulletin d' Information Bureau Gravimetrique International, 57 (1985), 82–96.

Time variations in the geometry and in the gravity field of the Earth, XIX-th General Assembly of IUGG Vancouver, Canada, Abstracts V.1 (Union, IAG, IASPEI), Vancouver, 1987, 252.

A 4 dimensional integrated model for geodynamic researches, Proceedings of the "International Symposium on Instrumentation, Theory and Analysis for Integrated Geodesy", Sopron (1988), 92–97.

What is gravity in fact? Report at the General Meeting of the International Association of Geodesy, Edinburgh, 1989. Gravity, Gradiometry and Gravimetry. Ed: R Rummel and R G Hipkin, IAG Symposia 103, Series Editor I I Mueller, Springer, New York 1990, 1–8.

A felsőgeodézia fizikai modelljei, "A földtudományok és a változó világ" továbbképző szeminárium előadásainak gyűjteménye, Sopron, 1990, 22–31.

D) Reviews

Újabb önműködő szintezőműszerek, *Geodézia és Kartográfia*, 10 (1958), 301–303.

A Drezdai Műszaki Egyetem II. Politechnikai Értekezlete, *Geodézia és Kartográfia*, 11 (1959), 143–144.

Új megoldások a magassági kör indexlibellájának kiküszöbölésére, *Geodézia és Kartográfia*, 12 (1960), 131.

Háromszögelés hanghullámokkal az óceánokon keresztül, *Geodézia és Kartográfia*, 12 (1960), 130.

Rakéták irányításának geodéziai feladatairól, *Geodézia és Kartográfia*, 12 (1960), 211–212.

A heerbruggi alapvonal, *Geodézia és Kartográfia*, 12 (1960), 292–295.

A Drezdai Műszaki Egyetem III. Politechnikai értekezlete, (co-author: T Lukács) *Geodézia és Kartográfia*, 13 (1961), 205–206.

A geodéta szakmérnökképzésről, *Geodézia és Kartográfia*, 17 (1965), 47–48.

Tanácskozás a Föld alakjának és méreteinek meghatározásáról Prágában, *Geodézia és Kartográfia*, 17 (1965), 51–53.

A Hold egyensúlyi alakja, *Geodézia és Kartográfia*, 18 (1966), 221–222.

"Földalak és refrakció" szimpózium Bécsben (co-author: L Homoródi) *Geodézia és Kartográfia*, 19 (1967), 299–300.

Az 1967. évi magyar geodéziai irodalom, *Geodézia és Kartográfia*, 20 (1968), 460–461.

BME Tudományos Ülésszaka. *Geodézia és Kartográfia*, 20 (1968), 68–69.

E) Others

Néhány korszerű eljárás a geodéziában, ÉKME előadó-hetek kiadványai (1958), 89–93.

Szputnyikokról geodéta szemmel, *Jövő Mérnöke*, 6 (1959).

Tudományos diákköri munka az ÉKME-n, *Jövő Mérnöke*, 8 (1961).

A hallgatók egyéni továbbképzése a tudományos diákkörökben, *Felsőoktatási Szemle*, (1962), 284–289.

A földmérőmérnöki szak, *Jövő Mérnöke*, 13 (1966).

ACTA GEODAETICA ET GEOPHYSICA HUNGARICA

A Quarterly Journal of the Hungarian Academy of Sciences

VOLUME 35

AKADÉMIAI KIADÓ, BUDAPEST
2000

CONTENTS

Number 2

Aeronomy and space physics

Electromagnetic studies

Seismology

History

Meeting of the IAG, Section I in Sopron, 1999

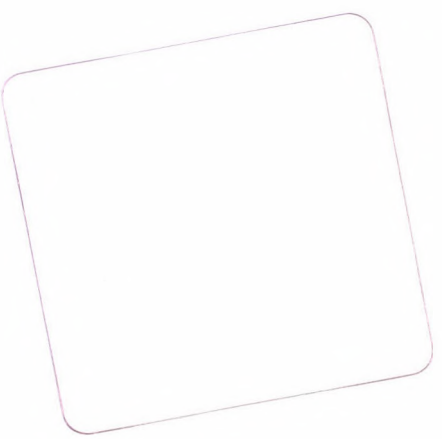
Book review

Number 3

Electromagnetic studies

AUTHOR INDEX

KEYWORD INDEX



INSTRUCTIONS TO AUTHORS

Manuscripts should be sent to the Editorial Office (address see above). Articles not published or submitted for publication elsewhere are only accepted.

Manuscripts should be preferably on disc (LATEX *.tex files or Word for Windows documents or text-only ASCII files) with a written copy, or via e-mail (actagg@ggki.hu). Files will be accepted, after previous agreement, via *anonymous ftp*, too. Typewritten manuscripts (in duplicate) are also accepted.

Manuscripts should include the following components which should be presented in the order listed (tables and illustrations should be separated in case of manuscripts on disc, too).

1. Title, name(s) of the author(s), affiliation, dateline, abstract, keywords
2. Text, acknowledgements
3. References
4. Footnotes
5. Legends
6. Tables
7. Illustrations

1. The *affiliation* should be as concise as possible and should include the complete mailing address of the authors. The *date of receipt* will be supplied by the editors. The *abstract* should not exceed 250 words and should clearly and simply summarize the most important methods and results. 5-10 significant expressions describing the content are used as *keywords*. Authors may recommend these keywords.

2. The *text* should be in English and as short and clear as possible. In case of typewritten manuscripts, please note the following:

- avoid possible confusion between o, O (letters) and 0 (zero), I (letter) and 1 (one), ν (Greek nu) and u, v (letters), etc.
- explain ambiguous and uncommon symbols by making marginal notes in pencil
- formulas are to be numbered consecutively with the number in parentheses to the right of the formula.
  Text references to the equations may then be made by the number in parenthesis. The word equation in this context is to be abbreviated to Eq. and Eqs in the plural
- the International System of Units (SI) should be used.

3. *References* are accepted only in the Harvard system. Citations in the text should be as:

... (Bomford 1971) ... or Bomford (1971) ...
... (Brosche and Sündermann 1976) ...
... (Gibbs et al. 1976b) ...

The list of references should contain names and initials of all authors (the abbreviation et al. is not accepted here); for journal articles, year of publication, title of the journal in abbreviated form, volume number, first and last page. For books or chapters in books, the title is followed by the publisher and place of publication.

All items must appear both in the text and in the list of references.

4. *Footnotes* should be typed on separate sheets.

5. *Legends* should be short and clear. The place of tables and figures should be indicated in the text on the margin.

6. *Tables* should be numbered serially with Roman numerals. Vertical lines are not used.

7. All the *illustrations* should contain the figure number and author's name in pencil on the reverse. The most important point with figures is clearness. Photographs and half-tone illustrations should be sharp and well contrasted. Colour photographs will be accepted, but the extra cost of reproduction in colour must be borne by the authors. Encapsulated postscript (EPS) files will be accepted on disk or via e-mail, too.

Only original papers will be published and a copy of the Publishing Agreement will be sent to the authors of papers accepted for publication. Manuscripts will be processed only after receiving the signed copy of the agreement. Information is sent to the first author if no other wish is expressed.