Acta Geodaetica, Geophysica et Montanistica Hungarica

VOLUME 26, NUMBERS 1–4, 1991

EDITOR-IN-CHIEF
J SOMOGYI

EDITOR
J VERŐ

EDITORIAL BOARD
A ÁDÁM, J ÁDÁM, A BAK, P BIRÓ, S DOLESCHALL, F KOVÁCS, A MESKÓ, A SCHMIEDER, F STEINER

Akadémiai Kiadó, Budapest
AGGM 26 (1–4) 1–481 (1991) HU ISSN 0236-5758

ACTA GEODAETICA, GEOPHYSICA et MONTANISTICA HUNGARICA
A Quarterly Journal of the Hungarian Academy of Sciences

Acta Geodaetica, Geophysica et Montanistica (AGGM) publishes original reports on geodesy, geophysics and minings in English.

AGGM is published in yearly volumes of four numbers by

AKADÉMIAI KIADÓ
Publishing House of the Hungarian Academy of Sciences
H-1117 Budapest, Prielle K. u. 19–35.

Manuscripts and editorial correspondence should be addressed to

AGGM Editorial Office
H-9401 Sopron P.O. Box 5

Subscription information

Orders should be addressed to

KULTURA Foreign Trading Company
H-1389 Budapest P.O. Box 149

INSTRUCTIONS TO AUTHORS

Manuscripts should be sent to the editors (MTA Geodéziai és Geofizikai Kutató Intézete, AGGM Editorial Office, H-9401 Sopron, P.O.Box 5, HUNGARY). Only articles not submitted for publication elsewhere are accepted.

Manuscripts should be typewritten in duplicate, double-spaced, 25 lines with 50 letters each. The papers generally include the following components, which should be presented in the order listed.

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

1. The *affiliation* should be as concise as possible and should include the complete mailing address of the authors. The *date of receipt* of the manuscript 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 generally in English and as short and clear as possible. From Hungarian authors papers are also accepted in Hungarian.

The section heading should *not* be underlined or in capitals.

Please note that underlining denotes special types:

— single underlining: italics
— double underlining: bold-face roman

ACTA GEODAETICA, GEOPHYSICA et MONTANISTICA
A Quarterly Journal of the Hungarian Academy of Sciences

EDITOR-IN-CHIEF
J SOMOGYI

EDITOR
J VERŐ

EDITORIAL BOARD
A ÁDÁM, J ÁDÁM, A BAK, P BIRÓ, S DOLESCHALL,
F KOVÁCS, A MESKÓ, A SCHMIEDER, F STEINER

VOLUME 26

AKADÉMIAI KIADÓ, BUDAPEST
1991

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 3-8 (1991)

SATELLITE GEODETIC STUDIES ON THE REGIONAL DYNAMICS OF THE CARPATHO-BALKAN AREA

I N Totomanov and N I Georgiev

Central Laboratory for Geodesy of the Bulgarian Academy of Sciences,
1000 Sofia, 7 Noemvri str. 1, Bulgaria

[Manuscript received October 8, 1987, revised: November 2, 1989]

The present paper expands the authors' investigations on the optimum laser location on artificial Earth's satellites (AES) for the determination of the absolute and relative kinematics of the regional lithosphere with emphasis on the Carpatho-Balkan Region (CBR), on its two main structural parts - the Middle European and Balkan ones. In earlier CBR studies, the accuracy of the Cartesian components of the geocentric vectors was investigated for the Balkan and the intercontinental chord vectors were independently considered for Middle Europe and for the Balkan. In this model study new results are presented for the accuracy of the determination of: a) the components of the geocentric vectors in Middle Europe, and b) the chord vectors for the entire CBR. Conclusions are drawn on the optimum AES orbit parameters, on the quality of the network of observatories and on the common and differring characteristics of CBR and its parts as related to the international project IDEAL.

Keywords: Carpatho-Balkan area; geodynamics; laser location; satellite geodesy

INTRODUCTION

Earlier model studies (Georgiev et al. 1987, 1988, Totomanov et al. 1987) aimed at the determination of the laser location of artificial Earth's satellites (AES) as related to the international geodynamic project IDEAL (Georgiev et al. 1985, 1986, Georgiev and Totomanov 1986) and were made both for the entire Carpatho-Balkan region (CBR) and for its two main parts: the Middle-European and the Balkan ones; results were presented for the satellite geodetic observatories Penc near Budapest (Hungary) and Plana near Sofia (Bulgaria). This paper continues this series and outlines the results of a report

(Totomanov and Georgiev 1987) at the conference "Structural development of the Carpatho-Balkan orogenic belt" (Bratislava, Czechoslovakia, October 12-15, 1987). 32 various computer-generated laser measurements are used which have an integral accuracy of 25 cm, and are between the terrestrial observatories and an hypothetical AES with a major semiaxis a (or height H = a-R) and an inclination i of the orbit.

ACCURACY OF GEOCENTRIC POSITIONING STUDIES

In earlier studies the root-mean square error m_s of the lengths of the CBR geocentric vectors $\vec{S}$ is used to measure the accuracy of their determination. In the most recent investigations related to the Balkan (Totomanov et al. 1987), the errors m_x, m_y and m_z of the Cartesian components of $\vec{S}$ are determined; they show that m_s is an approximate informative quantity. Thus m_x, m_y and m_z are calculated and mapped in the present work for the Middle-European part of the CBR (Fig. 1), too; here the laser location of an AES ensures the accuracy required by IDEAL, $m \leqslant 3.5$ cm in the hatched areas.

Even in this case a considerable difference is seen in the accuracy of the determination of the components of S the accuracy is the least (3-11 cm) for Y, the highest (2-6 cm) for Z, and an intermediate one (3-7 cm) for X. The optimum areas E are for the determination of a component of $\vec{S}$ with the condition given for m for Middle Europe:

$E_{ME}(m_x) \in (50 < i < 80^o,\ H > 4$ Mm; $80 < i < 90^o,\ 5 < H < 7$ Mm;
$100 < i < 110^o,\ H > 6$ Mm),
$E_{ME}(m_y) \in (50 < i < 60^o,\ H > 5$ Mm; $60 < i < 100^o,\ 5 < H < 6$ Mm;
$100 < i < 120^o,\ H > 6$ Mm),
$E_{ME}(m_z) \in (50 < i < 120^o,\ H > 4$ Mm),

or jointly

$E_{ME} = E_{ME}(m_x \wedge m_y \wedge m_z) \in (50 < i < 60^o,\ H\ \ 5 > $Mm; $60 < i < 100^o,$
$5 < H < 6$ Mm)

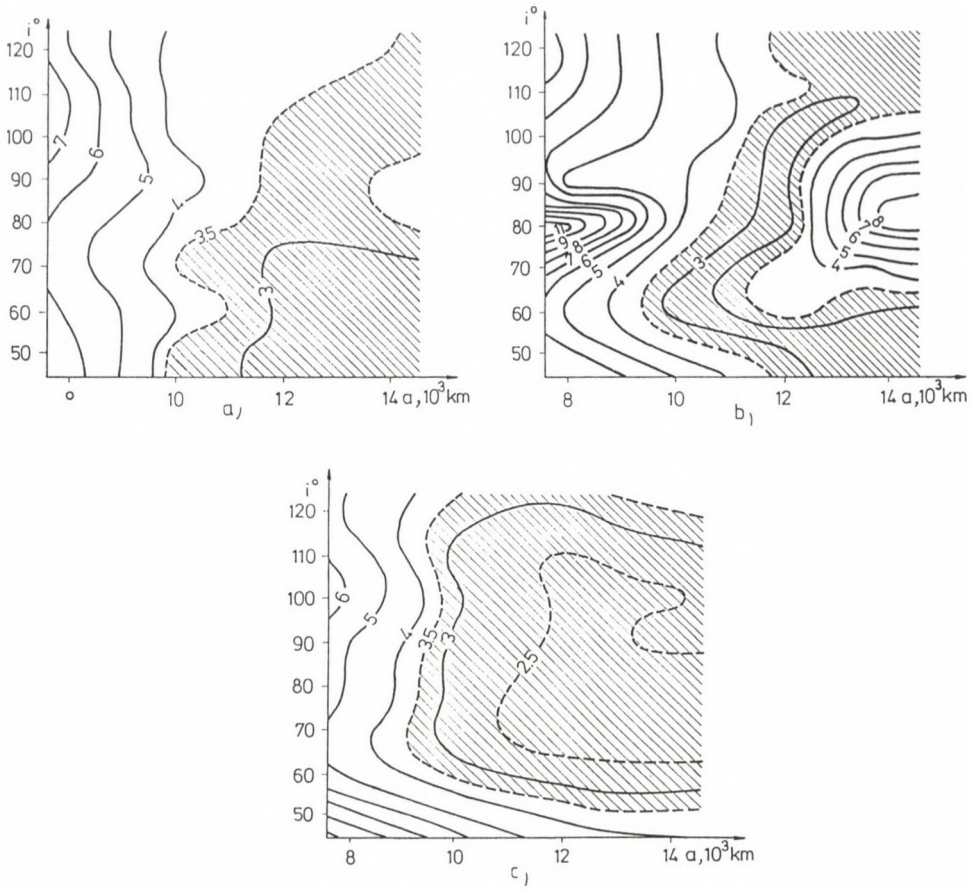


Fig. 1. Isolines of the root-mean square errors m of the Cartesian components of the geocentric positioning vectors $\bar{S}$ in Middle Europe: a - for m_x (cm), b - for m_y (cm), c - for m_z (cm). For details see the text

which are in accordance with the values given by Georgiev et al. (1988), through m_s, respective the area E.

Dirac's hypothesis on the expanding Earth can be only checked by means of the component Z within the area E_{ME} ($m_z \approx$ ≈ 2 cm) $\in (70 < i < 90^o$, $H > 5$ Mm).

A comparison of the new results E_{ME} obtained here for Middle Europe, with those for the Balkan region E_{BR} (Totomanov et al. 1987):

$E_{BR}(m_Z) \in (50 < i < 120^o,\ H > 3\ Mm)$,
$E_{BR} = E_{BR}(m_X\ m_Y\ m_Z) \in (50 < i < 60^o,\ 110 < i < 120^o,$
$H > 7\ Mm;\ 60 < i < 80^o,\ 5 < H < 6\ Mm;\ 100 < i < 110^o,$
$6 < H < 7\ Mm)$

shows the specific points of the two main CBR parts for the positioning of an AES laser to be considered.

ACCURACY OF INTERCONTINENTAL GEOKINEMATIC STUDIES

The root-mean square errors m_R, m_ϕ and m_L were calculated for the length R and for the orientations (ϕ by latitude, and L by longitude) of two groups of chord vectors $\vec{R}$ between observatories located in the stable parts of the African and Eurasian plates (Helwan in Egypt and Zvenigorod in the USSR, respectively), on the one hand, and the two parts of the mobile CBR, on the other hand. The results are plotted in Fig. 2 where the optimum areas E for the determination of $\vec{R}$ are hatched ($m \leq 3.5$ cm). This figure enables the following conclusions in addition to earlier ones (Georgiev et al. 1986b, 1987, 1988, Totomanov et al. 1987).

The determination of ϕ , by means of the location of a very large class of admissible AES, namely

$E(\phi)$ $(60 < i < 120^o,\ 4 < H < 6\ Mm;\ 60 < i < 70^o,$
$100 < i < 120^o,\ H > 6\ Mm)$

is characterized by the greatest reliability. When determining R, the requirements to the orbit are considerably stronger, the two parts of CBR differ significantly and the network of observatories within Middle Europe has a more favourable configuration. A need is indicated for the reconstruction of and the addition of new observatories be they of the stationary or of the mobile type to the network, because of the Balkan part. The optimum area E is for the determination of R

$E(R) \in (50 < i < 70^o,\ 100 < i < 110^o,\ 6 < H < 8\ Mm,$
$70 < i < 100^o,\ 5 < H < 6\ Mm).$

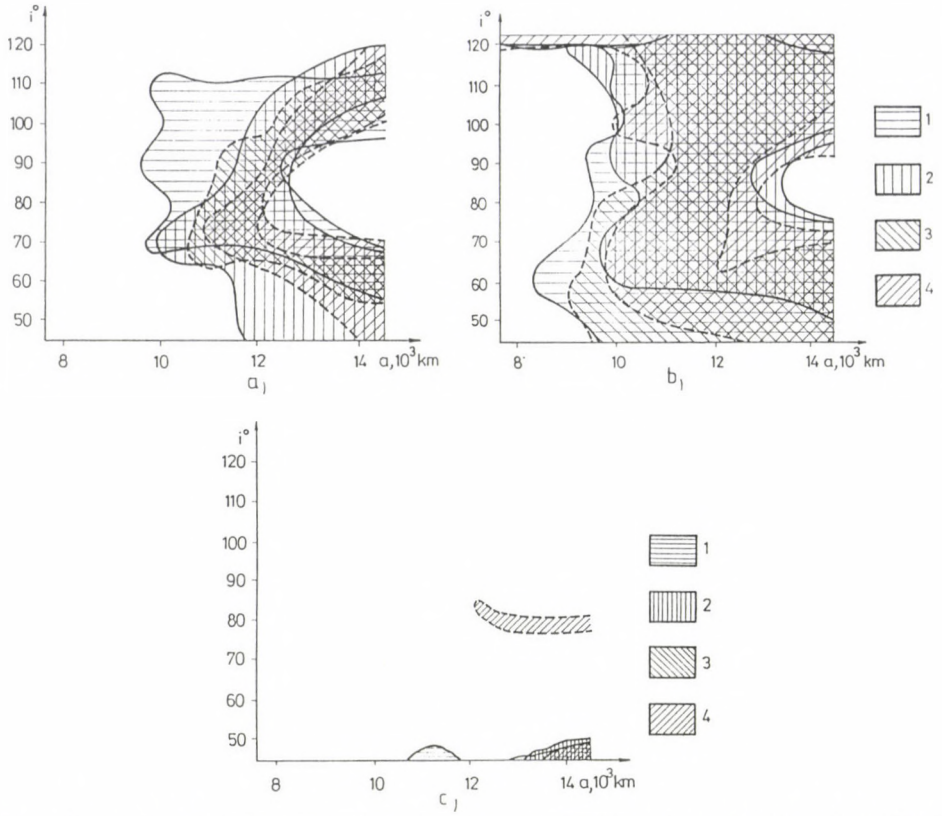


Fig. 2. Areas can be determined with a root-mean square error less than 3.5 cm where the polar components (a - for the length; for the orientation: b - by latitude, and c - by longitude) of the chord geokinematic vectors $\vec{R}$ between the lithospheric plates and CBR: Helwan-Penc (1), Zvenigorod-Penc (2), Helwan-Plana (3), Zvenigorod-Plana (4). For details see the text

The estimation of L completely differs from those of R and ϕ, by a new difference (in this case - according to the link of the vectors $\vec{R}$ to the lithospheric plates mentioned) in the two CBR parts. Nevertheless, a minimum area common for the four vectors $\vec{R}$, is represented in Fig. 2b:

$$E(L) \in (40 < i < 50^{o},\ 7 < H < 9\ \text{Mm}), \tag{8}$$

the realization of this requirement, however, will create additional difficulties for the project IDEAL.

In conclusion, it should be emphasized that the results obtained in the present study, and the inferences derived from them, aim to help the correct planning and the forthcoming realization of the laser location of AES within the entire CBR and its main structural parts.

REFERENCES

Georgiev N, Totomanov I 1986: CSTG Bulletin, No 9: Activities of CSTG subcommissions and projects, 85-87.

Georgiev N, Tomomanov I, Hadjijski A 1985: CSTG Bulletin, No 8: Future missions, systems and projects, 175-181.

Georgiev N, Totomanov I, Hadjijski A 1986a: Visša geodezija (Geodesy), No 12, 3-10.

Georgiev N, Totomanov I, Hadjijski A 1986b: Geodezia, Kartografiya, Zemeustroistvo, 26, No. 6, 3-6.

Georgiev N, Totomanov I, Hadjijski A 1987: In: Results of cooperation between the Central Laborabory for Geodesy at the Bulgarian Academy of Sciences in Sofia and the Geodetic and Geophysical Research Institute at the Hungarian Academy of Sciences in Sopron. Bulgarian Academy of Sciences, Sofia, 73-105.

Georgiev N , Totomanov I, Hadjijski A 1988: Journal of Geodynamics, 9, 97-110.

Totomanov I, Georgiev N 1987: In: Strukturny vyvoj Karpatsko-Balkanskeho orogenneho pasma. Československa vedecko-technicka spločnost, Bratislava, 158.

Totomanov I, Georgiev N, Hadjijski A 1987: Optimum parameters of the satellite geodetic network of project IDEAL for investigation of the Balkan regional geodynamics. In: XIX. Tagung der KAPG - Symposium A "Struktur der Lithosphäre" (GDR, Neubrandenburg)

ACTA GEODAETICA, GEOPHYSICA et MONTANISTICA
A Quarterly Journal of the Hungarian Academy of Sciences

EDITOR-IN-CHIEF
J SOMOGYI

EDITOR
J VERŐ

EDITORIAL BOARD
A ÁDÁM, J ÁDÁM, A BAK, P BIRÓ, S DOLESCHALL,
F KOVÁCS, A MESKÓ, A SCHMIEDER, F STEINER

VOLUME 26

AKADÉMIAI KIADÓ, BUDAPEST
1991

CONTENTS

AUTHOR INDEX

KEYWORD INDEX

Acta Geod. Geoph. Mont. Hung, Vol. 26 (1-4), pp. 9-17 (1991)

GRAVIMETRIC GEOID COMPUTATION FOR CANADA

D Nagy

Geological Survey of Canada
1 Observatory Crescent, Ottawa, Canada

Geological Survey of Canada Contribution No. 46989

[Manuscript received February 27, 1990]

Recently a gravimetric geoid (GEOID '88) has been calculated for Canada using 125 530 block averages covering the entire surface of the Earth. The computation for 6 398 points has been completed on a Cray 1-S supercomputer. Some details of data preparation and computations are given. Histograms for three selected points, which were also included in a geoid computation in 1972 using less gravity information, and the contour map of the calculated geoid are shown.

Keywords: Canada; FFT; gravimetric geoid; Stokes's formula

INTRODUCTION

Geoid computations follow a cyclic pattern. The first wave of gravimetric geoid computation was started by Hirvonen (1934) who, after discussing the theory, carried out a gravimetric geoid computation, with an estimation of the error of the computed result. Then, in the early 60's a new wave of computation began, which was tied to predicting the gravity anomalies over unsurveyed regions of the earth. In the next decade, the improvements in satellite tracking and better computational facilities for high degree spherical harmonic expansion of surface gravity data provided the foundation for a renewed interest in geoid computations. Recently, additional requirements from other disciplines (such as Global Positioning System, possible replacements of leveling, etc.) have given way to yet another set of geoid computations. After a brief discussion of theory, a summary of recent involvements with geoid computations is given.

THEORY

The gravimetric geoid in a spherical polar coordinate system is usually calculated from a spherical approximation, as derived by Stokes (1849):

$$N = \frac{R}{4\pi\gamma} \int_{\sigma} \Delta g \, S(\psi) \, d\sigma, \tag{1}$$

where N is the geoidal height,
R is the mean radius of the earth,
γ is the mean gravity,
Δg is the gravity anomaly, corresponding to $d\sigma$,
$S(\psi)$ is the Stokes function,
ψ is the spherical distance,
and $d\sigma$ is the surface element of the unit sphere.

Stokes' function is usually given in the following form:

$$S(\psi) = \operatorname{cosec} \tfrac{1}{2}\psi + 1 - 6 \sin \tfrac{1}{2}\psi - 5 \cos\psi - 3 \cos\psi \ln \left(\sin \tfrac{1}{2}\psi + \sin^2 \tfrac{1}{2}\psi \right).$$

Thus, the geoidal height can be calculated by carrying out the integration shown in Eq. (1). In practice the integration is replaced by double summation. The method obviously requires gravity everywhere over the surface of the Earth. The surface element, $d\sigma$, on the sphere is:

$$d\sigma = \sin\psi \, d\psi \, d\alpha,$$

where α is the azimuth measured from the north. Substituting $d\sigma$ into Eq. (1), N can be written as:

$$N = \frac{R}{4\pi\gamma} \int_{\alpha=0}^{2\pi} \int_{\psi=0}^{\pi} \Delta g \, F(\psi) \, d\psi \, d\alpha. \tag{2}$$

In Eq. (2), F is the weight which must be applied to gravity in order to obtain the contribution to geoidal height. The function F, defined as:

$$2F(\psi) = S(\psi) \sin\psi,$$

is shown in Fig. 1. It can be seen that even gravity anomalies far from the computation point, cannot be ignored and, in order to obtain meaningful geoidal heights using Eq. (2), the integration must be carried out to $\psi = \pi$.

When astro-geodetic deflections are available for an area, but not in sufficient detail for numerical integration to obtain relative geoidal heights, then deflections calculated from gravity can be used for densification. The starting point in this case is the formula derived by Vening-Meinesz (1928):

$$\xi = \frac{R}{4\pi\gamma} \int_{\sigma} \Delta g \, VM(\psi) \cos\alpha \, d\sigma, \tag{3}$$

$$\eta = \frac{R}{4\pi\gamma} \int_{\sigma} \Delta g \, VM(\psi) \sin\alpha \, d\sigma, \tag{4}$$

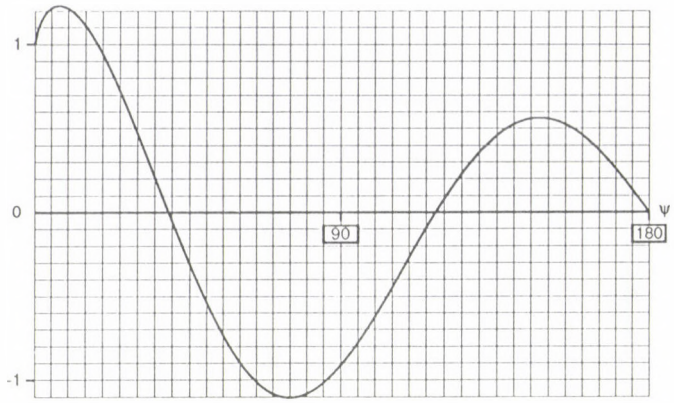


Fig. 1. The plot of F as the function of the spherical distance: ψ

where

$$VM(\psi) \equiv \frac{dS(\psi)}{d\psi} = -\tfrac{1}{2}\cos\tfrac{1}{2}\psi \cdot \operatorname{cosec}^2\tfrac{1}{2}\psi + 8\sin\psi - 6\cos\tfrac{1}{2}\psi$$
$$-3(1-\sin\tfrac{1}{2}\psi)\cdot\operatorname{cosec}\psi + 3\sin\psi\ln(\sin\tfrac{1}{2}\psi + \sin^2\tfrac{1}{2}\psi).$$

From Eqs. (3) and (4) computational formulae for the plane were derived for example, in Nagy (1963).

With the introduction of signal processing techniques for gravity interpretation and for physical geodesy, the Fourier transform became a useful tool. Developed by Tsuboi (1937) and later extended and used in Canada by Shimazu (1962), its wider application awaited the appearance of the Fast Fourier Transform (FFT) algorithm published by Cooley and Tukey (1965). The equivalence of the convolution integrals of both the Stokes and Vening-Meinesz forms and their Fourier transform and other details can be obtained from Nagy (1988a,1988b), Nagy and Fury (1988,1990), and Tsuboi and Fuchida (1937).

The geoid, for the plane, in the form of convolution integral is given as:

$$N(x,y) = \frac{1}{2\pi\gamma}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\sqrt{(x-x_p)^2+(y-y_p)^2}}\Delta g(x_p,y_p)\,dx\,dy.$$

Taking the Fourier transform of the kernel and the data (i.e. gravity anomaly), then the inverse Fourier transform provides the solution:

$$N = \frac{1}{2\pi\gamma}\,\Phi^{-1}\left[\frac{1}{u^2+v^2}\cdot\Delta G(u,v)\right], \tag{5}$$

where

$$\frac{1}{u^2+v^2} = \Phi\left[\frac{1}{\sqrt{(x-x_p)^2+(y-y_p)^2}}\right],$$

$$\text{and}\quad \Delta G(u,v) = \Phi\left[\Delta g(x_p,y_p)\right].$$

Here Φ and Φ^{-1} stand for the direct and inverse Fourier transform respectively.

The meridian and prime vertical components of the deflections of the vertical, (ξ,η), can be calculated from the expressions:

$$\xi = -\frac{1}{\gamma}\Phi^{-1}\left[\frac{iv}{\sqrt{u^2+v^2}}\cdot\Delta G(u,v)\right], \tag{6}$$

$$\eta = -\frac{1}{\gamma}\Phi^{-1}\left[\frac{iu}{\sqrt{u^2+v^2}}\cdot\Delta G(u,v)\right]. \tag{7}$$

Although the FFT technique provides an easy and efficient way of carrying out the necessary numerical computations, there are a number of characteristics that are particular to this technique (periodicity, leakage, sampling interval, etc.), which require special attention. To study some of these questions in detail, modeling can be used. For the model, the required computations can be generated both directly from analytical expressions and numerically by using FFT. The detailed investigations by the comparison of the results obtained from modeling provides the necessary understanding needed for the proper use of the FFT technique. For details and useful expressions see for example Nagy (1966,1980,1988a).

COMPUTATIONS

In this section, the numerical computations relevant to the use of gravity in geoid determination (including deflections of the vertical computations) are briefly summarized.

Gravimetric deflections

Nagy (1963) was possibly the first to implement the calculation of the gravimetric deflections of the vertical by digital computer. The computation has been carried out for an area of size 1 200×1 200 km. Where possible, the gravity values over unit areas (50×50 km) were represented by integral means, obtained from

second order surfaces (in two variables) fitted to the irregularly distributed gravity data by the methods of least squares. The computation was carried out at grid points for three overlapping areas, each with the origin placed at an astro station. The weighting function for the plane was derived from the Vening-Meinesz formula. For the purpose of comparison with astro-geodetic deflections available for the area, the necessary coordinate transformations were also programmed (from Clark's spheroid to the International Ellipsoid). A visual comparison, especially for the centres of regions, showed reasonable agreement.

Geoid 72

The first gravimetric geoid for Canada was computed in the early seventies by Nagy and Paul (1973). By that time, the gravity coverage had increased substantially, although there still existed large gaps around Hudson Bay and particularly in the western part of Canada. The Geodetic Reference System 1967 was used. For Canada, and a 3° belt at the Canada-U.S. border, $30' \times 30'$ block averages were calculated (from 111 997 points in Canada and from 27 976 points in the U.S.A). Large part of the United States was covered by 1° mean values. For the rest of the world, 5° block averages were used. This totaled up to 7 497 block averages, which served as input to calculate one geoidal height. The geoid was calculated at 1 050 points over Canada. Apart from the vertical shift (due to the use of a different reference ellipsoid) all the major features, as we know it today, are on this map (Fig. 2).

It is interesting to examine how the geoid builds up as a function of the spherical distance measured from the computation point to the contributing elementary block. This was examined in some detail in Nagy (1975). The results of such computations are presented in the form of histograms. Some of these histograms (Nos 12, 16 and 20 of Fig. 3 in Nagy (1975)) may be compared with more complete determinations (discussed below; locations are shown in Fig. 3).

Geoid '88

This computation is based upon Eq. (1). One of the major tasks has been to assemble the input required for the computation. Since the last geoid computation in 1972, the global gravity coverage and its quality is greatly improved. However, there are still large regions where gravity is poorly known.

The input file consisted of three parts. Over Canada and the continental U.S.A. 65 539 $15' \times 15'$ block averages were calculated (Nagy 1988c,1988d,1989a). For the world, 44 20 blocks of the $1° \times 1°$ data bank of the Ohio State University was used (Despotakis 1986). For the remaining areas, the GEM-T1 satellite model of NASA was used to compute 15 608 input values (Marsh et al. 1987). In total these 125 530 block values provided the input for each of the 6 398 computation points where the geoid was computed. The computation has been carried out on the Cray 1-S supercomputer requiring about 8.07 hours of CPU time (Nagy 1988b).

Naturally, the quality of gravity coverage affects the accuracy of every cal-

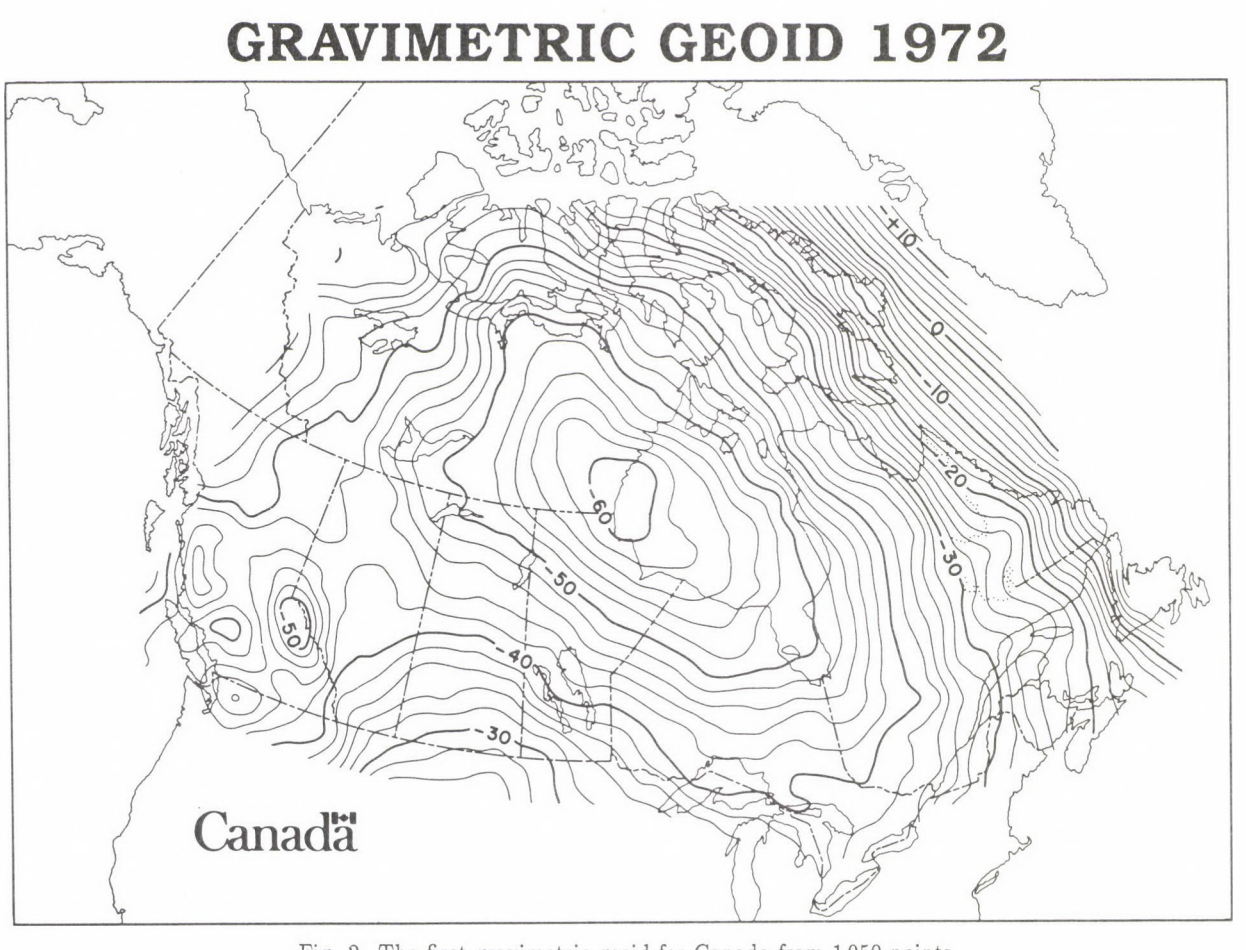


Fig. 2. The first gravimetric geoid for Canada from 1 050 points based on world-wide gravity data available in 1972

culated geoidal height and, thus, study of the error due to lack of coverage or its accuracy is important. To provide some insight into the mechanics of computation, three histograms show (Fig. 3) how the geoid contributions build up as a function of the spherical distance ψ. It is clear from the histograms, that no simple rule is available to obtain a *rough* estimate of geoidal height by using only *local* gravity data. The partial histograms obtained for the same stations from the 1972 computations (Nagy 1975 Fig. 3) are superimposed on the diagrams and are shown by heavy vertical bars. The scaling, quite arbitrarily, was done in such a way that the match was established at the last entry of the 1972 computed values. Apart from the uncertainty of scaling, the difference indicates the progress due to additional information about the gravity field both locally and globally.

A contour map representation of the geoid is shown in Fig. 4. The map is also printed as a part of the Geophysical Atlas Series of the Geological Survey of Canada (Nagy 1989a).

Throughout the years many computational procedures have been developed and refined for geoid computation, but the major obstacle, i.e. sufficient gravity coverage of the whole earth, still limits the determination of a high accuracy geoid.

REFERENCES

Cooley J W, Tukey J W 1965: Mathematics of Computation, **19**, 297-301.

Despotakis V K 1986: *The Development of the June 1986* 1° × 1° *and the August 1986*

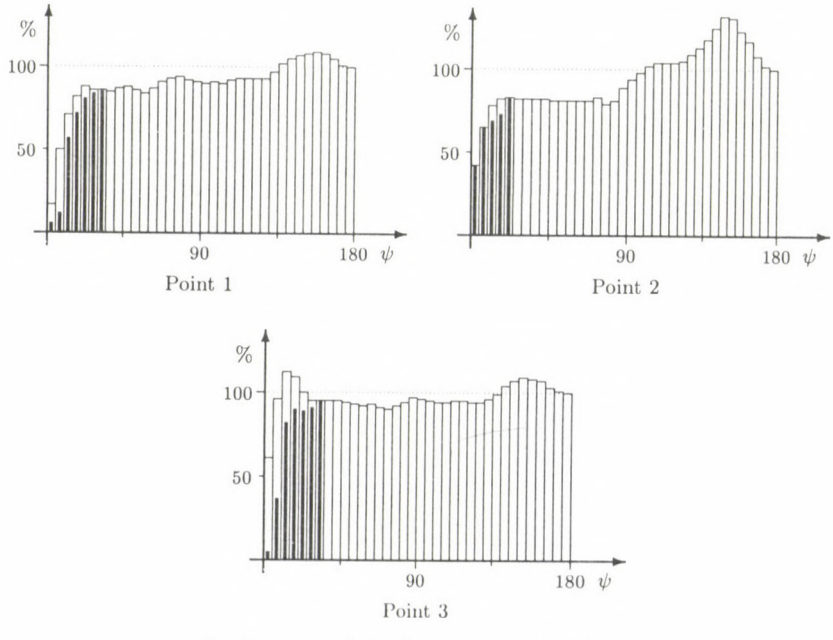


Fig. 3. Accumulative histogram of geoidal height.
For location of points see Fig. 4

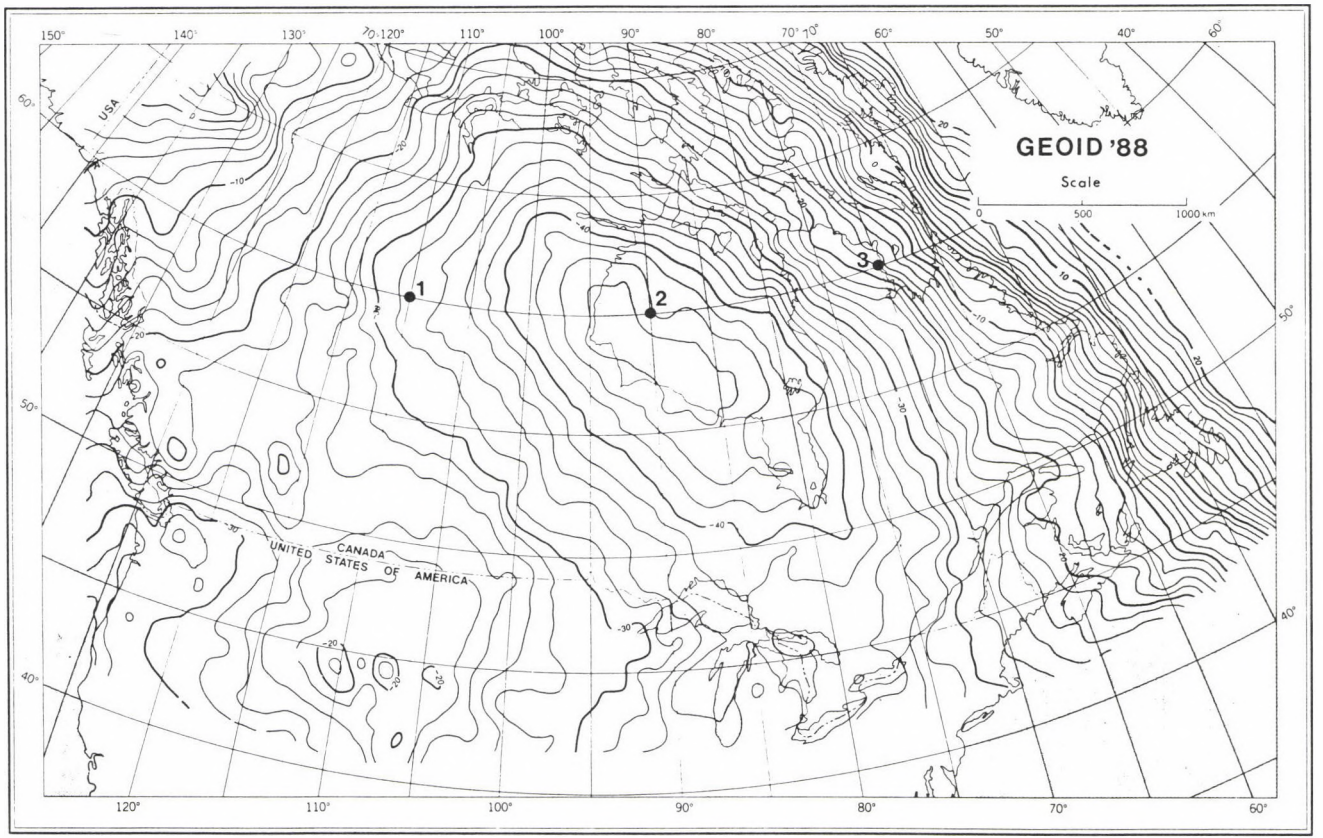


Fig. 4. Gravimetric geoid for Canada, calulated from world-wide gravity data at 6 398 points. Contour interval is 2 m

$30' \times 30'$ *Terrestrial Mean Free-Air Anomaly Data Bases.* Internal Report, Dept. of Geodetic Science and Surveying, The Ohio State University, Columbus

Hirvonen R A 1934: *The continental undulations of the geoid.* Suomen Geod. Laitaksen Julkaisuja, No. 19, Helsinki

Marsh J G, Lerch F J, Putney B H, Christodoulidis D C, Felsentreger T L, Sanchez B V, Smith D E, Klosko S M, Martin T V, Pavlis E C, Robbins S M, Williamson R G, Colombo O L, Chandler N L, Rachlin K E, Patel G B, Bhati S, Chinn D S 1987: *An Improved Model of the Earth's Gravitational Field :* ***GEM-T1*** NASA Technical Memorandum 4019.

Nagy D 1963: Publ. Dom. Obs. Ottawa, Canada.**27**, No. 1, 1–72.

Nagy D 1966: Geophysics, **31**, 362–371.

Nagy D 1975: Acta Geod. Geoph. et Mont. Hung., 10, 321-328.

Nagy D 1980: Acta Geod. Geoph. et Mont. Hung., 15 17–26.

Nagy D 1988a: Bolletino di Geodesia e Scienze Affini No.1, 33–43.

Nagy D 1988b: *Performance comparison of the CYBER 730 and the Cray 1-S Supercomputer.* Proceedings of the Supercomputing Symposium '88, University of Alberta, Edmonton, 267–273.

Nagy D 1988c: GEOID '88: *A gravimetric geoid for Canada.* Proceedings of the Chapman Conference on Progress in the Determination of the Earth's Gravity Field, Fort Lauderdale, 117–120.

Nagy D 1988d: *Gravity Field Representation Over Canada for Geoid Computation.* Proceedings of the Colloquium on Statistical Applications in the Earth Sciences, Geological Survey of Canada, Ottawa, 1988. Contribution No. 10989 (in press)

Nagy D 1989a: GEOID '88: *Gravimetric Geoid Map of Canada.* Geological Survey of Canada. Canadian Geophysical Atlas, Map 7.

Nagy D, Fury R J 1988: *FFT—Local Gravimetric Geoid Computation.* Proceedings of the Chapman Conference on Progress in the Determination of the Earth's Gravity Field. Fort Lauderdale, 92–95.

Nagy D, Fury R J 1990: *Local Geoid Computation from Gravity Using the Fast Fourier Technique.* Bulletin Géodésique (in press)

Nagy D, Paul M K 1973: *Gravimetric geoid of Canada.* Proceedings of the Symposium on the Earth's Gravitational Field and Secular Variations in Position. Sydney, Australia, p. 188.

Shimazu Y 1962: Publ. Dom. Obs. Ottawa, Canada. **26**, 323–354.

Stokes G G 1849: *On the variation of gravity on the surface of the earth.* Trans. Cambridge Phil. Soc., 8, 672–695.

Tsuboi C 1937: Earthquake Res. Inst. Bull. Tokyo, 15,650–653.

Tsuboi C, Fuchida T 1937: Earthquake Res. Inst. Bull., Tokyo, 15, 636–649.

Vening-Meinesz, F A 1928: Proc. Köninkl. Ned. Akad. Wetenschap., Amsterdam, 31, 315–331.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 19-28 (1991)

INTRODUCTION TO BÉZIER CURVES*

D Nagy
Geological Survey of Canada
1 Observatory Crescent, Ottawa, Canada

Geological Survey of Canada Contribution No. 46889

[Manuscript received February 27, 1990]

In many branches of science and engineering, the mathematical representation of curves and surfaces plays an increasingly important role. The widespread availability of personal computers (PC-s) with good graphic capabilities (resolution of up to 1024×768) now allows the development of interactive graphic design in a very simple environment. One of the most useful techniques for these developments is based on the use of the Bézier curves, which will be introduced.A comprehensive list of references is also provided.

Keywords: Bernstein; Bézier; METAFONT; PostScript; TEX

INTRODUCTION

In most geodetic and geophysical applications, the graphic representation of data (in two dimensions) requires:

- a simple mathematical form to *approximate* the given data points, which contain errors. This in many cases leads to the use of the method of least squares. This procedure selects some simple basis functions (in many cases polynomials) and constructs from them a model function to be used for interpolation. This function, in general, will not pass through any of the data points!
- a mathematical representation to data points such that the input is *reproduced*, i.e. the function satisfies all data points! This requirements leads to spline type of an approximation.

In many other applications such as Computer Aided Manufacturing (CAM), Computer Aided Design (CAD), typography, generalization in mappings, etc. the

*Dedicated to the 100th birthday of Professor Albert Stasney

purpose is quite different. In these cases, curves are sought which express mathematically what the designer has in mind. These curves should be *smooth* in the definition of the user. They also should be simple to develop, i.e. the designer should anticipate the result of changes in the shape of the curve. Today, with the availability of PC-s, good graphic boards (EGA, VGA, superVGA, etc.) and suitable programming languages (such as C), interactive design is a must. The use of Bézier curves in large part fulfills these requirements.

BÉZIER CURVES

The technique to be introduced was developed in the early 1960s mostly by people involved in automobile manufacturing for the purpose of defining various shapes of automobiles, which were then machined by computer driven software. Because of secrecy, little of the details of the technique became available. Although Bézier may not have been the first one to use the curves named now after him, it was his contribution which was widely published and serves as the standard reference on the subject.

THEORY

Before discussing the mathematical details, some basic properties of the Bézier curve will be summarized:

- it requires only data points (no derivatives are needed),
- it passes only through the end points,
- it has tangency at the end points (discussed later),
- for $t \in [0, 1]$, it always remains within the polygon formed by the data points,
- the degree can be varied from segment to segment,
- the segments can be joined with various degree of continuity,
- it is suitable for interactive development,
- extension to higher dimensions is simple,
- implementation for PCs is easy.

To simplify the discussion, only two-dimensional cubic Bézier curves defined over four points will be treated here. To use higher degree curves or to join the piecewise approximations, the reader is referred to the list of references.

The cubic Bézier curve is based on the Bernstein (1912) cubic basis functions. In vector notation such a Bézier curve can be written as:

$$\mathbf{r}(t) = (1-t)^3\mathbf{r}_0 + 3t(1-t)^2\mathbf{r}_1 + 3t^2(1-t)\mathbf{r}_2 + t^3\mathbf{r}_3, \tag{1}$$

where $\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ are the position vectors which form a tetrahedron, such that $\mathbf{r}$, for $t \in [0,1]$, lies within the same tetrahedron. Known as the *convex hull property* of the Bézier curves, this is a most useful property, where bounds in various computations, such as hidden line removal, are required.

Expanding the terms, for example for the coefficient of $\mathbf{r}_0$ one finds:

$$(1-t)^3 = 1 - 3t + 3t^2 - t^3.$$

Thus Eq. (1) can be expressed in matrix form:

$$\mathbf{r} = \begin{bmatrix} 1 & t & t^2 & t^3 \end{bmatrix} \mathbf{A} \begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix}^T$$

where T designates the transpose and

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{pmatrix}.$$

The Bernstein basis functions in Eq. (1), can be generated easily from the relation:

$$B_k^m = \frac{m!}{k!(m-k)!} t^k (1-t)^{m-k}.$$

B_k^m is an m-th degree polynomial, in our case a cubic, with the property:

$$\sum_{k=0}^{m} B_k^m = 1.$$

They are coefficients of the expansion $[t + 1 - t]^m = 1$.

By substituting 0 and 1 for t into Eq. (1), it can be seen that the curve passes through the first and the last points. Examining the first and the second derivatives with respect to t, $\mathbf{r}'$ and $\mathbf{r}''$ respectively, further insight into the behavior of the Bézier curves can be gained. Thus:

$$\mathbf{r}'(t) = 3(1-t)^2[\mathbf{r}_1 - \mathbf{r}_0] + 6t(1-t)[\mathbf{r}_2 - \mathbf{r}_1] + 3t^2[\mathbf{r}_3 - \mathbf{r}_2]. \tag{2}$$

Substituting 0 and 1 for t into Eq. (2), one gets for the tangents:

$$\begin{aligned} \mathbf{r}'(0) &= 3[\mathbf{r}_1 - \mathbf{r}_0], \\ \mathbf{r}'(1) &= 3[\mathbf{r}_3 - \mathbf{r}_2]. \end{aligned}$$

This shows that tangents at end points $(\mathbf{r}_0, \mathbf{r}_3)$ pass through other points $(\mathbf{r}_1, \mathbf{r}_2)$ respectively.

Taking the second derivative of $\mathbf{r}$:

$$\mathbf{r}''(t) = 6(1-t)[\mathbf{r}_0 - 2\mathbf{r}_1 + \mathbf{r}_2] + 6t[\mathbf{r}_3 - 2\mathbf{r}_2 + \mathbf{r}_1]$$

shows that at each end point $\mathbf{r}''$ depends only on three position vectors. This is important information if control of continuity is sought when making a composite Bézier curve by joining segments together.

We shall now return to the study of the Bernstein basis functions (also known as *blending* functions). Fig. 1 depicts these functions for $m = 3$. It can be seen that at $t = 0$ only B_0^3 contributes to the curve. Similarly the curve at $t = 1$ is controlled completely by B_3^3. The other two functions reach their maxima at $\frac{1}{3}$ and $\frac{2}{3}$ of the interval respectively.

COMPUTATIONS

Eq. (1) for the 2-D case is given explicitly as:

$$\begin{aligned} x &= (1-t)^3 x_0 + 3t(1-t)^2 x_1 + 3t^2(1-t)x_2 + t^3 x_3, \\ y &= (1-t)^3 y_0 + 3t(1-t)^2 y_1 + 3t^2(1-t)y_2 + t^3 y_3. \end{aligned}$$

Here

$$(x_k,\ y_k) \qquad k = 0, 1, 2, 3$$

are the coordinates of the four points which control the shape of the Bézier curve. In Fig. 2 some examples are given to show the kind of shapes which can be generated easily. For those who want to duplicate (and check their programs) Table I gives the corresponding coordinates.

Table I. Coordinates of polygon points to draw Bézier curves. Unit: 1 pt (0.3515 mm)

	x_0	y_0	x_1	y_1	x_2	y_2	x_3	y_3
1	0	0	50	250	400	300	450	25
2	50	100	200	350	450	0	500	125
3	50	100	200	350	450	0	500	125
4	125	100	500	375	50	250	450	150

Table II lists the coordinates for the curve **4** as a function of t. For space saving, the coordinates (x, y) are rounded up to the nearest integer point value.

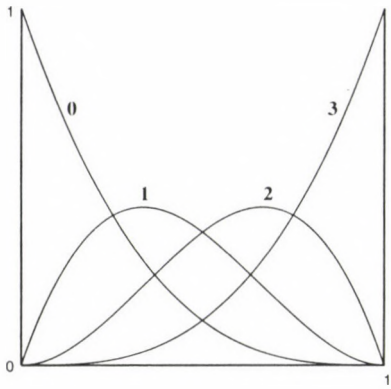


Figure 1. The Bernstein basis functions for degree equal to 3

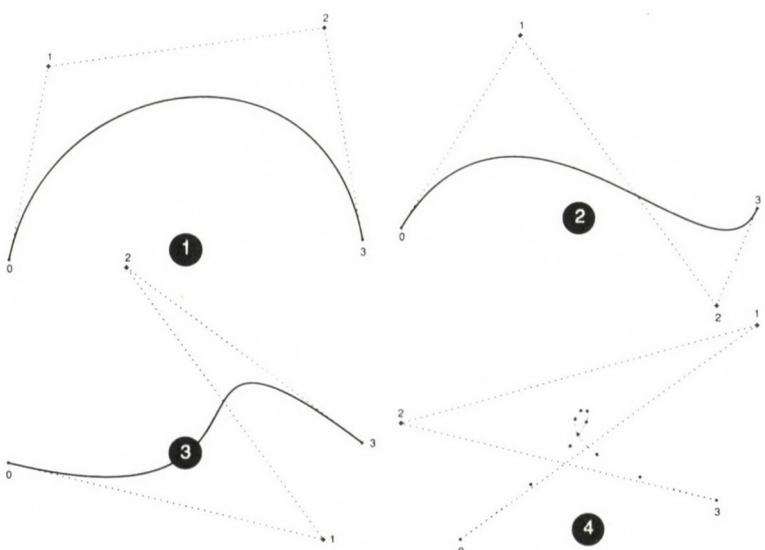


Figure 2. Some typical examples of using Bézier curves to draw lines.
The coordinates of control points for each curve are given in Table I

Figure 3. Times-Roman characters from PostScript (300 pt size)

Table II. Coordinates of points shown for the Bézier curve **4** in Fig. 2

t	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
x	125	214	264	285	286	278	271	274	299	354	450
y	100	170	220	251	265	266	255	235	210	180	150

APPLICATIONS

There are many applications of Bézier curves as can be seen from the references provided. Here only two major applications will be mentioned:

Typography: Font making is one of the major task in desktop publishing. Here the particular use of METAFONT designed for TEX and LATEX by Knuth (1986c,1986d) is mentioned. The other major application is PostScript. As an example in Fig. 3 the characters S and A are shown. The success of PostScript to be able to handle practically any size of characters (from points to several feet) is due to the use of Bézier curves.

Generalization: Reduction of maps brings in the problems of how to represent the boundary in a smoother manner. It is easy to see, for example, how the southern border of Hungary could be generalized with a few Bézier curves to provide a smooth form for page size diagrams.

SUMMARY

In this short introduction to Bézier curves, sufficient detail is provided to enable the reader to investigate the usefulness of the technique for a number of problems. Implementation of Bézier curve drawing on a PC is straightforward, and the interactive mode of operation provides an easy mastering of the design principles. Obviously for some applications one may have to use either higher degree Bézier curves and/or joining segments. One of the major advantages of using segments is, that changes within the segments have no effect at all on any other part of the development. This means that once a sectionally good design has been achieved it will not be changed by any other segment. Another advantage of using only the cubic Bézier curves is, that very quickly, the user picks up the idea how to change the control points to achieve the desired change in the shape. For higher and (variable) degree curves, this knowledge is more difficult to achieve.

REFERENCES

– Bézier curve related –

Bernstein S 1912: *Démonstration du théorème de Weierstrass fondeé sur le calcul des probabilités.* Harkov Soobs. Matem ob-va, **13**, 1-2.

Bézier P 1966: *Definition numérique des courbes et surfaces I.* Automatisme, **11**, 625–632.

Bézier P 1967: *Definition numérique des courbes et surfaces II.* Automatisme, **12**, 17–21.

Bézier P 1972: *Numerical Control; Mathematics and Applications.* Translated by R. Forrest, Wiley, New York

Bézier P 1977: *Essay de définition numérique des courbes et des surfaces expérimentales.* PhD thesis, University of Paris VI.

Bézier P 1986: *The Mathematical Basis of the UNISURF CAD System.*, Butterworths, London

Brueckner I 1980: *Construction of Bézier points of quadrilaterals from those of triangles.* Computer Aided Design, **12**(1), 21–24.

Chang G 1982: *Matrix formulation of Bézier technique.* Computer Aided Design, **14**(6).

Chang G, Feng Y 1985: *An improved condition for the convexity of Bernstein-Bézier surfaces over triangles.* Computer Aided Geometric Design, **1**, 279–298.

Coons S 1964: *Surfaces for computer aided design.* Technical Report, M.I.T.

de Boor C 1978: *A Practical Guide to Splines.* Springer

de Casteljau P 1959: *Outillages méthodes calcul.* Technical Report, A. Citroen, Paris

de Casteljau P 1963: *Courbes et surfaces à poles.* Technical Report, A. Citroen, Paris

de Casteljau P 1986: *Shape Mathematics and CAD.* Kogan Page, London

do Carmo M 1976: *Differential Geometry of Curves and Surfaces.* Prentice Hall

Farin G 1977: Konstruktion *und Eigenschaften von Bézier-Kurven und Flächen.* Master's thesis, Technical University Braunschweig, FRG

Farin G 1983: *Algorithms for rational Bézier curves.* Computer Aided Design, **15**(2), 73–77.

Farin G, Barry P 1986: *A link between Lagrange and Bézier curve and surface schemes.* Computer Aided Design, **18**, 525–528.

Farouki R, Rajan V 1988: *Algorithms for polynomials in Bernstein form.* Computer Aided Geometric Design, **18**(1), 15-21.

Foley J, Van Dam A 1982: *Fundamentals of Interactive Computer Graphics.* Addison-Wesley

Forrest A 1968: *Curves and surfaces for computer-aided design.* PhD thesis, Cambridge

Forrest A 1972: *Interactive interpolation and approximation by Bézier polynomials.* Computer Journal, **15**, 71–79.

Goldman R 1982: *Using degenerate Bézier triangles and tetrahedra to subdivide Bézier curves.* Computer Aided Design, **14**(6), 307-312.

Gordon W J 1969: *Free-form Surface Interpolation through Curve Networks.* Technical Report GMR-921, General Motors Research Laboratories, Warren, Michigan

Gordon W J, Riesenfeld R F 1974: *Bernstein-Bézier Methods for the Computer-Aided Design of Free-Form Curves and Surfaces.* Journal of the Association of Computing Machinery, **23**(2), 293–310.

Lasser D 1987: *Bernstein-Bézier Darstellung trivariater Splines.* PhD thesis, TH Darmstadt, FRG

Overhauser A 1968: *Analytic definition of curves and surfaces by parabolic blending.* Technical Report, Ford Motor Company

Piegl L 1986: *The sphere as a rational Bézier surface.* Computer Aided Geometric Design, **3**(1), 45–52.

Piegl L 1987: *A geometric investigation of the rational Bézier scheme in computer aided geometric design.* Computers in Industry, **7**, 401–410.

Sabin M 1976: *The use of piecewise forms for the numerical representation of shape.* PhD thesis, Hungarian Academy of Sciences, Budapest, Hungary

Shirman L, Séquin C 1988: *Local surface interpolation with Bézier patches.* Computer Aided Geometric Design, **4**(4), 279-296.

Staerk E 1976: *Mehrfach differenzierbare Bézierkurven und Bézierflächen.* PhD thesis, Technical University, Braunschweig

– TEX and PostScript related –

Braswell F M 1989: Inside PostScript. Peachpit Press, Inc. Berkeley California

Glover G 1989: Running PostScript from MS–DOS. Windcrest Books Division of TAB BOOKS Inc. Blue Ridge Summit, PA

Holzgang D A 1988: Understanding PostScript programming. Second Edition. SYBEX Inc. San Francisco

Holzgang D A 1989: PostScript Programmer's Reference Guide. Scott, Foresman and Company, Glenview, Illinois

Knuth D E 1986a: Volume A *The TEXbook*, Addison-Wesley Publishing Company, Inc.

Knuth D E 1986b: Volume B *TEX: The Program*, Addison-Wesley Publishing Company, Inc.

Knuth D E 1986c: Volume C *The METAFONTbook*, Addison-Wesley Publishing Company, Inc.

Knuth D E 1986d: Volume D *METAFONT: The Program*, Addison-Wesley Publishing Company, Inc.

Knuth D E 1986e: Volume E *Computer Modern Typefaces*, Addison-Wesley Publishing Company, Inc.

Kunkel G 1990: Graphic Design with PostScript. Scott, Foresman and Company, Glenview, Illinois

Lamport L 1986: LATEX: *A Document Preparation System*, Addison-Wesley Publishing Company, Inc.

Nagy D 1989a: *A bar chart in* LATEX. TUGBOAT Volume 10, Number 2. P.O. Box 9506 Providence, RI, U.S.A.

Nagy D 1989b: *Vertical centering for transparencies.* TUGBOAT Volume 10, Number 2. P.O. Box 9506 Providence, RI, U.S.A.

Nagy D 1989c: *An introduction to TEX.* Electronic Composition & Imaging, Vol.3, Issue 4, 200 Yorkland Blvd., Willowdale, Canada

Nagy D 1989d: *Mathematical typesetting.* Electronic Composition & Imaging, Vol.3, Issue 5, 200 Yorkland Blvd., Willowdale, Canada

Nagy D 1989e: *Mathematical typesetting.* Electronic Composition & Imaging, Vol.3, Issue 6, 200 Yorkland Blvd., Willowdale, Canada

Spivak M D 1986: *The Joy of TEX*, American Mathematical Society, Providence, Rhode Island

PostScript Language–Reference Manual 1985a: Addison-Wesley Publishing Company, Inc.

PostScript Language–Tutorial and Cookbook 1985b: Addison-Wesley Publishing Company, Inc.

PostScript Language–Program Design 1988: Addison-Wesley Publishing Company, Inc.

Roth S F ed. 1988: Real World PostScript, Addison-Wesley Publishing Company, Inc.

Smith R 1990: Learning PostScript. A Visual Approach. Peachpit Press, Inc. Berkeley California

Thomas B 1988: A PostScript Cookbook. Van Nostrand Reinhold New York

Acta Geod. Geoph. Mont. Hung., Vol. (1-4), pp. 29-31 (1991)

ENCLOSED AREA OF A POLYGON

D Nagy
Geological Survey of Canada
1 Observatory Crescent, Ottawa, Canada

[Manuscript received February 27, 1990]

Following a brief review of methods suggested recently by Stolk and Ettershank, and also by Messeder, another derivation is given for the calculation of the area enclosed by a polygon. It is shown, that after some simplification, the number of multiplication can be reduced by a factor of two.

Keywords: Area; enclosed polygon; minimum

INTRODUCTION

In a recent article, Stolk and Ettershank (1987b) commented on the calculation of the area enclosed by a closed polygon. Also on the same topic a letter was published by Messeder (1987a). His statement about a *more general* solution and no overlapping areas may require clarification.

Following the tradition well established in surveying, we return to the same topic and give a *different* and simpler derivation for the same problem: that is the computation of the area enclosed by an n-sided polygon. Our derivation obviously gets the same result, but in a different form, and by rearrangement, gives an alternate expression as well, which may be used to check out the correctness of the computation (admitedly, this computation is a simple one, however, the provision for alternate expression may build good programming habits). In addition, the derived expression has the *minimum* number of arithmetic operations required in the computation.

THEORY

Let us start from Fig. 1, and write out twice the area (in order to avoid the use of fractions) under the first polygon side[1].

[1]For simplicity, only the minimum polygon i.e. a triangle is used in the derivation.

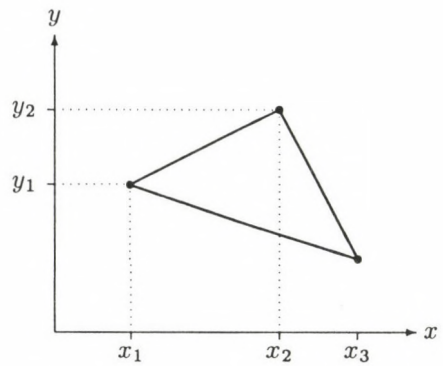


Fig. 1. Notation used in derivation

The area of this trapezoid is the base×height, i.e.:

$$2a = (x_2 - x_1)(y_1 + y_2).$$

Clock-wise direction gives a positive value for the area. The sum of areas for each polygon side actually gives the area inside of the polygon. Thus the general expression for an n-sided polygon can be written immediately:

$$2A = \sum_{i=1}^{i=n} (x_{i+1} - x_i)(y_i + y_{i+1}). \tag{1}$$

In order to get the simplified form, the area for a triangle, ($n = 3$), will be written out in detail, then regrouped:

$$\begin{aligned}
2A &= (x_2 - x_1)(y_1 + y_2) + (x_3 - x_2)(y_2 + y_3) + (x_4 - x_3)(y_3 + y_4) \\
&= (x_2 - x_1)y_1 + (x_2 - x_1)y_2 + (x_3 - x_2)y_2 + (x_3 - x_2)y_3 + \\
&\quad + (x_4 - x_3)y_3 + (x_4 - x_3)y_4 \\
&= [(x_2 - x_3) + (x_1 - x_3)]\, y_1 + [(x_2 - x_1) + (x_3 - x_2)]\, y_2 + \\
&\quad + [(x_3 - x_2) + (x_1 - x_3)]\, y_3 \\
&= [x_2 - x_3]y_1 + [x_3 - x_1]y_2 + [x_1 - x_2]y_3.
\end{aligned}$$

Here we made use of the property, that:

$$x_{n+1} \equiv x_1 \quad \text{and } y_{n+1} \equiv y_1.$$

Thus the general expression to calculate the area of an n-sided polygon is:

$$2A = \sum_{i=1}^{i=n} (x_{i+1} - x_{i+2})y_i. \tag{2}$$

Regrouping the expression in a different manner, or forming the trapezoids with respect to the y axis, the following alternate form can be obtained for the area:

$$2A = \sum_{i=1}^{i=n} (y_{i+2} - y_{i+1})x_i. \tag{3}$$

Since the area is the absolute value of the right-hand side, multiplication by -1 is permitted, which then results in the form:

$$2A = \sum_{i=1}^{i=n} (y_{i+1} - y_{i+2})x_i, \tag{4}$$

which can be obtained directly from Eq. (2) by simply exchanging x and y.

It is noted here, that all forms, including those of the authors quoted earlier, give the same result. The difference is in the derivation and in the number of algebraic operations. All forms require $2n-1$ additions or subtractions, but whereas the other forms quoted, require $2n$ multiplications, (which is normally a more time consuming operation than addition or subtraction), the form I give requires *only half of the multiplications*, i.e. only n.

In conclusion, we return to the statment of Messeder, that his *method may provide a more general solution analysis than the Stolk and Ettershank approach.*

Since all methods are equivalent, it is hard to see the validity of this statement. Furthermore, Messeder showed his derivation on a *special* (definitely not a general) triangle, with one point at the origin. Writing out the area for a triangle in a general position results exactly in the form given by Stolk and Ettershank. Of course in this case there will be overlapping areas as well.

REFERENCES

Messeder R L 1987a: BYTE, June, 28.

Stolk R, Ettershank G 1987b: BYTE, February, 135–136

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 33-43 (1991)

TEST COMPUTATIONS FOR A LOCAL QUASIGEOID IN HUNGARY USING FFT

J Ádám[1] and H Denker[2]

[1]Institute of Geodesy, Cartography and Remote Sensing, Satellite Geodetic Observatory, H-1373 Budapest, POB 546, Hungary
Temporarily: Dept. of Geodetic Science and Surveying, The Ohio State University, 1958 Neil Avenue, Columbus, Ohio 43210-1247, USA

[2]Institut für Erdmessung (IFE), Universität Hannover, D-3000 Hannover 1, Nienburger Str. 6, FRG

[Manuscript received April 3, 1990]

The Fast Fourier Transform (FFT) technique is an efficient and practical tool in geodesy for solving local gravity field approximation problems, among others for local geoid (quasigeoid) determination (Schwarz et al. 1989). In the future we plan to use the FFT procedure for its efficiency in computing a detailed gridded set of height anomalies for Hungary (Ádám et al. 1988).

As a first step, a test calculation was carried out for the determination of height anomalies in Hungary using point gravity data published by Renner (1959) and Renner and Szilárd (1959) as well as the global geopotential model OSU86F (Rapp and Cruz 1986).

Keywords: FFT; geoid; gravity anomalies; Hungary; quasigeoid

1. COMPUTATION METHOD

Height anomalies have been determined for Hungary using FFT techniques. The predicted height anomalies are obtained by

$$\zeta = \zeta_1 + \zeta_2 \ , \tag{1}$$

where ζ_1 is the influence of the spherical harmonic model, and ζ_2 is the contribution from terrestrial gravity field observations. The spherical harmonic model is used as a reference field and yields the major part of the quasigeoid. Improved results can in general be obtained by using terrain reductions as shown by Kearsley et al. (1985).

After subtracting the effect of a global geopotential

model from all observations, the contribution of terrestrial gravity field observations (ζ_2) can be computed by integral formulas evaluated by FFT techniques. The spectral computation of the disturbing potential and its functionals by FFT is based on flat-earth approximations. Thus, Stokes' integral formula can be written as a two-dimensional convolution in the form (Schwarz et al. 1989)

$$\zeta_2 = \frac{1}{2\pi\gamma}\, s * \Delta g_r, \qquad s = \left(x^2+y^2\right)^{-1/2}, \tag{2}$$

where Δg_r are the gravity anomalies reduced for the effect of a geopotential model. The convolution of the kernel function s with the data Δg_r is most easily done in the frequency domain. Using the analytical transform of s, formula (2) can be written as

$$\zeta_2 = \frac{1}{2\pi\gamma}\, F^{-1}\left\{ \frac{2\pi}{\omega}\, \Delta\tilde{g}(u,v) \right\}, \qquad \omega=\sqrt{u^2+v^2}, \tag{3}$$

where F^{-1} denotes the inverse Fourier transform, (u,v) are the frequencies and $\Delta\tilde{g}$ is the Fourier transform of Δg_r.

This computation procedure and numerical results have been discussed in a number of papers, see e.g., Denker (1988, 1989), Forsberg and Solheim (1989), Kearsley et al. (1985), Sideris (1987), Schwarz et al. (1989) and Tziavos (1987).

2. DATA COLLECTION AND EVALUATION

For the determination of the long wavelength part of the Earth's gravity field, the geopotential model OSU86F has been used (Rapp and Cruz 1986). This model is complete to degree and order 360 corresponding to a spatial resolution of approx. 50 km. The geopotential solution OSU86F has an accuracy estimate of roughly ±0.5-1.0 m including omission errors.

For the generation of the reference height anomalies ζ_1

and gravity anomalies (Δg_{GM}) from the OSU86F potential model complete to degree and order 360, the efficient algorithm of Rizos has been used (see e.g., Tscherning et al. 1983). The reference height anomalies and gravity anomalies were computed for a $0^{o}25$ grid in the area [$40^{o} = \phi = 55^{o}$; $10^{o} = \lambda = 30^{o}$]. Altogether, 4941 gravity anomalies respectively height anomalies were computed on this $0^{o}25$ grid. These gridded values were then interpolated to the actual station location using a 4*4 spline window.

For the determination of the medium and short wavelength part of the earth's gravity field, the point gravity data reported by Renner (1959) and Renner and Szilárd (1959) have been used for the test calculation. These gravity data were obtained by means of a Heiland gravimeter (type GSC-3, No. 40) on 16 first order base stations and on 493 second order base stations while establishing a fundamental gravity network across Hungary in the years 1950-1955. The first order network was adjusted by least squares. The adjusted gravity values of the first order network stations were fixed in the adjustment of the second order network. The second order network was adjusted in one stage to form a unified network as a whole. The precision of gravity values in the first order network was estimated as ± 0.02 mgal (std. dev.). The precision of the gravity data in the second order networks is estimated to be ± 0.029 mgal. These gravity values are given in the Potsdam system. The stations of the first order network are distributed all over the country, some 100 to 120 kilometres apart, while the average spacing between second order gravity stations is 15 to 20 km. Figure 1 shows the location of the gravity stations.

The adjusted gravity values were transformed into free air gravity anomalies at 508 of the 509 points using the International Gravity Formula (1930). The free air gravity anomalies computed in this way and referring to the Potsdam gravity system are shown in Fig. 2.

The gravity anomaly data reported by Renner (1959) and Renner and Szilárd (1959) were converted as follows in order to make them suitable for the test computation. The appropriate formula for the conversion from the International Gravity

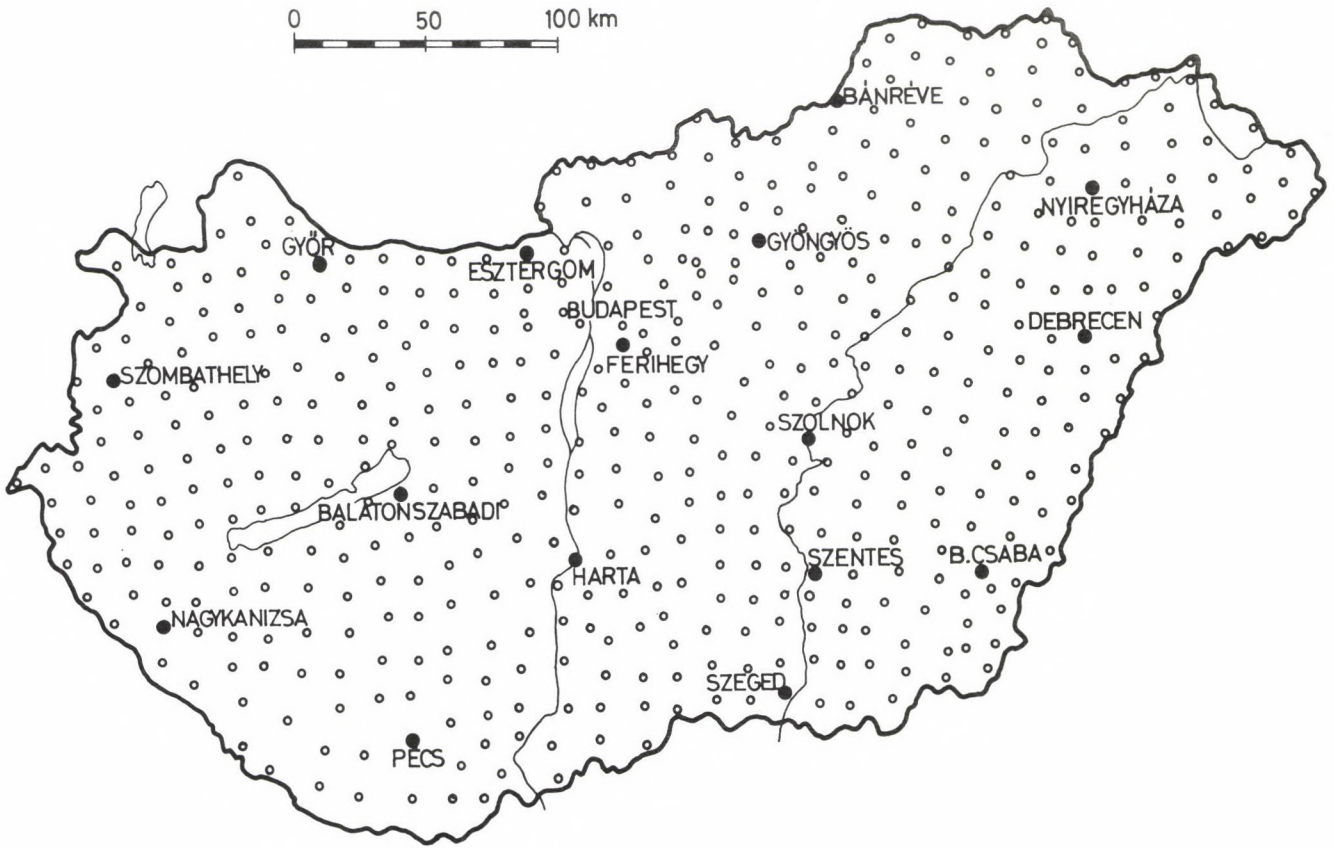


Fig. 1. Distribution of point gravity data in Hungary used for test computation (after Renner 1959)

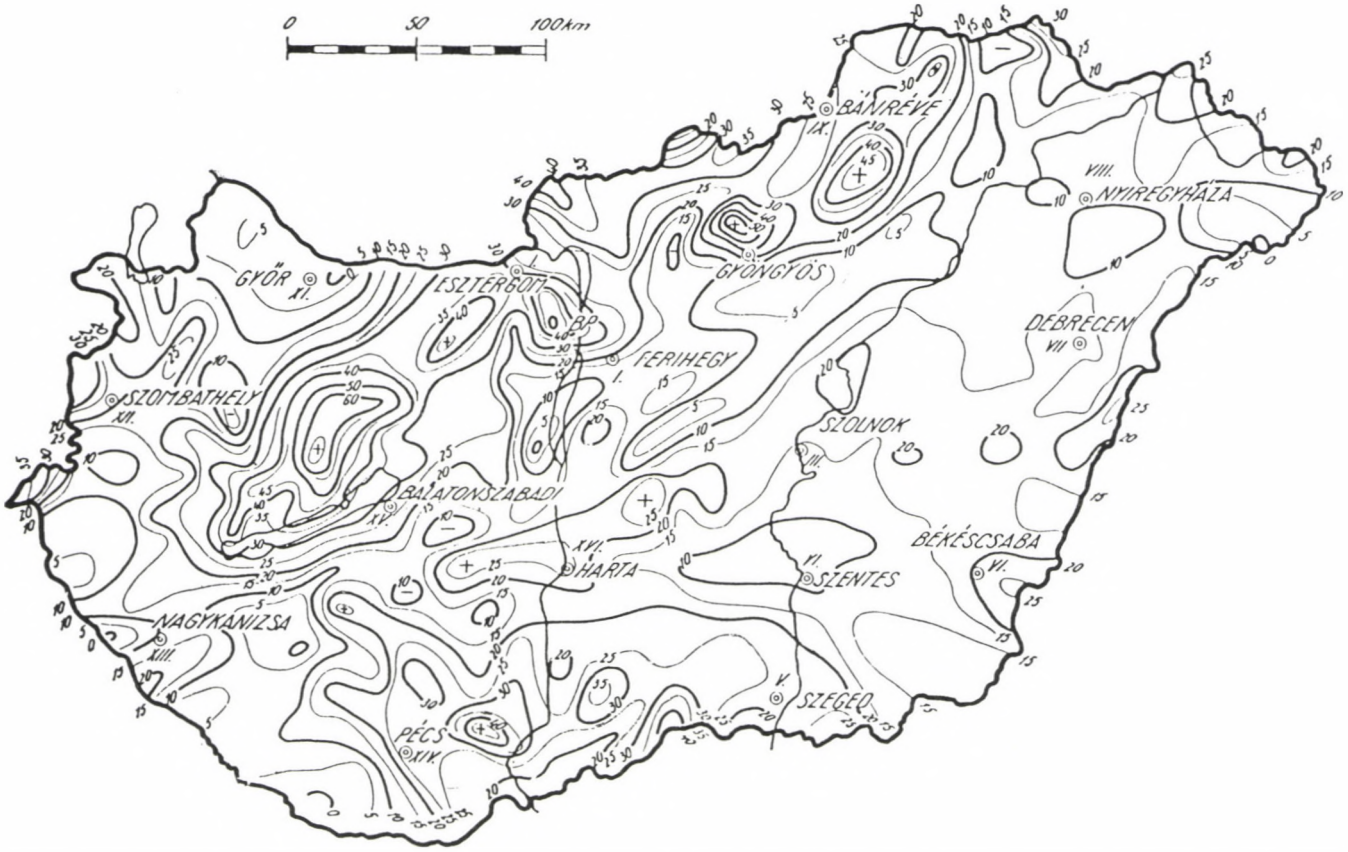


Fig. 2. Representation of the free air gravity anomalies computed from point gravity values using the Int. Gravity Formula (1930) and the Potsdam gravity system. The contour interval is 5 mgal (after Renner 1959)

Formula (1930) to the Geodetic Reference System 1980 (GRS80) is given by Moritz (1984):

$$\gamma_{1980} - \gamma_{1930} = (-16.3 + 13.7*\sin^2 \phi) \quad \text{mgal.} \qquad (4)$$

Taking into account the -14 mgal for the difference between the Potsdam gravity system and the IGSN71 system, we get the following equation

$$\Delta g_{1980} = (\Delta g_{1930} + 2.3 - 13.7*\sin^2 \phi) \quad \text{mgal.} \qquad (5)$$

Equation (5) was applied for the data set of Renner (1959) and Renner and Szilárd (1959).

3. RESULTS OF THE TEST CALCULATION

The first step in the processing was to calculate the residual gravity anomalies Δg_r. These values were found by simply subtracting the interpolated model gravity anomaly values Δg_{GM} from the terrestrial gravity anomaly values Δg_{1980}. These residual gravity anomalies (Δg_r) were gridded in a 0^o1*0^o1 grid for the area [$43^o5 = \phi = 50^o5$; $13^o0 = \lambda = 26^o0$] using the GEOGRID program. This prediction area was chosen 2^o in latitude and 3^o in longitude larger than the area for which the „final" geoid was to be retained. In total, 9301 0^o1 grid values were calculated. All data are gridded in a geographic grid. The gridding was done using collocation (kriging). The number of closest points per quadrant used for the prediction is 4.

The residual gravity anomalies in Hungary referring to GRS80 are shown in Fig. 3. The contour interval is 5 mgal.

The residual height anomalies ζ_2 were then calculated using the GEOFOUR program. A total of 8192 values were computed on a 0^o1 grid in the area [$43^o8 = \phi = 50^o1$; $13^o1 = \lambda = 25^o8$]. Figure 4 shows the residual height anomalies computed by the FFT method. Values are given on a 0^o1*0^o1 grid. Contour interval is 10 cm.

For generating the actual reference height anomalies ζ_1,

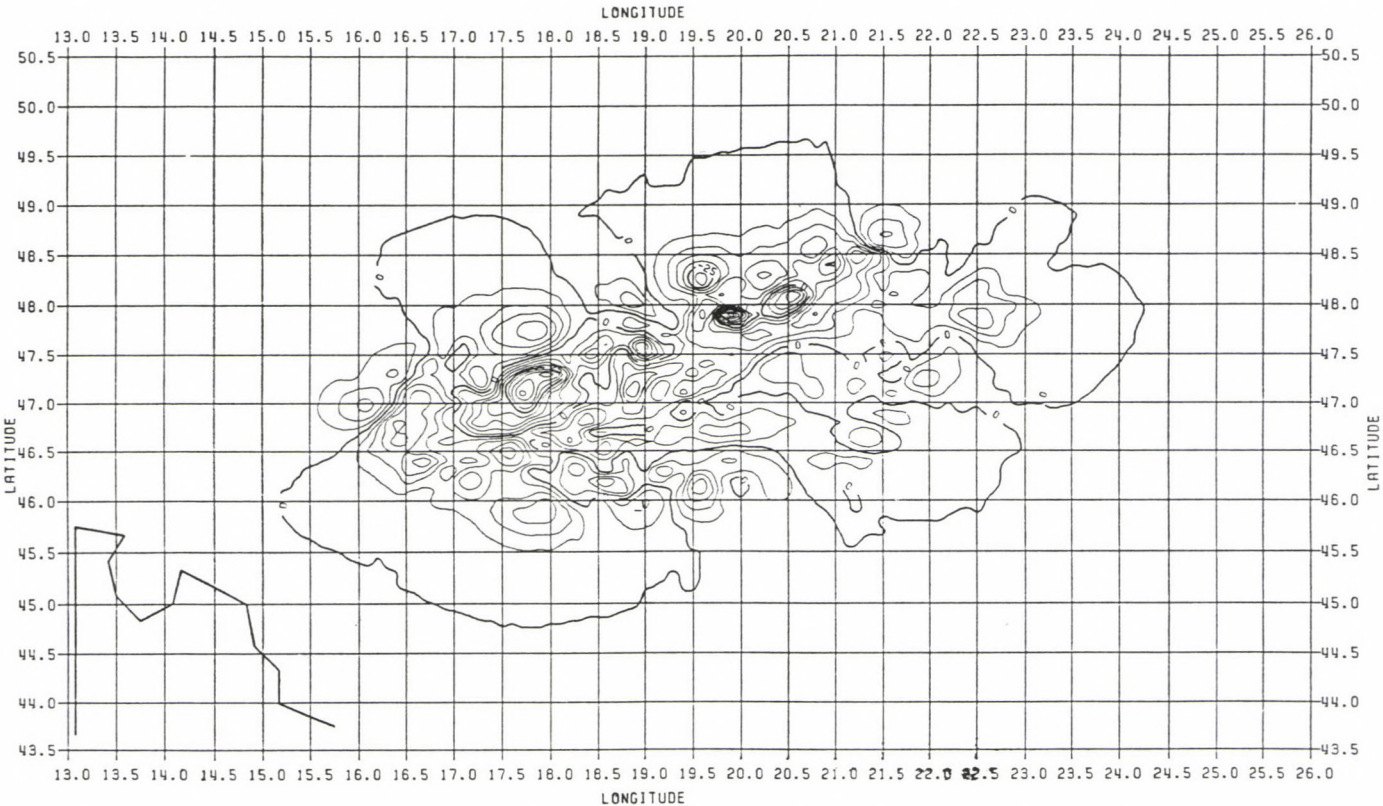


Fig. 3. A representation of the residual gravity anomaly field. Values given on $0.1^{o} \times 0.1^{o}$ grid. Contour interval: 5 mgal. Gravity anomalies refer to the Geodetic Reference System 1980 (GRS80)

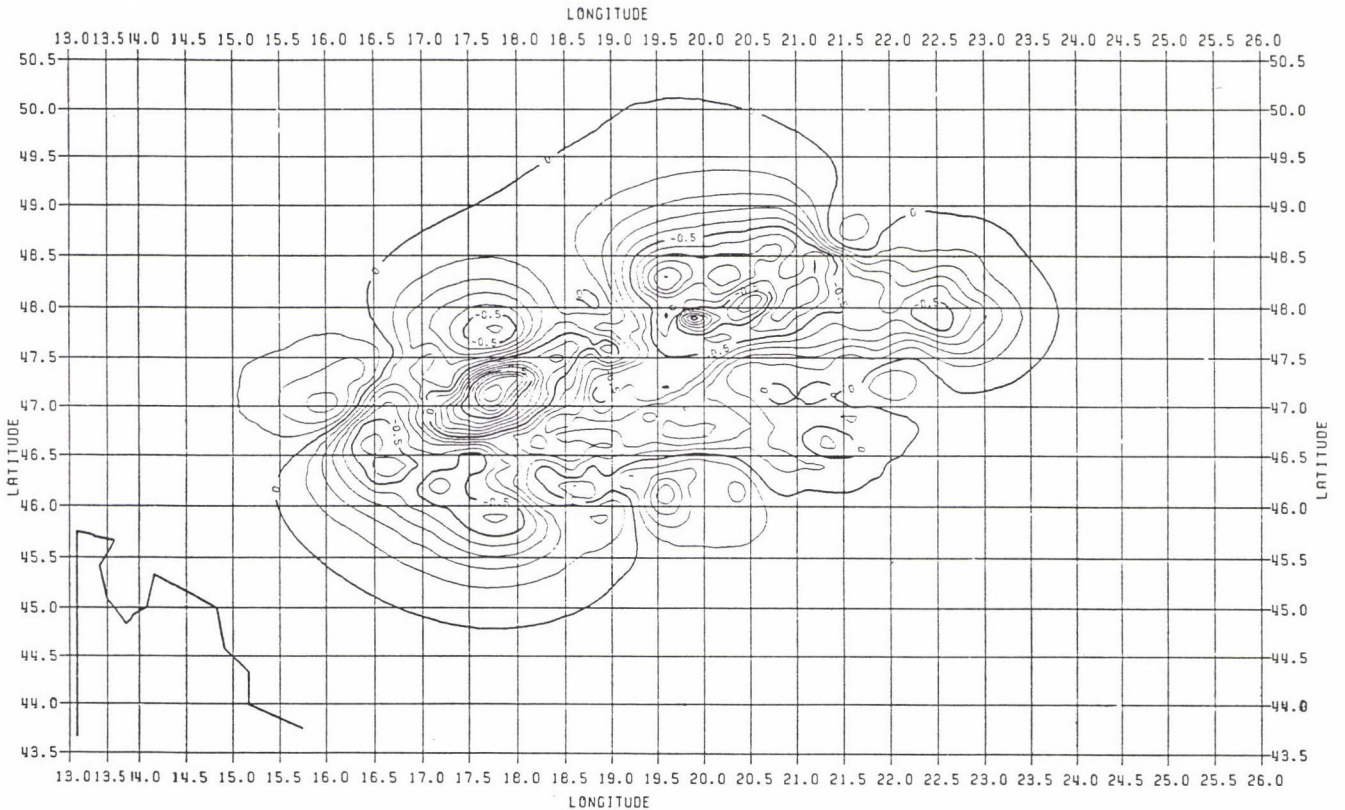


Fig. 4. Residual height anomalies computed by FFT method. Values given on $0.1^{O} \times 0.1^{O}$ grid. Contour interval: 10 cm

the computed values generated from the OSU86F field to degree and order 360 on a $0^{\circ}.25$ grid (see Chapter 2) were interpolated to the detailed $0^{\circ}.1$ grid in the same area of [$43^{\circ}.8 = \phi 50^{\circ}.1$; $13^{\circ}.1 = \lambda = 25^{\circ}.8$]. A total of 8192 values were calculated. Finally, the total height anomaly ζ was obtained by adding the reference height anomaly ζ_1 to the residual height anomaly ζ_2.

The geoid map for Hungary obtained from the FFT analysis and referring to the Geodetic Reference System GRS80 is shown in Fig. 5. The significant details that can be seen in Figs 4 and 5, lying in the northeast/southwest direction in two separated areas, primarily reflect the anomaly field associated with the topography in that regions.

The FFT method as used herein does not yield accuracy estimates although such estimates are possible (Sideris 1987). The effect of the terrain on the quasigeoid has not been taken into account. A way to do this is to calculate the residual terrain model (RTM) effect on both anomalies and undulations. This can be done by using a remove-restore technique as proposed by Forsberg and Tscherning (1981).

The FFT procedure described in this report can be used to construct a high resolution local quasigeoid for Hungary using all the available gravity data. Although the amount and distribution of the point gravity data (Renner 1959, Renner and Szilárd 1959) used in this test calculation is not sufficient for a detailed quasigeoid determination, however, these gravity values may be useful to involve them in future geopotential solutions.

ACKNOWLEDGEMENTS

We express our thanks to Dr. R H Rapp of the Ohio State University for providing the OSU86F geopotential solution, and Dr. R Forsberg of the Danish Geodetic Institute for making available his computer programs GEOGRID and GEOFOUR. Figures 3-5 were prepared with the GSPP plot package made by H Sünkel (1980).

Computer support was provided by the Instruction and

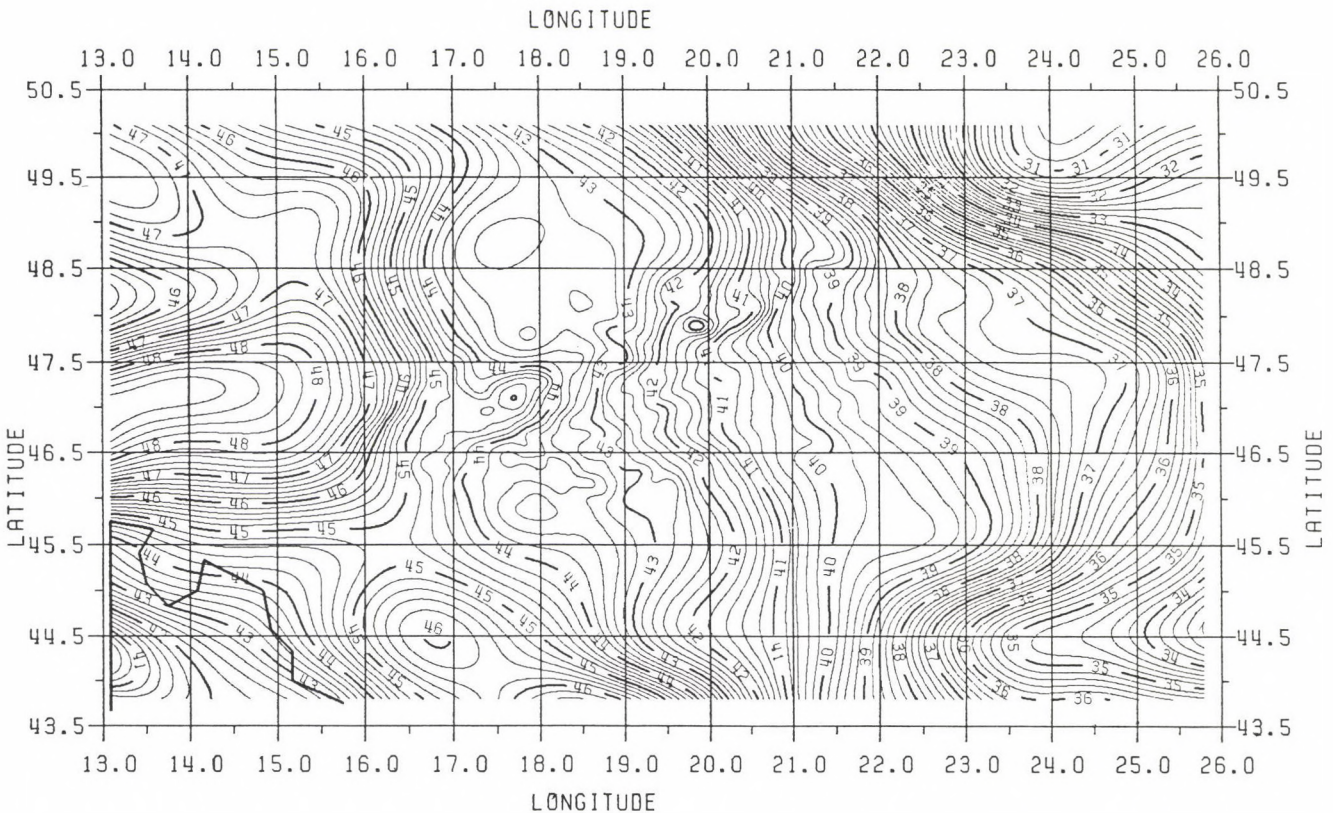


Fig. 5. Geoid undulations in the area $[43^{\circ}.8 = \phi = 50^{\circ}.1;\ 13^{\circ}.1 = \lambda = 25^{\circ}.8]$ based on the OSU86F geopotential model plus contributions from gravity anomaly data of Hungary through FFT analysis. Values given on a $0.1^{\circ} \times 0.1^{\circ}$ grid. Contour interval: 20 cm

Research Computer Center of The Ohio State University through the Department of Geodetic Science and Surveying.

REFERENCES

Ádám J, Bölcsvölgyiné Bán M, Gazsó M, Kenyeres A, Sárhidai A 1988: In: Proceedings of the Int. Symp. on „Instrumentation, Theory and Analysis for Integrated Geodesy", Sopron, Hungary, Vol. 1, 90-109.

Denker H 1988: Hochauflösende regionale Schwerefeldbestimmung mit gravimetrischen und topographischen Daten. Wiss. Arb. Univ. Hannover, Nr. 156, Hannover

Denker H 1989: In: Progress in the Determination of the Earth's Gravity Field, R H Rapp (ed.), Report No. 397, 121-124, Dept. of Geodetic Science and Surveying, The Ohio State Univ., Columbus, Ohio, USA

Forsberg F, Solheim D 1989: In: Progress in the Determination of the Earth's Gravity Field, R H Rapp (ed.), Report No. 397, 100-103, Dept. of Geodetic Science and Surveying, The Ohio State Univ., Columbus, Ohio, USA

Forsberg R, Tscherning C C 1981: J. Geophys. Research, 86, 7843-7854.

Kearsley A H W, Siders M G, Krynski J, Forsberg R, Schwarz K P 1985: White Sands Revisited — A Comparison of Techniques to Predict Deflections of the Vertical. UCSE Reports, No. 30007. Department of Surveying Engineering, The University of Calgary, Calgary, Alberta, Canada

Moritz H 1984: Bull. Geod., 58, 388-398.

Rapp R H, Cruz J Y 1986: Spherical Harmonic Expansions of the Earth's Gravitational Potential to Degree 360 Using 30' Mean Anomalies. Report No. 376, Dept. of Geodetic Science and Surveying, The Ohio State Univ., Columbus, Ohio, USA

Renner J 1959: Geofizikai Közlemények, 8, No. 3, 105-141.

Renner J, Szilárd J 1959: Acta Technica Acad. Sci. Hung., 23, 365-395.

Schwarz K P, Sideris M G, Forsberg R 1989: The Use of FFT Techniques in Physical Geodesy. Geophysical Journal (in press)

Sideris M G 1987: In: Proc. of the XIXth IUGG General Assembly, Vancouver, B.C., 2, 428-442.

Sünkel H 1980: Manual to the General Surface Representation Module Designed for Geodesy (GSPP). Internal Report, Dept. of Geodetic Science, The Ohio State Univ., Columbus, Ohio, USA

Tscherning C C, Rapp R H, Goad C C 1983: Manuscripta Geodaetica, 8, 249-272.

Tziavos I N 1987: In: Proc. of the XIXth IUGG General Assembly, Vancouver, B.C., 2, 177-193.

Acta Geod. Geoph. Mont. Hung., Vol. (1-4), pp. 45-55 (1991)

THE FR-1 FIELD INSTRUMENT FOR AUTOMATIC RECOGNIZING AND RECORDING OF NEAR EARTHQUAKES

Gy Mentes

Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H-9401 Sopron, POB 5, Hungary

[Manuscript received July 9, 1990]

The construction and the operation principle of an automatic earthquake recognizing and recording field instrument is presented. The instrument is intended to record near week earthquakes to know the local seismic activity in the vicinity of dangerous industrial establishments. A very simple algorithm for earthquake recognition and a simulated example to show the efficiency of the algorithm are given.

Keywords: earthquake recognition; field instrument; recognition algorithm

INTRODUCTION

The destroying effect of large earthquakes is well-known. The demage can be intensified if dangerous industrial projects e.g. chemical factories, nuclear power-stations etc. are destroyed and an additional high-grade and long-term demage of the surroundings is caused by an earthquake. Therefore it is very important to know the local seismic activitiy in the district of dangerous establishments. For this purpose it is recommended to place more seismographs onto the field closely around the dangerous object. However, for earthquake recording in more points in the field low-cost instruments with high reliability are needed. The existing instruments (e.g. KINEMETRICS) are expensive and have much more ability then we need for this purpose. That was the reason why we developed an automatic earthquake recognizing and recording instrument which has the following main properties:

1. The instrument has a very low energy consumption and can operate from a battery or a solar-cell.

2. The instruments placed in the field can be synchronized by means of a built-in DCF-77 time-signal receiver.

3. The instrument stores the recognized earthquake data in its semiconductor memory. The recorded data of the individual instruments placed in the field are collected by means of a portable cassette recorder or a laptop computer.

THE OPERATION OF THE FR-1 INSTRUMENT

Figure 1 shows the block diagram of the FR-1 automatic earthquake recognizing and recording instrument. It is built with a Z80 CMOS microprocessor. All of the functions of the instrument which can be realized by software are programmed for the microprocessor to minimize the number of elements of the instrument. Thus the instrument is very reliable when put it into action in the field. The controlling program of the instrument is burnt into the ROM memory. The instrument has a CMOS solid-state memory for data storage. The capacity of this memory is 64K bytes and is expandable to 1024K bytes on two extra boards of 256K bytes each.

The instrument has an inner clock with calendar for identification of the recognized earthquakes. The built-in DCF-77 receiver is used to refresh the inner clock. When switching-on the instrument, the microprocessor decodes the received time-signal and writes the correct time into the inner clock. After that the FR-1 is ready for operation. The inner clock causes an interrupt once per day and the microprocessor refreshes the inner clock in the above described way. This solution provides a reliable timing when receiving disturbances, too.

The inner clock controls the 8 bits A/D converter. The sampling rate of the converter is 32 samples per second. The sampling is made continuously and data are stored only if an earthquake is detected by the microprocessor. The low-noise amplifier has a gain of more then 80 dB. The anti-alias filter built together to the amplifier has a fixed cut-off frequency of 8 Hz.

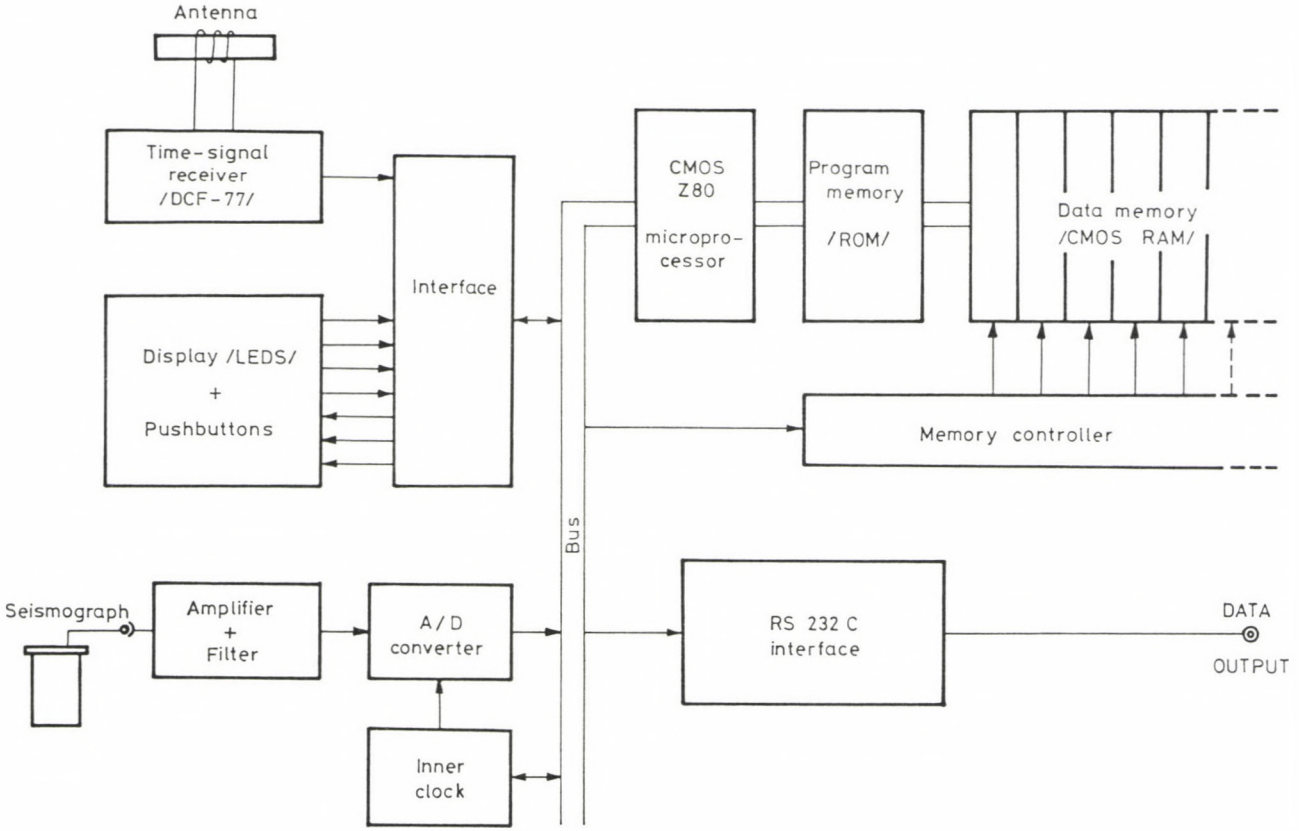


Fig. 1. The block diagram of the FR-1 automatic earthquake recognizing and recording instrument

The instrument works fully automatically. It has three pushbuttons for handling, one to reset the microprocessor, one to control the read-out of the data from its memory and one for enabling the four LED indicators. For sake of low energy consumption the LEDs light until this last push-button is pushed. The LEDs indicate the power supply, the time syncronization, the full memory and the recording of a recognized earthquake.

The data are read-out from the CMOS memory via a RS 232 C port into a portable laptop computer. By means of an adapter and a portable cassette recorder the data can be stored on a cassette, too. By means of this adapter the data can be backplayed from the cassette into a computer having a RS 232 C port (Fig. 2).

Figure 3 shows the global flow-chart of the controlling program. The main part of the program is the trigger algorithm which recognizes the earthquakes.

THE RECOGNIZING PROCESS OF AN EARTHQUAKE

Many algorithms are already known for earthquake recognition (Stewart 1977, Allen 1978, Gofort and Herrin 1981). For the FR-1 instrument an algorithm was developped which performs the recognition very rapidly by means of simple arithmetic operations and the calculation do not need a lot of memory locations.

The sampled raw data are stored as pre-event values continuously in the RAM memory for earthquake data. The number of pre-event values is 320 what corresponds to the time interval of 10 seconds. If the number of the sampled and stored pre-event data is more than 320, then the first value is shifted out of the memory and all data are shifted forward by one and the latest sampled data will be the 320th data if no earthquake had occurred.

In the case of a recognized earthquake the data storage is made continuously for 30 seconds without shifting. So the stored data block consists of 320 pre-event and 960 post-event

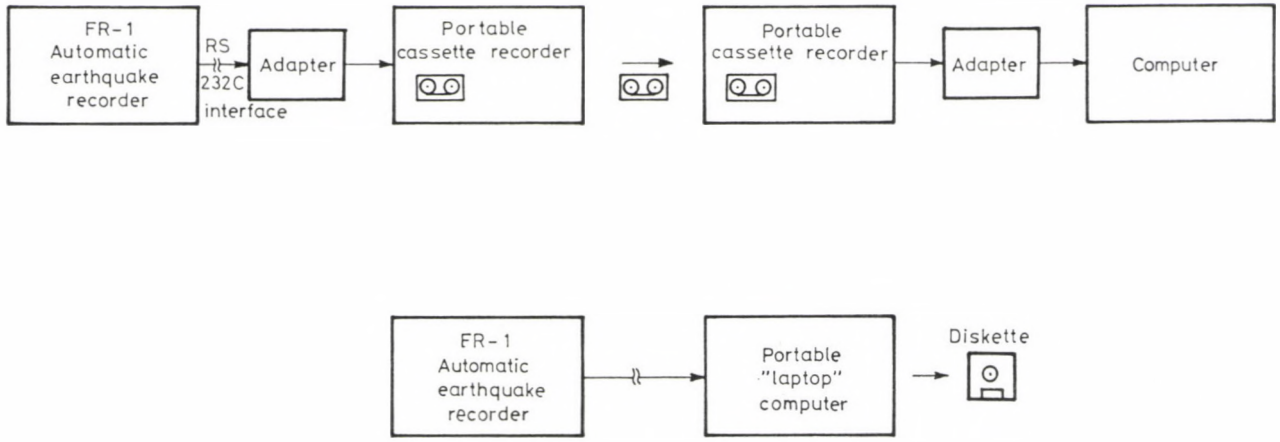


Fig. 2. The possibilities of the data transfer from the FR-1 instrument into a computer

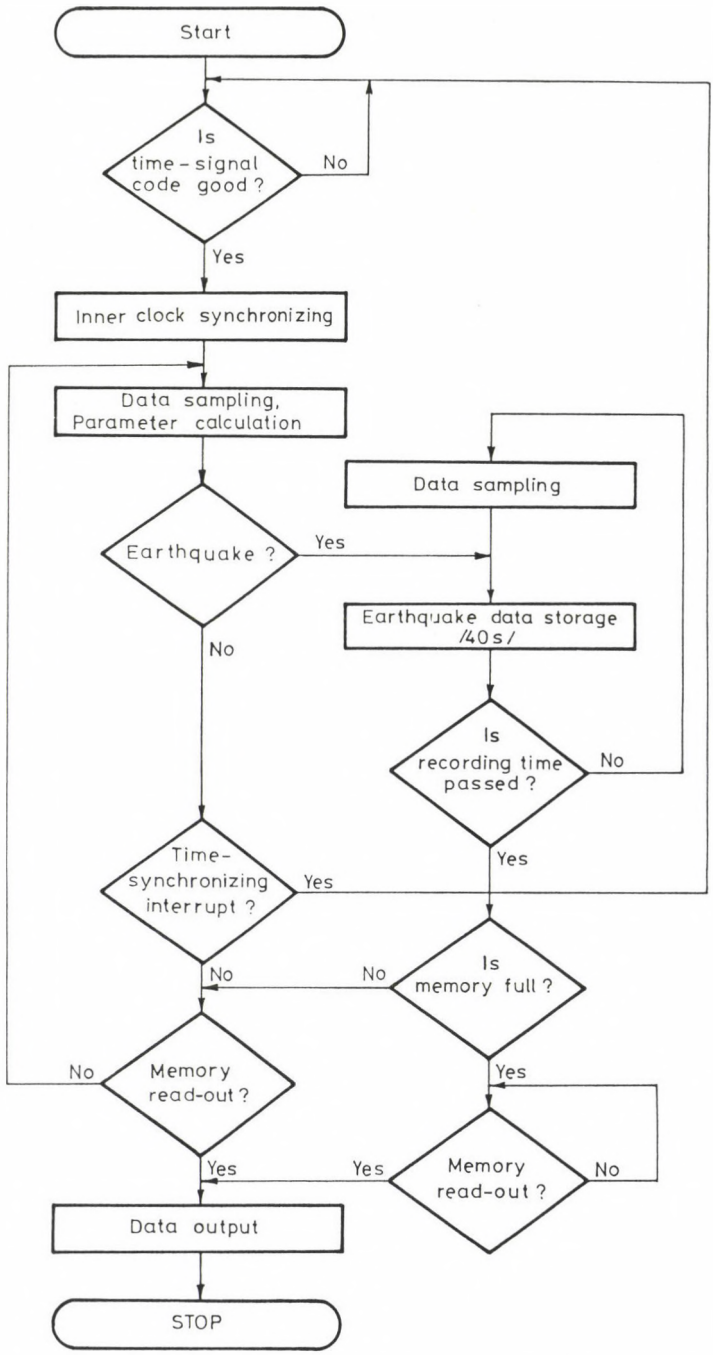


Fig. 3. The flow-chart of the controlling program of the FR-1 instrument

data.

The earthquake recognition is based upon a comparison of the Short Term Average (STA) and the Long Term Average (LTA) of the digitized seismic data. The STA is computed by averaging the absolute value of each sample over the STA interval of 4 seconds. The calculation of the absolute value, that is, the rectification of the seismic signal is performed digitally in parallel with the raw data storing and is used only for triggering computations. There is a 128 bytes length buffer for storage of the absolute values for calculation of the STA. When a new value is coming all bytes of the buffer are shifted forward. The first value is lost and the new value is written into the last byte. Before the first data is lost, the new STA is calculated in the following way:

$$STA_i = \frac{STA_{i-1} \cdot 128 - DATA_{i-128} + DATA_i}{128} \quad , \quad \text{if} \quad i > 128 \; . \qquad (1)$$

The LTA is computed as follows:

$$LTA_i = \frac{LTA_{i-1} \cdot 32 - LTA_{i-1} + STA}{32} \quad , \quad \text{if} \quad i > 4096 \; . \qquad (2)$$

The initial LTA is established by averaging 32 STA's. The triggering is inhibited during the calculation of the initial LTA. These calculation methods are very effective for microprocessor programming because the multiplications and divisions by 32 and 128 can be performed very simply by shifting of register contents.

The program calculates the STA_i/LTA_{i-1} ratio. If this is greater than a pregiven trigger level, the program detects an earthquake and the instrument stores altogether 1280 seismic data and the exact date and time of the last sample in its memory. During this time the LTA will be so high that a long earthquake cannot trigger the instrument again. The trigger level can be changed from 2 to 5 in the FR-1 instrument. Figure 4 shows the variations of STA, LTA and the ratio of STA/LTA during an earthquake. The data were recorded with the FR-1 instrument and then processed by a simulation program to

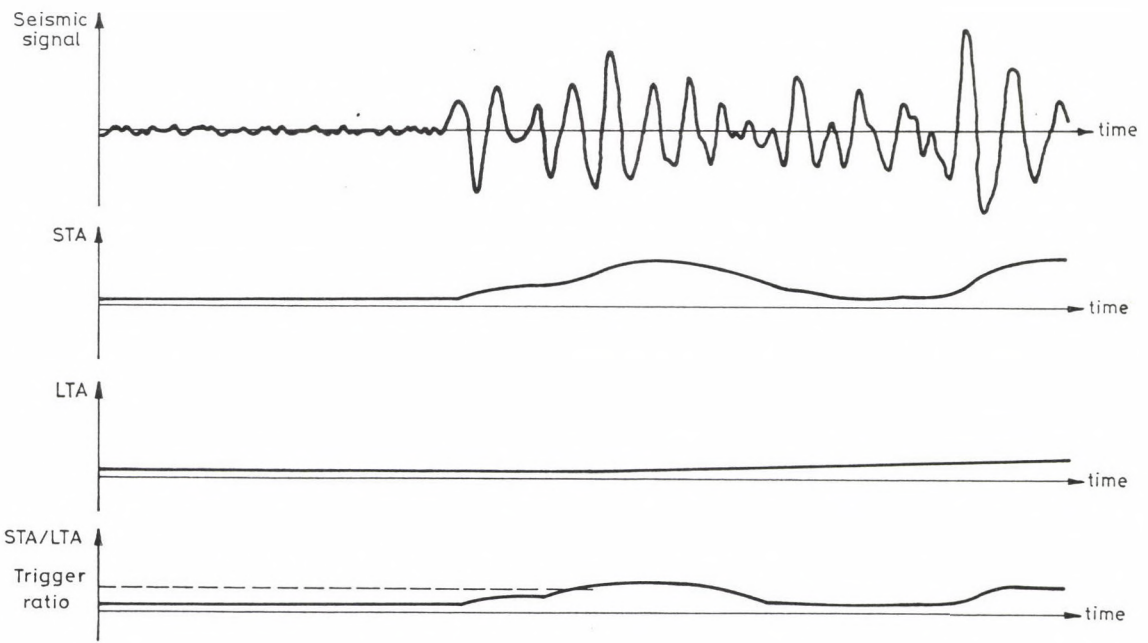


Fig. 4. The variations of the STA, LTA and the ratio of STA/LTA

indicate the change of the above mentioned values.

THE FIRST PRACTICAL EXPERIENCES WITH THE FR-1 INSTRUMENT

The prototype of the instrument has been operating in the Geodynamical Observatory of the Geodetic and Geophysical Research Institute of HAS in Sopron for a half year. The sensor is an OYO McSEIS TM-202 vertical seismometer with 2 Hz natural frequency. The chosen STA/LTA ratio is 2. In this time interval about 100 earthquakes or microseismic activities were recorded. A program was developed for an IBM computer to process the recorded earthquake data. The dates of earthquakes can be listed and the recorded data plotted by the program.

Though the FR-1 instrument was developed for recognizing and recording of near earthquakes, in the following the plots of some large far earthquakes are demonstrated to illustrate the operation of the instrument. Figure 5 shows the first part

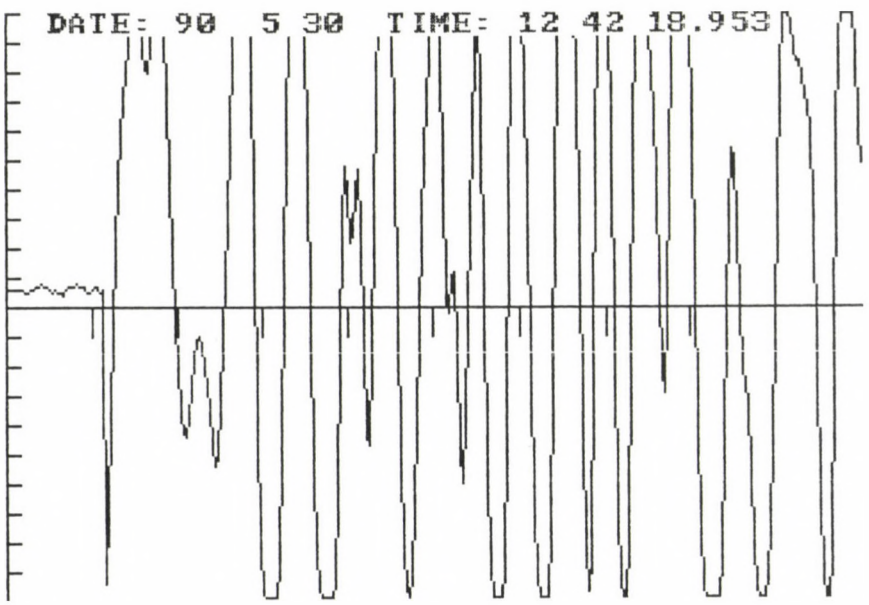


Fig. 5. The Rumanian earthquake (Richter magnitude 6.8)

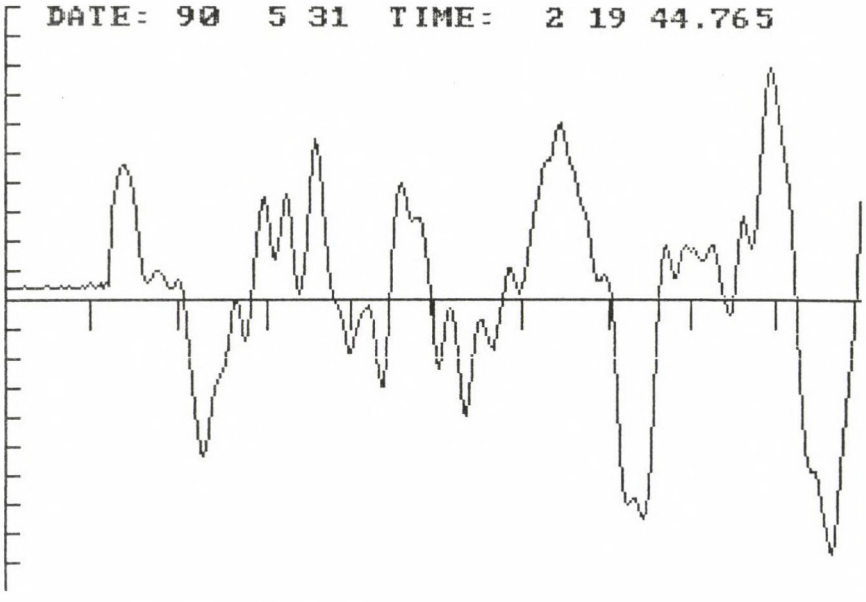


Fig. 6. An afterquake of the Rumanian earthquake

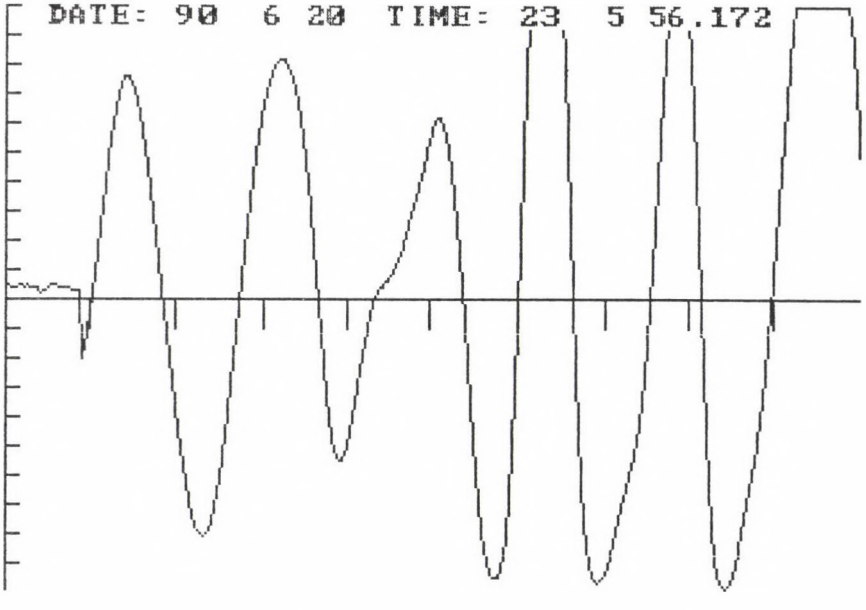


Fig. 7. The Iranian earthquake (Richter magnitude 7.3-7.7)

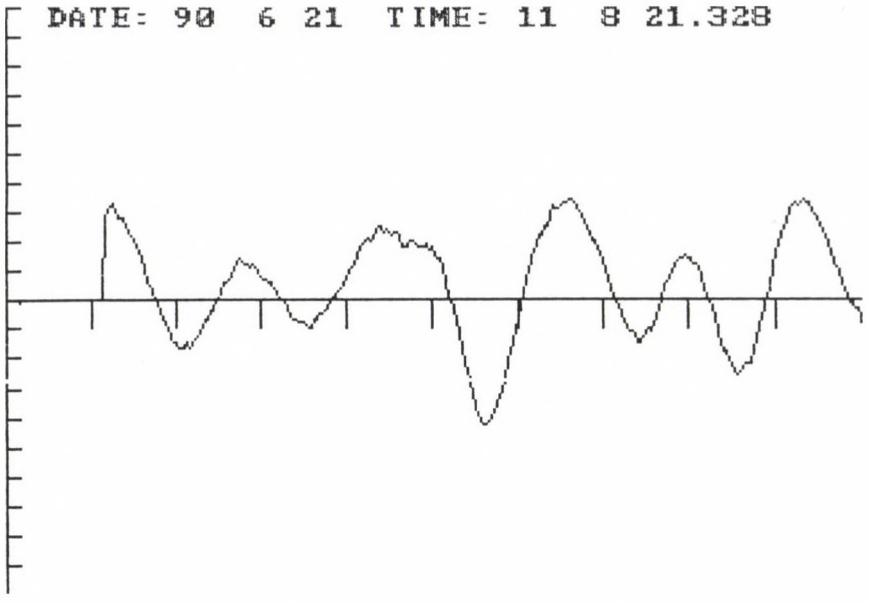


Fig. 8. An afterquake of the Iranian earthquake (Richter magnitude 6.5)

of the plot of the Rumanian earthquake (Richter magnitude 6.8). From this figure it can be seen that the very sensitive amplifier was saturated at this strong but not near earthquake. Figure 6 shows an afterquake of the Rumanian earthquake on the following day. Figure 7 shows the beginning of the Iranian earthquake (Richter magnitude 7.3-7.7) and Fig. 8 shows one of its afterquakes (Richter magnitude 6.5). The date and time written on figures are the arrival times of the first wave. The amplitudes of the curves are in relative units, the time is in seconds. Thus, the figures show only a 10 seconds part of the recorded 40 seconds long curve.

REFERENCES

Allen R V 1978: Bull. Seism. Soc. Am., 68, 1521-1532.
Gofort T, Herrin E 1981: Bull. Seism. Soc. Am., 71, 1351-1360.
Stewart S W 1977: Bull. Seism. Soc. Am., 67, 433-453.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 57-68 (1991)

THE USE OF ROBUST ESTIMATION IN THE GEODETIC DATA PROCESSING

J Somogyi and J Kalmár

Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H-9401 Sopron, POB 5, Hungary

[Manuscript received October 30, 1990]

The LSM yields an unbiased estimation in the case when the measuring data are normally distributed. Some alternatives to the least squares estimation are available when the measurements are contaminated by gross errors or they are not fully normally distributed. The paper deals with different robust estimation methods for geodetical application.

Keywords: distributions; Hampel's estimator; Huber's function; likelihood function; M-estimation; object function; Sopron method; W-estimation

1. INTRODUCTION

In geodetic data processing, the least squares methods is most widely used. This estimation process connected with Gauss implies the supposition about the normal distribution of the measurement data and leads to a very simple mathematical solution of the problem. Nevertheless, the least squares estimator is very sensitive to gross errors caused by outliers and other contaminated values. They result in uncertain estimators and in erroneous consequences of the analysis. That is why robust methods have been developed to reduce the effect of outliers and contaminations in the estimation.

In the last decades significant results have been achieved in the development of robust methods and in theoretical problems connected with them.

Basic works are those by Tukey (1960) about the distribution of observations, by Huber (1964) about the M-estimation and by Hampel (1968) about the influence curve.

At present several robust estimation methods are at

disposal. Their practical use increases, however, slower than expected, as most of the users sticks to the least squares method supposing a normal distribution of the errors. This is especially true for geodesy, nevertheless, there were some experiments to introduce robust methods there, too, e.g. by Carosio (1979), Krarup et al. (1980), Caspary and Borutta (1987), Somogyi and Kalmár (1988), Xu Peiliang (1989) and Koch (1989).

In the present paper a method is described which belongs to the group of M-estimations for the determination of the location parameter, and it is compared to Hampel's (1974) method.

2. ESTIMATE FUNCTIONS

The estimate function $\varrho(x)$ expresses the mathematical connection between the observations x_i and the parameters Θ to be determined. If x_1, x_2 ... x_n is a random sample with the density $f(x-\Theta)$, then the location parameter Θ can be estimated using the popular maximum likelihood method. The logarithm of the likelihood function $L(\Theta)$ of the location parameter Θ is:

$$\ln L(\Theta) = \sum_{i=1}^{n} \ln f(x_i - \Theta) = - \sum_{i=1}^{n} \varrho(x_i - \Theta) \qquad (1)$$

where $-\ln f(x) = \varrho(x)$.

The aim of the M-estimation is to maximize the function $L(\Theta)$ by minimizing the object function $\sum \varrho(x-\Theta)$. This aim is reached by deriving the object function after Θ :

$$\frac{\partial}{\partial \Theta} \sum_{i=1}^{n} \varrho(x_i - \Theta) = \sum_{i=1}^{n} \psi(x_i - \Theta) \qquad (2)$$

where $\varrho'(x) = \psi(x)$.

The solution

$$\sum_{i=1}^{n} \psi(x_i - \Theta) = 0 \qquad (3)$$

is the maximum likelihood estimator of the parameter Θ. An unambiguous solution is obtained if Eq. (1) is convex. The weight function $w(x-\Theta)$ of the function $\psi(x-\Theta)$ is defined as follows:

$$w(x-\Theta) = \frac{\psi(x-\Theta)}{(x-\Theta)} \quad . \tag{4}$$

The M-estimation gets a robust one if the object function $\varrho(x)$ is chosen to have robust properties, i.e. it is unsensitive for deviations from the supposed distribution. The solution would be easy if the distribution of the random sample would be known, as in such a case the generalized maximum likelihood estimators would lead in every case to the most acceptable solution. To show this three solutions are presented for different types of distributions.

a) Normal distribution

$$\varrho(x) = x^2/2 + c, \quad \psi(x) = x, \quad \sum_{i=1}^{n} (x_i - \Theta) = 0$$

gives the least squares estimator $\hat{\Theta} = \frac{\Sigma x}{n}$.

b) Double exponential distribution

$$\varrho(x) = |x| + c, \quad \psi(x) = -1 \text{ if } x < 0, \psi(x) = 1 \text{ if } x > 0,$$

$\Sigma\psi(x_i - \Theta) = 0$ gives the median of the sample $\hat{\Theta}$.

c) Cauchy-distribution

$$\varrho(x) = \ln(1+x^2) + c, \quad \psi(x) = 2x/(1+x^2), \quad \sum_{i=1}^{n} \psi(x_i - \Theta) = 0$$

this is a non-linear system of equations and $\hat{\Theta}$ = can be obtained by iteration.

Figure 1 shows the functions ψ of the three distributions. The least squares method is fully influenced by the outliers and by other contaminations, i.e. by deviations from the supposed distribution, whereas in the case of the other two distributions their effect is much smaller. The functions ψ of the second and third examples have an upper

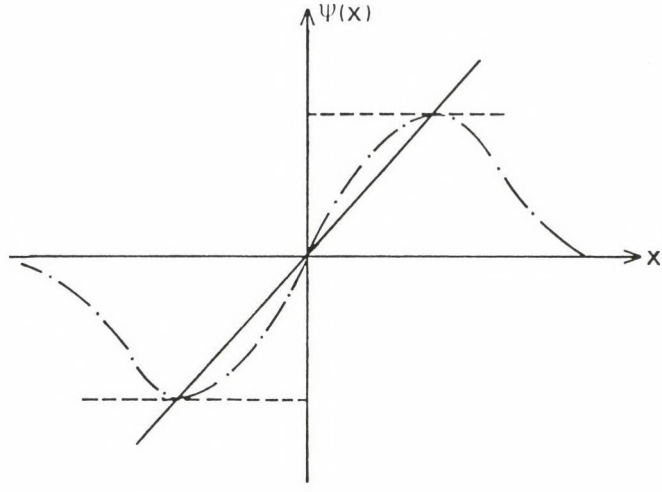


Fig. 1. $\psi(x)$ curves of normal, double exponential and Cauchy distributions

limit, furthermore the third is descending and approaches asymptotically zero.

As the least squares method does not yield a reliable estimator if the sample distribution has a long tail being a common case in the surveying practice, it is recommendable to use robust methods which are effective both if the distribution is normal and if the distribution has a long tail. Starting from this point Huber proposed a robust function $\varrho(x-\Theta)$, which has a density function normal in its central part and double exponential at the tails. Huber's function ψ is the following:

$$\begin{aligned} \psi(x) &= x \quad \text{if} \quad -k < x < k \\ &= -k \quad \text{if} \quad x \leqslant -k \\ &= k \quad \text{if} \quad x \geqslant k \end{aligned} \tag{5}$$

where $x = (x-\Theta)$.

Figure 2 shows the ψ and w functions for Huber's estimator.

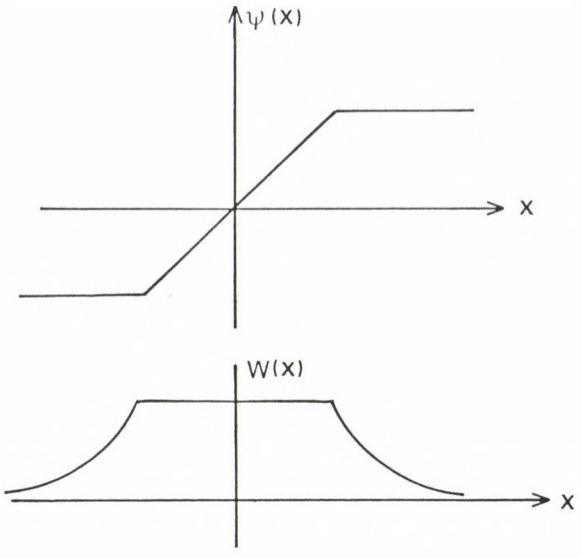


Fig. 2. $\psi(x)$ and w(x) functions for Huber's estimator

3. HAMPEL'S ESTIMATOR

Huber's estimator is a so-called no-descending one due to the form of the function ψ. It is characteristic for these solutions that the estimator is influenced by all observations. In order to exclude outliers, Hampel (1974) proposed a redescending estimate function which ensures in addition to its robust character the exclusion of strongly deviating observations. Hampel's ψ and w functions are the following:

$$\begin{aligned} \psi(x) &= x && \text{if} \quad |x| < a \\ &= a \ \text{sign}\ (x) && \text{if} \quad a \leqslant |x| < b \\ &= a(c-|x|)/(c-b) \ \text{sign}\ (x) && \text{if} \quad b \leqslant |x| < c \\ &= 0 && \text{else} \end{aligned} \tag{6}$$

$$\begin{aligned} w(x) &= 1 && \text{if} \quad |x| < a \\ &= a/|x| && \text{if} \quad a \leqslant |x| < b \\ &= a(1-c/|x|)/(b-c) && \text{if} \quad b \leqslant |x| < c \\ &= 0 && \text{else} \end{aligned} \tag{7}$$

where $x = (x - \Theta)$.

Figure 3 shows the ψ and w functions of Hampel's estimator.

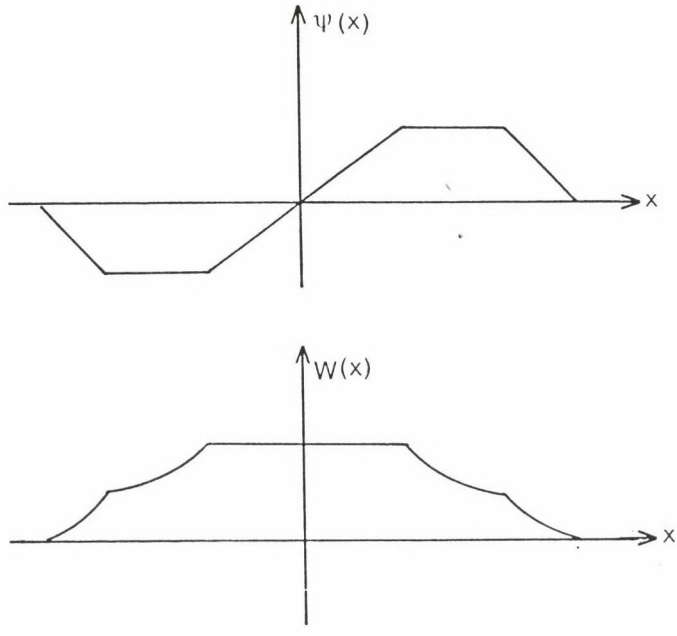


Fig. 3. ψ (x) and w(x) functions for Hampel's estimator

As function ψ is descending, the corresponding object function is not convex. That is why problems of convergence may appear in numerical solutions. The tuning constants a, b and c are to be chosen in accordance with the actual problem. Hampel's propositions are a=2, b=4, c=8.

3.1 Working formulas for Hampel's method for the adjustment of indirect observations

The task is to estimate the parameter vector Θ (with r elements) on the basis of

$$\underline{V} = A\Theta - \underline{l} \tag{8}$$

The solution is:

$$\sum_{i=1}^{n} \varrho(x_i) \longrightarrow \min \qquad (9)$$

instead of Eq. (9), the following system of equation is solved:

$$\sum_{i=1}^{n} \frac{\partial \varrho(x_i)}{\partial \Theta j} = 0 \qquad j = 1, \ldots r \; . \qquad (10)$$

One gets:

$$\sum_{i=1}^{n} \frac{\partial \varrho(x_i)}{\partial x_i} \cdot \frac{\partial x_i}{\partial \Theta j} = \sum_{i=1}^{n} a_{ij} \frac{\partial \varrho(x_i)}{\partial x_i} = \sum_{i=1}^{n} a_{ij} \psi(x_i) = 0 \qquad (11)$$

with $x = (x - \Theta)$.

By substituting ψ and x_i one has:

$$B \quad \Theta = \underline{d} \qquad \text{where} \qquad B = A^T P A \qquad \text{and}$$

$$d = A^T P \underline{l} + a A^T Q \; \text{sign}\,(x) \; \text{further}$$

P and Q are diagonal weight matrices which fulfil the equations:

$$\begin{aligned} p_{ii} &= 1 && \text{if} && |x_i| < a \\ &= a/(b-c) && \text{if} && b \leqslant |x_i| < c \\ &= 0 && \text{else} \end{aligned} \qquad (12)$$

$$\begin{aligned} q_{ii} &= -1 && \text{if} && a \leqslant |x_i| < b \\ &= c/(b-c) && \text{if} && b \leqslant |x_i| < c \\ &= 0 && \text{else.} \end{aligned} \qquad (13)$$

The above algorithm yields Huber's solution if $b = c = \infty$.

4. THE SOPRON METHOD

As a variant of Hampel's solution, a descending estimator function has been proposed by us, where the density distribution is at the tails double exponential.

The functions ψ and w of the estimator are the following:

$$\begin{aligned}
\psi(x_i) &= x && \text{if } |x_i| < a \\
&= a(x+(a-2b)\ \mathrm{sign}(x_i))/2/(a-b) && \text{if } a \leqslant |x_i| < b \\
&= a\ \mathrm{sign}(x_i)/2 && \text{if } b \leqslant |x_i| < c \\
&= 0 && \text{else}
\end{aligned} \tag{14}$$

$$\begin{aligned}
w(x_i) &= 1 && \text{if } |x_i| < a \\
&= a(1+(a-2b)/|x_i|)/2/(a-b) && \text{if } a \leqslant |x_i| < b \\
&= a/2/|x_i| && \text{if } b \leqslant |x_i| < c \\
&= 0 && \text{else}
\end{aligned} \tag{15}$$

where $x = (x-\Theta)$.

Figure 4 shows the function ψ and w of the estimator.

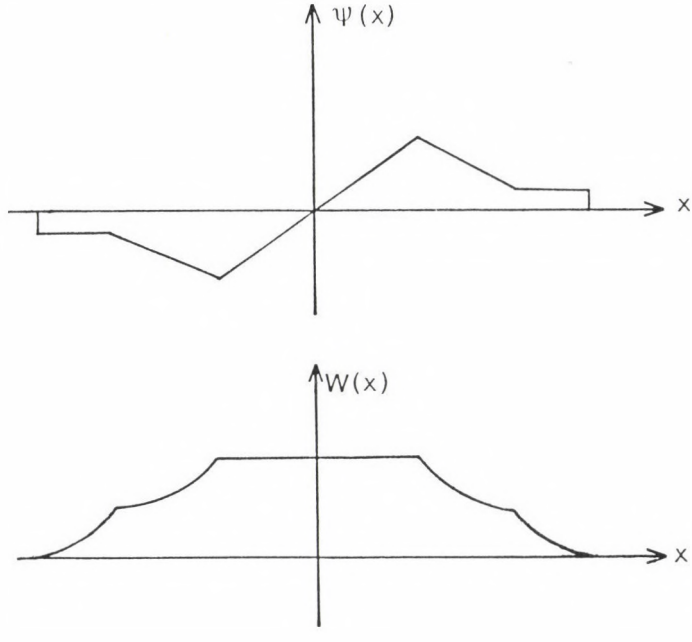


Fig. 4. ψ (x) and w(x) functions for Sopron estimator

4.1 Working formulas of the Sopron method for the adjustment of indirect observations

By substituting ψ and x_i into the system of equations:

$$\sum_{i=1}^{n} a_{ij}\psi(x_i) = 0 \qquad j = 1, \ldots, r \tag{16}$$

one gets:

$$B' \Theta = d';$$

where $B' = A^T P' A$ and

$d' = A^T P' l + a A^T Q' \operatorname{sign}(x)$

where the diagonal weight matrices P' and Q' fulfil the following:

$$\begin{aligned} p'_{ii} &= 1 && \text{if} \quad |x_i| < a \\ &= a/[(a-b)2] && \text{if} \quad a \leqslant |x_i| < b \\ &= 0 && \text{else} \end{aligned} \tag{17}$$

$$\begin{aligned} q'_{ii} &= (2b-a)/[(a-b)2] && \text{if} \quad a \leqslant |x_i| < b \\ &= -1/2 && \text{if} \quad b \leqslant |x_i| < c \\ &= 0 && \text{else} \end{aligned} \tag{18}$$

where $x = (x - \Theta)$.

5. W-ESTIMATION

An alternative for the M-estimations is the W-estimation which leads to the weighted average as to a solution and the weights depend on the data. Thus the robust estimation can be reduced to the re-weighted iterative least squares adjustment. Each weight depends on the deviations in the previous iteration. Starting from a convenient approximation, the iteration is to be continued till the solution converges with a prescribed accuracy. In the iterations, weights refer to single data, e.g. outliers or contaminated values get zero or low weights.

The weights deduced from the M-estimators in the previous chapters are in the method of the re-weighted least squares the following:

Θ is determined from the equations $V = A\Theta - l$ so that the condiction

$$\sum_{i=1}^{n} p_i v_i^2 \rightarrow \min$$

be also fulfilled, i.e. by solving the system of equations $A^T P A \Theta = A^T P l$, where P is the diagonal weight matrix and $p_{ii} = p_i$.

By substituting the previously deduced weight functions, the following formulas are obtained:

1. Hampel's method

$$\begin{aligned} p_{ii} &= 1 && \text{if} \quad |v_i| < a \\ &= a/|v_i| && \text{if} \quad a \leqslant |v_i| < b \\ &= a(1-c/|v_i|)/(b-c) && \text{if} \quad b \leqslant |v_i| < c \\ &= 0 && \text{else} \end{aligned} \tag{19}$$

2. The Sopron method

$$\begin{aligned} p'_{ii} &= 1 && \text{if} \quad |v_i| < a \\ &= a(1+(a-2b)/|v_i|)/[2(a-b)] && \text{if} \quad a \leqslant |v_i| < b \\ &= a/(2|v_i|) && \text{if} \quad b \leqslant |v_i| < c \\ &= 0 && \text{else} \end{aligned} \tag{20}$$

where naturally $v_i = (x_i - \Theta)$.

In order to avoid computational difficulties, it is advisable to start both in the cases of the M- and W-estimations from suitable approximative values. As approximative values, the results of a least-squares adjustment can be used or even more advantageously the sample median as by it the outliers can be already found and the robust estimation following the first phase may concentrate on the elimination of the effects of smaller contaminations.

6. NUMERICAL EXAMPLE

Finally we compare the robust estimation methods described in the previous sections. The x and y coordinates are obtained from the X and Y coordinates of a simulated point field consisting of 125 points so that the former are rotated and multiplied with a scale factor. For the determination of the transformation parameters 15 points were chosen. The coordinates x, y of the points were contaminated by normally distributed N(0,3) errors and in 20 percent, by gross errors. The results of the different robust methods are presented in Table I. Using the transformation parameters from the different

Table I

	H		S		W_H		W_S	
	v_x	v_y	v_x	v_y	v_x	v_y	v_x	v_y
1	-5	1	-5	0	-6	-2	-6	-2
2	1	-4	0	-4	0	-4	0	-4
3	44	-8	44	-8	47	-8	47	-8
4	-5	1	-6	2	-8	4	-8	4
5	1	-5	0	-5	-1	-4	-1	-4
6	-8	11	-8	12	-9	15	-9	15
7	-2	-2	-2	-2	-3	-1	-3	1
8	-5	-82	-5	-81	-4	-79	-4	-79
9	-2	-9	-1	-8	0	-7	0	-7
10	13	3	13	3	14	3	14	3
11	5	-1	6	-1	8	-2	8	-2
12	-2	-3	-1	-4	1	-5	1	-5
13	163	-29	164	-30	165	-30	165	-30
14	3	6	3	6	2	4	2	4
15	0	2	1	1	1	-1	1	-1

H = Hampel's method
S = Sopron method
W_H = Weighted Hampel
W_S = Weighted Sopron

robust methods the X and Y coordinates of the remaining 110 points were computed. Mean square errors from the difference between original and computed coordinates are contained in Table II.

Table II

H		S		W_H		W_S	
Δx	Δy	Δx	Δy	Δx	Δy	Δx	Δy
1.58	1.54	1.58	1.46	2.28	2.64	2.28	2.64

It is evident that different solution yield similar results and the effect of the points with gross errors does not appear in other points.

REFERENCES

Carosio A 1979: Vermessung, Photogrammetrie, Kulturtechnik, 11/79, 293-297.

Caspary W, Borutta H 1987: Survey Review, 1/87, 29-45.

Hampel F R 1968: Contributions to the Theory of Robust Estimation. Ph.D. Thesis, University of California, Berkeley

Hampel F R 1974: Journal of the American Statistical Association, 69, 383-393.

Huber P J 1964: Annales of Mathematical Statistics, 35, 73-101.

Koch K R 1989: Bayesian inference with Geodetic Applications (Lecture Notes). Institut für Theoretische Geodäsie der Rheinischen Friedrich-Wilhelms-Universität Bonn, 142-152.

Krarup T, Juhl J, Kubik K 1980: In: International Archives of Photogrammetry, Vol. 23 (B3), 369-378.

Somogyi J, Kalmár J 1988: Allgemeine Vermessungsnachrichten, 4/88, 141-146.

Tukey J W 1960: A Survey of Sampling from Contaminated Distributions. Contributions to Probability and Statistics, I Olkin, ed., Standford University Press, Standford, California

Xu Peiliang 1989: Bulletin Géodésique, 63, 237-252.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 69-76 (1991)

STUDYING THE ATTENUATION OF ELASTIC WAVES IN HOMOGENEOUS MATERIALS UNDER HIGH PRESSURES BY USING THE METHOD OF FREQUENCY ANALYSIS

V A Kalinin, I V Zhukov, G A Yefimova

Institute of Physics of the Earth, USSR Academy of Sciences, 123810 Moscow, Gruzinskaya 10, USSR

[Manuscript received June 20, 1988]

Attenuation characteristics of ultrasonic waves for homogeneous materials were investigated under high pressure in a piston-type apparatus. The approach is based on Fourier analysis of a pulsed signal taking into account the scale factor. For a test of the method frequency-attenuation relationships were determined in some metals (copper, brass) undergoing pressure up to 0.95 GPa.

Keywords: attenuation; elastic waves; high pressure; piston

A high sensitivity of ultrasonic pulse-frequency spectra to variations in physical properties of materials prompted the development of spectral mehods for determining the attenuation of elastic waves in defectoscopy and in solid state physics (Truell and Elbaum 1969, Merkulov and Tokarev 1970). In studying elastic wave attenuation in high-pressure setups the traditional methods must be modified, because in this case the ultrasonic wave channel respresents a system of abutting elements and the recorded waveform turns out to be extremely complex for interpretation. Moreover, in the operating frequency range one should assess the effect of the scale factor when using specimens of limited dimensions. The proposed methodological solution, which belongs to the category of so-called variable-distance methods (Brazhnikov 1965), permits us to determine the attenuation coefficient of compressional waves α_p in the frequency function f for various materials sufficiently homogeneous with respect to structure and substance composition in the case of quasi-hydrostatic compression in "piston-type" apparatus (Volarovich et al. 1974).

First our studies of attenuation by the proposed method were made for polycrystalline copper and brass (40 % Zn + 60 % Cu) loaded up to 1 GPa pressure. Ultrasonic attenuation in these metals under atmospheric pressure was studied in some works (Papadakis 1969, Kobelev and Soifer 1985). The effect of high pressure on ultrasonic attenuation in metals due to methodological difficulties has not yet been adequately studied. In some isolated works (Pilipenko 1977) it was noted that attenuation depends on the type of crystal lattice, the presence of defects, the properties of the medium transmitting pressure and certain other factors.

EXPERIMENT

Determination of α_p at different frequencies was achieved by comparing the amplitudes of the spectral components of ultrasonic pulses passing through specimens of one and the same metal having a different length and alternately placed in the system. The principal computation formula for α_p at a given frequency f_i has the form:

$$\alpha_p = \frac{1}{l_n - l_m} \ln \frac{A_m}{A_n} \qquad (1)$$

where l_n and l_m are the lengths of specimens placed in the high-pressure cell; A_n and A_m are the correspoding amplitudes of spectral components at the frequency f_i.

To estimate the magnitude of an error in determining α_p, which is due to a difference in specimen lengths, we studied copper and brass specimens of 19.8, 14.9, 12.0 and 8.0 mm length (all specimens having the form of a cylinder of 17.5 mm diameter); thus, we had 6 combinations of $l_n - l_m$.

Ultrasonic investigations were conducted on standard equipment; the block-diagram of the setup is described in (Volarovich et al. 1974). For the excitation and reception of ultrasound we used resonance pickups made of lead zirconate-titanate (PZT).

Registration of pulses in the system with different-length

metal specimens placed in it was achieved under pressures ranging from 0.15 to 1.0 GPa. Loading was made in a stepped regime. An adequate quality and stability of acoustic contacts between ultrasonic channel elements, as was necessary for comparison of the wave forms, were guaranteed by the condition of a complete closure of "gaps" formed by roughnesses at contact boundaries: this condition was fulfilled in loading the system up to over 0.15 GPa pressures. The given criterion was established experimentally.

AMPLITUDE-FREQUENCY ANALYSIS

General theoretical principles (2) pertaining to the spectral method for determining α_p require that for the purpose of analysis a complete pulse should be singled out from the moment of arrival of the front of a direct compressional wave till a complete attenuation of oscillations against a noise background. In practice this requirement is realizable only in limited cases, particularly, in a boundless mass of a homogeneous isotropic material. The presence of acoustic boundaries inavoidably leads to interference effects introducing complications in the resultant waveform. Thus, under conditions of the ultrasonic channel of the high-pressure setup used by us we usually succeed in registering an oscillation free from distortion, which corresponds to a direct compressional wave, in a time window from t=0 to $t=\tau$. Apparently, for the purpose of analysis we should use only this part of the pulse, discarding the uneasily interpretable pulse "tail". The spectra of incomplete pulses obtained in the frequency analysis, as is known from works in the field of seismology (Berzon et al. 1962), with respect to component width and amplitudes are different from that of a pulse (corresponding to a given type of wave) analysed from the moment of first arrival till a complete attenuation. To estimate errors in determining frequency component amplitudes and, consequently, α_p, we studied the effect of the width of time window on quantitative spectral parameters. Proceeding

from a reasonable physical assumption that in using resonance ultrasonic transducers the pulse tail must have a form similar to an attenuating sinusoidal signal at the fundamental resonance harmonic frequency of the emitter (and its identical receiver), the amplitude spectrum of a complete pulse was constructed on the basis of the formula:

$$S(f)=\sqrt{S_1^2(f)+S_2^2(f)+2S_1(f)S_2(f)\cos[\Phi_1(f)-\Phi_2(f)-2\pi f\cdot\tau]}, \quad (2)$$

where $S_1(f)$ and $\Phi_1(f)$ are the amplitude and phase spectra of the head portion of the pulse, corresponding to the time window τ, both spectra being determined directly by analysing the pulse oscillogram; $S_2(f)$ and $\Phi_2(f)$ are the amplitude and phase spectra of the attenuating portion of the pulse represented in an analytical form by an attenuating sinusoid with the parameters (initial amplitude, fundamental frequency and time attenuating coefficient) likewise to be determined from the oscillogram.

The pulse-frequency spectra computed after the formula (2) in a copper specimen for time windows τ = 1.3, 2.0, 2.4 and 3.0 μs and the frequency-dependent spectral ratios S_1/S_2 obtained for these τ are presented in Figs 1, 2, respectively (for 0.5 GPa pressure). As follows from the above estimates, the maximum error in analysis, i.e. up to 25 %, occurs at low amplitudes of the spectral components, and that at the frequencies in the vicinity of the resonant frequency of the pickups, i.e. the fundamental frequency of the attenuating sinusoid (4.2 MHz), it is up to 20 %. In the rest, sufficiently wide frequency interval the amplitudes can be determined to high accuracy and used for the purpose of a quantitative interpretation, including also the computation of α_p from the formula (1).

RESULTS

The obtained α_p versus $\Delta l_{nm} = l_n - l_m$ at all of the frequencies under study are observed to be practically

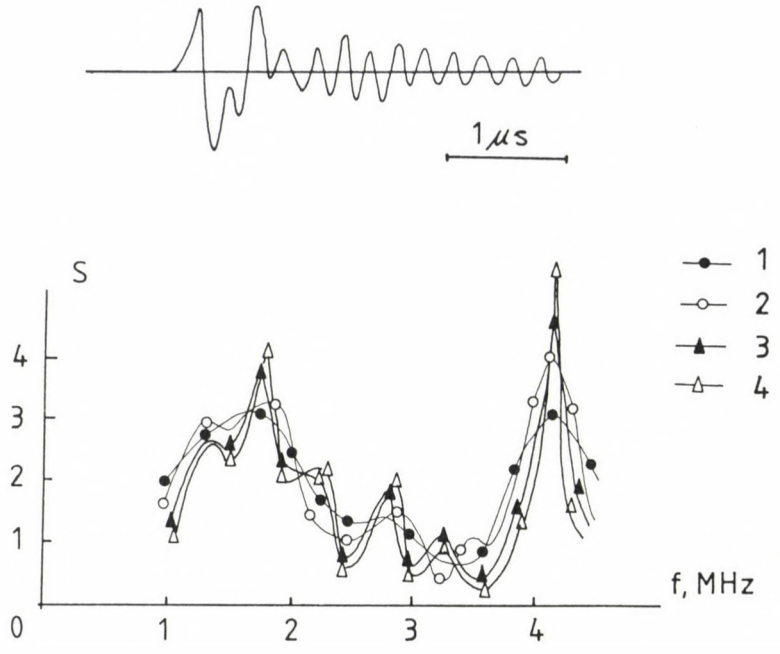


Fig. 1. Form and amplitude spectra of pulses recorded in the system with specimens of polycrystalline copper (length prior to loading 19.8 mm) under 0.5 GPa pressure. The S(f) spectra computed after the formula (2) for different time windows: 1.3 μs - 1; 2.0 μs - 2; 2.4 μs - 3; 3.0 μs - 4. The amplitude is given in arbitary units

insensitive to variations of Δl_{nm}, i.e. $\alpha_p(\Delta l_{nm}) = \text{const}$. Consequently, the resultant waveform at the selected frequencies is not subject to perturbations due to restricted specimen dimensions and α_p to be determined thus characterize the properties of the mass of the material under investigation.

Quasi-hydrostatic pressure was found to alter α_p in the investigated copper within the frequency interval where spectral component amplitudes are determinable with a minimum error (1.0-3.1 MHz); moreover, with an increase in pressure from 0.15 to 0.50 GPa α_p is observed to decrease in copper at all frequencies under study averagely by 50 %. As loading is further increased from 0.50 to 1.0 GPa α_p display a tendency for growth, Fig. 3. Possibly this effect is associated with the appearance of microplastic zones and needs to be interpreted

from the viewpoint of a dislocation model (Merkulov 1963, Strel'tsov 1983).

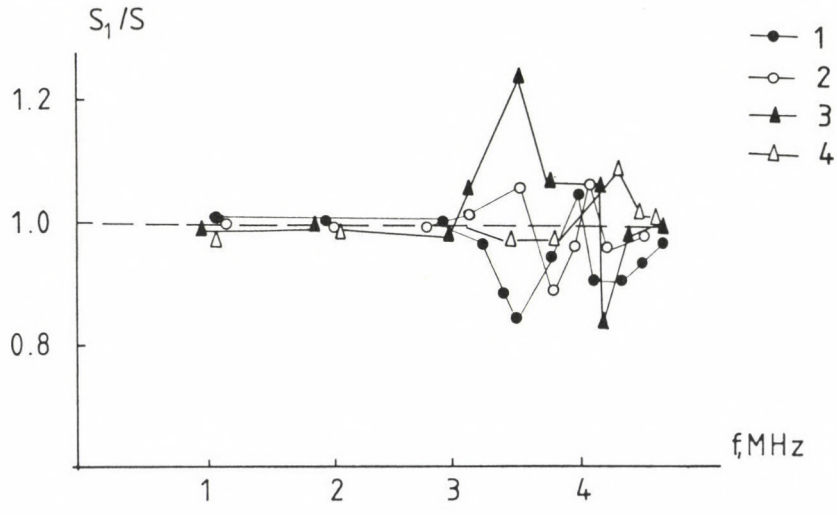


Fig. 2. The ratio of the spectra $S_1(f)$ of an incomplete pulse for different τ: 1.3 μs - 1; 2.0 μs - 2; 2.4 μs - 3; 3.0 μs - 4 to the corresponding complete spectra S(f) computed after the formula (2), which characterizes the accuracy of the quantitative frequency analysis

As regards brass, variations in α_p with pressure are less distinct and do not go beyond accuracy limits in determinig α_p($\pm$10 %). This could be expected, because in brass, similarly as in a solid substitution solution where Zn atoms occupy the places of Cu atoms in the FCC-lattice, plastic properties are expressed much weaker than is the case with pure copper.

Thus, the qualitative character of α_p versus pressure relationships obtained by us confirms the validity of the chosen methodological approach. Data on elastic waves attenuation under high pressures in metal, apart from an independent scientific significance, may prove useful in studying the impact of high pressure on substances with a more complex composition and structure, particularly rocks. On the basis of the above investigations we may choose brass as a reference material, since the effect of high pressure on its

acoustic characteristics is expressed much weaker compared to the majority of rocks.

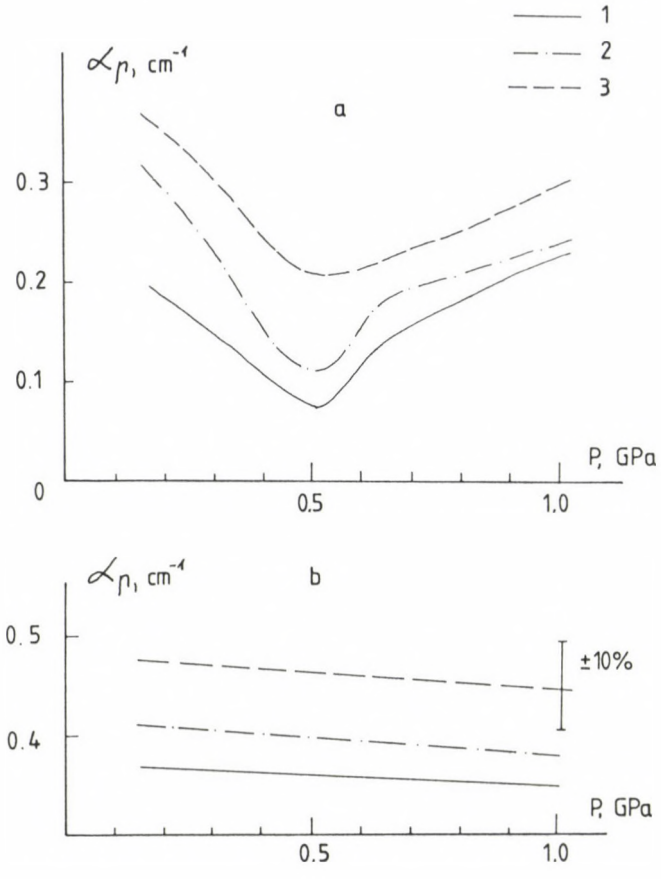


Fig. 3. Attenuation coefficient of compressional waves α_p versus pressure in copper (a) and brass (b) for the frequency components 1.0 MHz - 1; 1.5 MHz - 2 and 3.0 MHz - 3

REFERENCES

Berzon I S, Yepinat'eva A M, Pariiskaya R N, Starodubovskaya S P 1962: Dynamic Characteristics of Seismic Waves. USSR Academy of Sciences Press, M

Brazhnikov N I 1965: Ultrasonic Methods. "Energiya" Publishers, M-L

Merkulov L G ed. 1963: Ultrasonic methods for investigation of

dislocations. A collection of papers (in Russian). Foreign Literature Press, M

Merkulov L G, Tokarev V A 1970: Defektoskopiya (Defectoscopy), No. 4, 3-11.

Kobelev N P, Soifer Ya M 1985: FTT, 27, No. 4, 1256-1259.

Papadakis E P 1969: In: Physical Acoustic, Ed. by W Mason, "Mir" Publishers, M, Vol. 4, Part B, 317-381.

Pilipenko N P 1977: Working out the technique and investigating the effect of hydrostatic pressure on attenuation of ultrasound in metals. Author's abstract for candidate's thesis, Donetsk

Strel'tsov V A 1983: High-Pressure Physics and Technology, No. 11, 21-24.

Truell R Ch, Elbaum B 1969: Chick Ultrasonic Methods in Solid State Physics. Academic Press, N-Y, London

Volarovich M P, Bayuk E I, Levykin A I, Tomashevskaya I S 1974: Physico-Mechanical Properties of Rocks and Minerals Under High Pressures and Temperatures. "Nauka" Publishers, M

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 77-87 (1991)

ELASTIC MODULI AND ULTRASONIC WAVE VELOCITIES IN A TWO-PHASE MEDIUM UNDER HIGH PRESSURES

I O Bayuk, R M Nasimov, V A Kalinin, A I Levykin

Institute of Physics of the Earth, USSR Academy of Sciences, 123810 Moscow, Gruzinskaya 10, USSR

[Manuscript received June 20, 1988]

Ultrasonic velocities and density of the two-phases media (quartz sand mixed up with stearin) were experimentally determined at high pressure and temperature (up to 1.2 GPa and 400 K). Elastic moduli of the mixture are calculated by the use of the generalized singular approach method based on the physically more correct model of strain and stress fields in elastic media, then by the Voight-Reuss-Hill method. The theoretical results are in agreement with the experimental data.

Keywords: elastic moduli; high pressure; two-phase media; ultrasonic wave velocity

In determining the substance composition of the Earth there arises the necessity of theoretically determining the elastic properties of rocks from the properties of the minerals composing them. To achieve this purpose the Voight-Reuss-Hill method (Belikov 1970) is widely used in geophysics. The method is based on the following inequalities proved by Hill (1952):

$$K_R \leqslant K \leqslant K_V$$
$$\mu_R \leqslant \mu \leqslant \mu_V \qquad (1)$$

where K and μ are the bulk and shear moduli of a heterogeneous body.

The moduli K_V, μ_V were computed after the Voight method (with an assumption of the constancy of the deformation field) and K_R, μ_R - after the Reuss method (with an assumption of the constancy of the stress field). On the basis of (1) Hill proposed that elastic moduli of a heterogeneous body be

computed from the fomulae

$$K_{VRH} = \frac{K_V+K_R}{2} , \qquad \mu_{VRH} = \frac{\mu_V+\mu_R}{2} . \tag{2}$$

If we denote the true value of the bulk and shear modulus X_{true}, then by virtue of the inequalities (1) $X_{true} = X_V - \delta_V^X$ or $X_{true} = X_R + \delta_R^X$. Here δ_V^X is an error in determining the modulus, which results from an assumption about the constancy of the deformation field, i.e. the Voight method. Similarly, δ_R^X is an error in determining the modulus, resulting from an assumption of the constancy of the stress field, i.e. the Reuss method. According to the definition,

$$X_{VRH} = X_{true} + (\delta_V^X - \delta_R^X)/2 . \tag{3}$$

In the case of low values of δ_V^X and δ_R^X X_{VRH} is close to X_{true}, which is valid for the case of weakly-contrasting composite media, the anisotropy of elastic properties of which is insignificant. However, in the lithosphere and astenosphere there are present partial melts, the elastic moduli of which are much smaller than the corresponding moduli of the solid component. In such a medium the melt fills in interstices, spreads out along grain boundaries in the form of films. Under the impact of external loading the melt is subject to a greater deformation than are the solid components. This brings about a strongly heterogeneous deformation field. Moreover, it is to be expected in this case that the stress field will be close to a homogeneous one. Therefore, $\delta_R^X \leqslant \delta_V^X$ and $X_{VRX} \approx X_{true} + \delta_V^X/2$. Thus, in this case the VRH method yields overstated values of the moduli. On the other hand, in the case of liquid ($K \neq 0$, $\mu = 0$) and gaseous ($K \approx 0$, $\mu = 0$) components the VRH method is inapplicable, since the formulae involved lead to a necessity of dividing by zero. Accordingly, there arises a need of using a theoretical method taking a more correct account of the distribution of the stress and deformation fields in a heterogeneous body and furnishing a possibility of estimating the effect of liquid and gaseous inclusions on its elastic

properties.

In the given work for determining elastic properties of such heterogeneous media use was made of the method of generalized singular approach (GSA) which was extended to the case of multicomponent systems, such as rocks (Shermergor 1977, Kalinin and Bayuk 1987). The GSA method is based on an analysis of a difference in the deformation fields which arise under similar boundary conditions in the investigated heterogeneous body and in a certain homogeneous body which is termed the body comparison. Deformations arising in a heterogeneous body satisfy an equilibrium equation by means of which is established a relationship between the magnitude of the deformation arising in each of the components and its value averaged for the volume. This relationship is determined by the elastic properties of the given component and the comparison body. Making use of a definition of the effective elasticity tensor

$$C^* = \langle \sigma \rangle^{-1} \langle \varepsilon \rangle \qquad (4)$$

where triangular brackets stand for averaging by volume, we obtain an expression for the effective elasticity tensor in the operator form. Obtaining the formula in a tensor form is possible solely by accepting an assumption about the constancy of the stress and deformation fields inside each of the components. In the case of a rock we presume the constancy of the respective fields in the grains of one and the same mineral having a similar orientation of crystallographic axes. Whereas the fields in the other minerals or in the grains of the same mineral, but differently oriented may be different.

In the case of isotropic rocks the principal computation formulae of the method concerned have the form

$$K^* = \left[\sum_{\nu=1}^{N} \frac{C_\nu}{K_\nu^* + b_K^C} \right]^{-1} - b_K^C, \quad \mu^* = \left[\sum_{\nu=1}^{N} \frac{C_\nu}{\mu_\nu + b_\mu^C} \right]^{-1} - B_\mu^C . \qquad (5)$$

Here C_ν is a volume concentration of the ν-th phase, N is the number of phases, K_ν^*, μ_ν^* are elastic moduli of the ν-th phase, K_C, μ_C are elastic moduli of the comparison body.

The method under discussion permits us at least partly to take into account the structural specifics of a heterogeneous body. This is achieved by the choice of a comparison body. In the choice of a comparison body we should proceed from physical assumptions about the distribution of the stress and deformation fields in a heterogeneous body. For composites of the "matrix-inclusion" type in the case of a soft matrix and hard inclusions a comparison body should be chosen with elastic moduli after Reuss, i.e. on an assumption of the constancy of the stress field. In an opposite case, i.e. "hard matrix - soft inclusions", a comparison body will be chosen with elastic moduli after Voight. And in the case of polycrystalline aggregates we presume that $K_C = K^*$, $\mu_C = \mu^*$, which corresponds to elastic fields in an aggregate, the components of which have an isometric form. An analysis of the formulae (5) enables us to draw a conclusion that for determining the elastic moduli of isotropic rocks instead of determining total elasticity tensors of the minerals composing the rock it is sufficient to know only the elastic moduli of polycrystalline aggregates of the corresponding minerals.

Elastic properties of multicomponent systems, as determined from the formulae (5), are in good agreement with the experiment. Figure 1 shows velocities of elastic waves versus pressure in granites, the latter being 4- and 5-component systems. Moreover, it can be seen that by taking account of isometric pores as a 5th component in the Westerley granite enables calculated values of elastic wave velocities to be brought closer to the experimental ones. In this rock pores account for 1.2 % of the volume (8). In making computations to secure an equality of elastic moduli of the comparison body quite effective proved the use of the iteration process. The values obtained after the VRH method were used as the initial approximation. The result was obtained after 3 iterations, as a maximum.

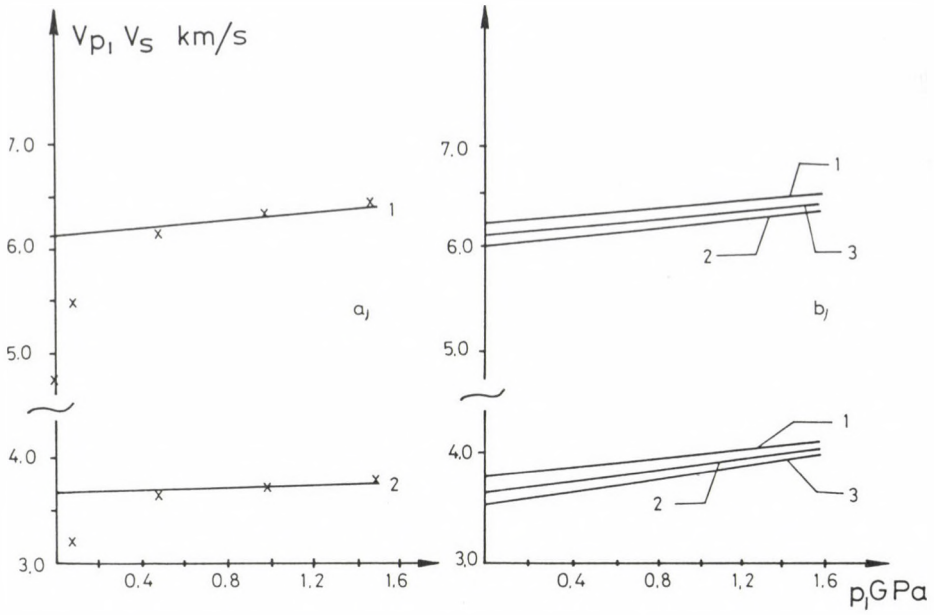


Fig. 1. Velocities of compressional and shear elastic waves in granites. a) composition (5) (in percent by volume): oligoclase-0.32, microcline-0.29, quartz-0.35, biotite-0.04; calculation after GSA method of velocities of compressional V_p (1) and shear V_S (2) elastic wave, x - experimental data; b) composition (Westerley granite): oligoclase-0.32, microcline-0.36, quartz-0.28, biotite-0.04, 1 - calculation after GSA method of the velocities of compressional V_p and shear V_S waves disregarding pores, 2 - calculation after GSA method of the velocities of elastic waves, account being taken of pores, 3 - linear approximation of experimental data (6), (7)

It is worth noting that a relatively rapid decrease of velocities under a declining pressure in the 0.2-0.3 GPa region is usually attributed to cracks. The GSA method departs from the experiment precisely in this region. It is yet unable to take account of the effect of jointing.

As a rule, elastic properties of the mineral skeleton of rocks, determined after the VRH and GSA methods, are close, this fact being due to a similarity of elastic properties of the minerals composing the rock. However, once there appears a component the elastic properties of which greatly differ from those of the other components a comparison of the computation results after the VRH method with experimental ones points to

their poor agreement. Thus were calculated the elastic properties of an artificial mixture consisting of a polyether matrix with glass microspheres of 210-297 μm diameter impregnated into it. The volume concentration of inclusions was subject to change. The ratio of elastic moduli of the mixture components is as follows: $K_I/K_M = 7$, $\mu_I/\mu_M = 49$; I is the microspheres, M is the matrix. Elastic moduli of the reference body were taken after Reuss. An analysis of Fig. 2 indicates a good agreement between the results obtained by the GSA method and the experiment, the latter being borrowed from Christen (1979). Furthermore, we may speak of practical inapplicability of the VRH method in determining elastic properties of mixtures whose components have sharply different properties.

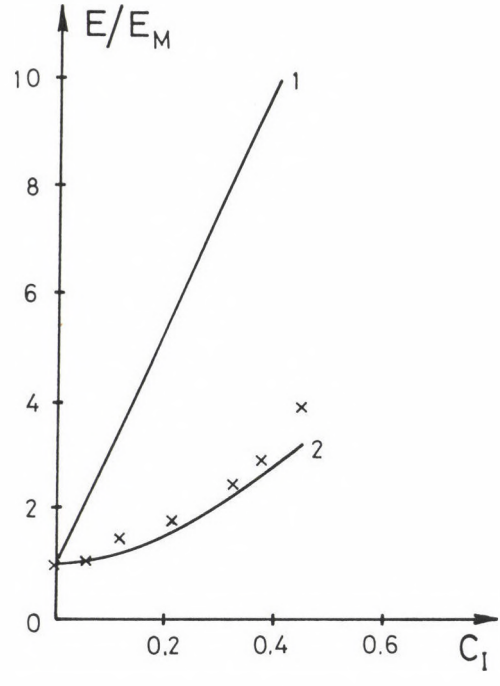


Fig. 2. Glass microspheres in a polyether matrix. E - Young's modulus, C_I - a volume concentration of microspheres in the matrix, 1 - calculation after VRH method, 2 - calculation after GSA method, x - experimental data (9)

A similar comparison was made for a mixture of quartz sand

with a fusible component - stearin - under pressures up to 1.2 GPa and temperatures - 20°C (under these conditions strearin beeing in solid state) and 95°C (stearin being in molten state up to 0.6 GPa pressure). The composites were obtained by intermixing a fractionated quartz sand with molten stearin. Average grain sizes of the sand fraction were 400, 200 and 50 μm. Stearin fills in voids between grains and forms films at the contacts of separate grains, i.e. quartz is well wetted with liquid stearin. The film thickness is estimated to be 10-20 μm. Experiments were conducted in a piston-type

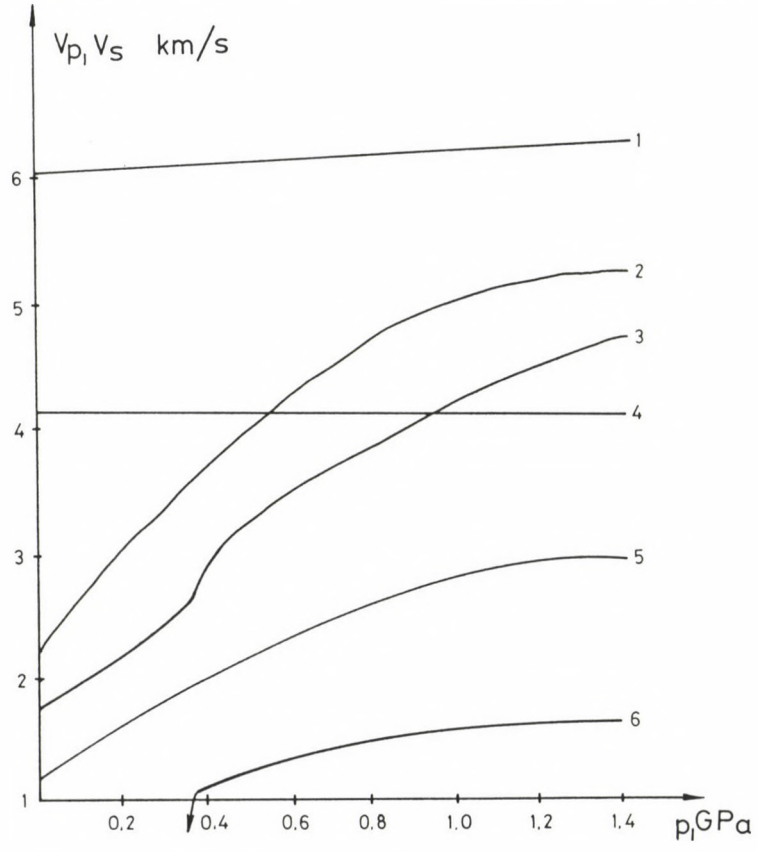


Fig. 3. Elastic wave velocities in quartz and stearin. Compressional wave V_p velocities: 1 - in quartz at 20°C and 95°C, 2 - in stearin at 20°C, 3 - in stearin at 95°C; shear V_s wave velocities: 4 - in quartz at 20°C and 95°C, 5 - in stearin at 20°C, 6 - in stearin at 95°C

apparatus with external heating (Levykin 1981). The selected ultrasonic frequency (1 MHz) permits the wave to be considered long.

An error in determining velocity values in stearin and mixtures has been estimated: for velocities of compressional waves not more than 3 %, for shear waves not more than 5 %. Figure 3 shows ultrasonic wave velocities in stearin obtained at the High-Pressure Laboratory, Institute of Physics of the Earth (Levykin 1981) as well as the corresponding velocities in quartz (Anderson 1968). Compressional wave velocities in stearing (20^{o}) are observed to be sharply rising with pressure, this being associated with the structure of organic matter consisting of large crystals. Under high pressures this is a compact medium.

If stearin fills in contacts between quartz grains, good results are to be expected if we take as elastic moduli of the comparison body the corresponding moduli after Reuss. A good agreement between calculations after GSA and the chosen comparison body was obtained for mixtures of 200 and 400 μm ($T = 20^{o}C$). It is to be noted that the experimentally determined velocities of compressional and shear waves in these mixtures are practically coincident (cf. Fig. 4). The fact of coincidence between calculated and experimental values in such a choice of the comparison body confirms our assumption about the structure of the mixture. However, under compression of the mixture with the fusible component being in the molten state the internal structure is subject to alteration, thus leading to the joining of quartz grains before stearin grows solid. Therefore, in calculations the elastic moduli of the reference body were taken after Voight (i.e. on an assumption of the homogeneity of the deformation field), as a result of which in the high-pressure region under $T = 95^{o}C$ there was obtained a good agreement between calculations and the experiment (cf. Fig. 5).

Comparing elastic wave velocities calculated after the VRH method with experimental ones in the low-pressure region, i.e. where elastic properties of the components are particularly

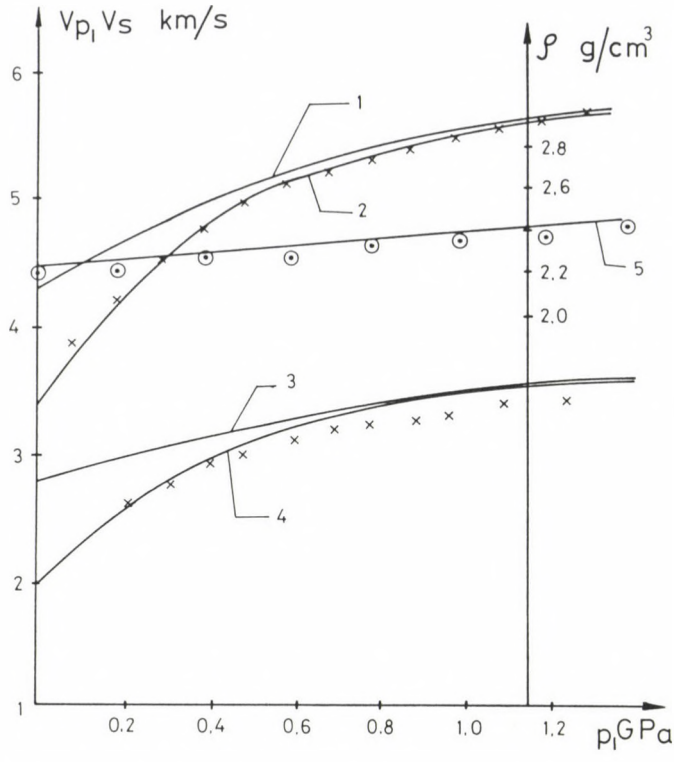


Fig. 4. Elastic wave velocities in a mixture of quartz (400 μm) and stearin at 20°C. Mass concentration of stearin in the mixture - 0.15. Compressional wave velocities: 1 - calculation after VRH method, 2 - calculation after GSA method; shear wave velocities: 3 - calculation after VRH method, 4 - calculation after GSA method, x - experimental velocity values, 5 - calculated density of the mixture, $\varrho = \sum_{\nu=1}^{n} \varrho_\nu c_\nu$, ⊙ - experimental density values

contrasting, we may draw a conclusion about practical inapplicability of the VRH method in this region.

It follows from what has been said above that for a theoretical determination of elastic properties of rocks either endowed with porosity or containing partly and completely molten components the GSA method yields fairly good results, whereas the VRH method in these cases is either inapplicable or gives values very much different from experimental ones.

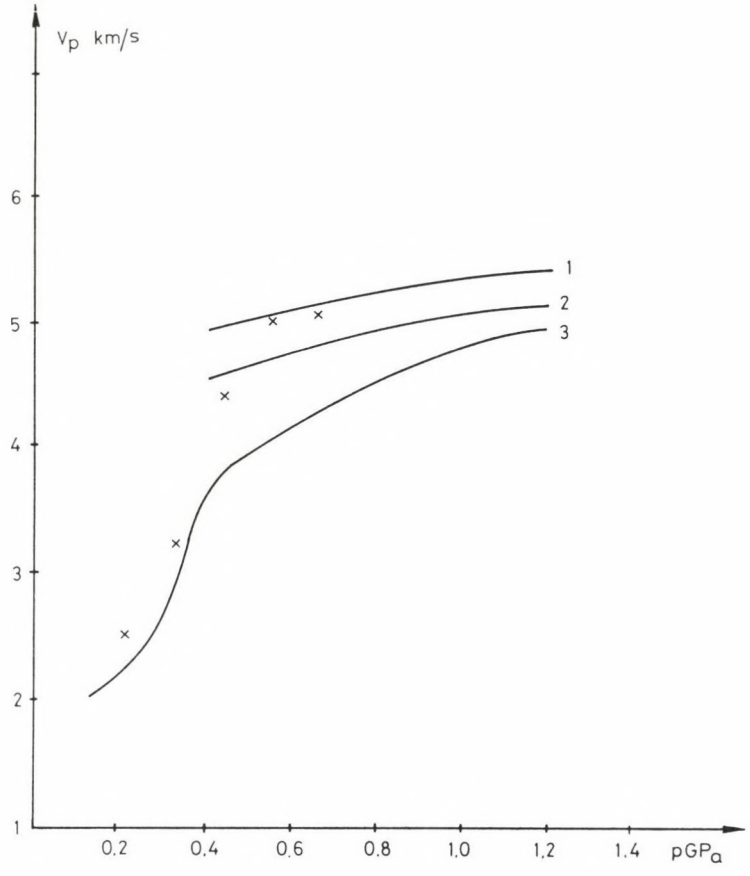


Fig. 5. Elastic wave velocities in a mixture of quartz (400 μm) and stearin at 95^{o}C. Compressional wave velocities: 1 - calculation after GSA method with a comparison body after Voight; 2 - calculation after VRH, 3 - calculation after GSA method with a comparison body after Reuss, x - experimental data

REFERENCES

Anderson O L, Shreiber E, Liebermann R, Soga N 1968: Rev. Geoph., 6, No. 4, 491-523.

Belikov B P, Aleksandrov K S, Ryzhova T V 1970: Elastic Properties of Rock-Forming minerals in Rocks. Nauka Publishers

Birch F I 1961: J. Geoph. Res., 66, No. 7, 2199-2224.

Christensen R M 1979: Mechanics of Composite Materials. New York, Chichester, Brisbane, Toronto, A Wiley Intersc. Publ. J. Wiley and Sons

Hill R 1952: Proc. Phys. Soc. A65, No. 389, 349-353.

Kalinin V A, Bayuk I O 1987: Geofiz. hz., 9, No. 2, 69-75.

Levykin A I, Farberov A I, Nasimov R M, Gliko A O 1981: Phys. of the Earth and Planet Let., 26, 163-170.

Manual on Physical Properties of Minerals and Rocks Under High Thermodynamic Parameters. Ed. by M P Volarovich. Nedra Publishers, 1978

Shermergor T D 1970: The Theory of Elasticity of Microinhomogeneous Media. Nauka Publishers

Simmons G 1964: Velocity of Shear Waves in Rocks to 10 kbars. Ibid, 69, No. 6, 1123-1130.

Walsh I B 1965: J. Geoph. Res., 70, No. 2, 381-389.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 89-98 (1991)

ACOUSTIC EMISSION IN SOLID-STATE TRANSFORMATIONS

V A Kalinin, I S Tomashevskaya, M I Greblov, A D Rasskazov

Institute of Physics of the Earth, USSR Academy of Sciences, 123810 Moscow, Gruzinskaya 10, USSR

[Manuscript received June 20, 1988]

The acoustic emission activity was investigated during a process of solid-state transformation (phase transition in KNO_3 at $\sim$ 128°C and reaction dehydration in gypsum). Experiments were carried out in a range from 20°C to 160°C. The sample of the investigated material was slowly heated up to 160°C, then it was cooled. The biggest acoustic emission peak was found during the cooling of the KNO_3 sample at $\sim$ 128°C.

Keywords: acoustic emissions; dehydration; gypsum; solid-state transformation

Proceeding from laboratory experiments under high pressures and temperatures, as well as from theoretical estimates, it appears to be universally acknowledged today that practically all rock-forming minerals under thermodynamic conditions of the Earth's crust and mantle are experiencing diverse solid-state transformations (polymorphous transitions, dehydration, metamorphism, etc.). More and more new data have been reported in recent years, indicating that during the process of solid-state transformations accompanied by microstructural variations of materials, solids are growing more deformable, less strong, propagation velocities of elastic waves in them decrease, thermal expansion coefficients are subject to abrupt alterations. Our experiments in determining the deformation-strength properties of quartz-containing rocks under thermodynamic conditions of the α - β transition in quartz (Tomashevskaya and Kalinin 1985, Kalinin and Tomashevskaya 1985) and Murrell's (Murrell and Ismail 1976) and Paterson's (Raleigh and Paterson 1965) works in experimental studies of strength properties during the process of

dehydration of various water-containing minerals have revealed that during these processes subject to anomalous changes are the physical properties not merely of individual minerals undergoing transformation but of the rock as a whole.

Studies of heat flows and the estimates of temperature distributions with depth obtained from them indicate that these distributions significantly differ for tectonically active and passive regions. Figure 1 shows by way of illustration the geotherms plotted on the basis of (Lyubimova et al. 1983) for placid regions (lower hatched zone) and for seismically active regions (upper hatched zone). The same Fig. 1 presents p and T values for which under laboratory conditions phase transformations were observed in $CaCO_3$ (Rapoport 1966) and in

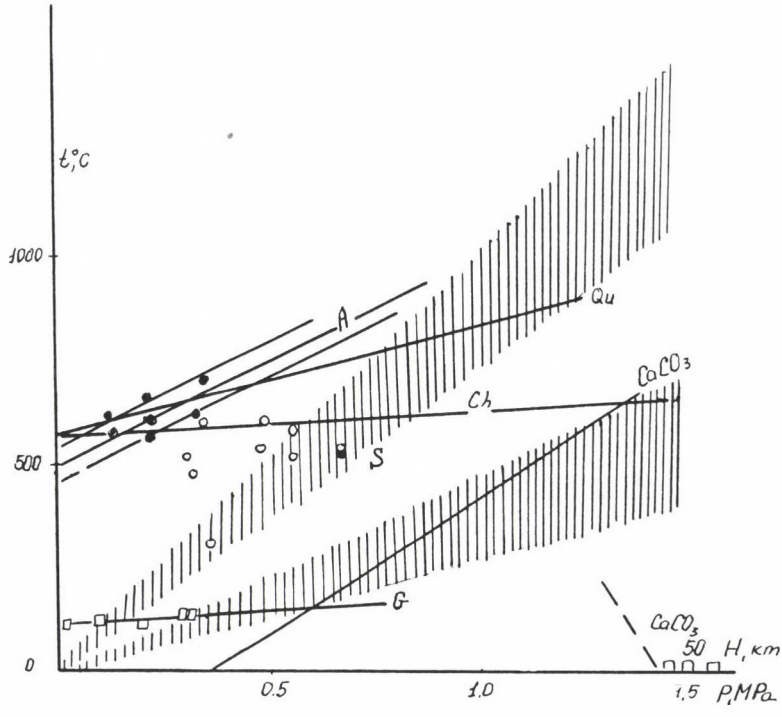


Fig. 1. p,T-conditions at down to 50 km depths and p,T-conditions for some solid-state reactions (cf. text)

quartz (QO) (Rapoport 1966), as well as a dehydration process was observed in gypsums (G), amphiboles (A), chlorites (Ch) and in a number of serpentinous rocks (S) (Murrell and Ismail 1976, Raleigh and Paterson 1965, Reference physical constants of the rocks 1969).

It is worth noting that in placid or stable regions thermodynamic conditions practically do not change with time, whereas in tectonically active zones these conditions are changing fairly rapidly. Hence, in seismically passive regions the boundaries of solid-state transformations do not change their position and, consequently, the specifics of the physical properties of minerals and rocks are not manifested, because anomalous properties arise solely when the process of lattice restructuring is in progress. In seismically active regions on account of variations in thermodynamic conditions the phase boundaries do not remain stationary, i.e. phase transformations are taking place along the boundaries and, as a result of this, there arise ductile layers, which significantly affect the progress of tectonic processes - a highly important fact as far as geophysics is concerned.

What has been said above indicates that the study of the kinetics of solid-state transformations and of the nature of these phenomena proves to be an important task for geophysics.

As is known, under mechanical loading or under temperature variation there arise acoustic pulses in polycrystals, which are due to the fissuring of the material, this fact being widely turned to account in technology for the purpose of detectoscopy. Intensity variation of acoustic emission (AE) under polymorphous transformations was reported only in some isolated works (Beattie 1973, Livshits et al. 1973, Kalitenko et al. 1980). We presume that AE during solid-state transformations can be employed for studying the kinetics of these transformations and, consequently, for an analysis of the causes behind anomalous alterations of physical properties of solids in the process of their transformation.

Investigations of AE during solid-state transformations of geophysical objects meet with great difficulties, because these

transformations usually occur under high pressures and temperatures and the available equipment restricts the possibility of recording weak acoustic signals. Accordingly, at least at the initial stage of work, investigations should preferably be conducted on model materials in which thermodynamic conditions of solid-state reactions permit us to employ the widespread standard piezoelectric transducers. Furthermore, this significantly improves conditions for recording weak signals because of the absence of noise from mechanical equipment setting up high pressure.

At the first stage of work, the results of which are reported herein, we worked out the procedure and software for recording AE signals, studied the characteristics of AE in three materials: sodium nitrate $NaNO_3$, potassium nitrate KNO_3 and gypsum - in the process of heating from room temperature to 200°C. These materials had been chosen for the following reasons: sodium nitrate in the above temperature interval does not experience any transformations and this makes it possible to bring to light the level and character of variation of acoustic noise in the system and the nature of AE under different heating rates; potassium nitrate with respect to its phase diagram and polymorphous structures happens to be a close analogue of $CaCO_3$ (Rapoport 1966), but polymorphous transformations in it take place under lower temperatures and pressures; on samples of gypsum we studied AE during the process of dehydration.

The registering apparatus represented a complex comprising an "Elektronika-60" computer and a modular, program-controlled system "KAMAK". Figure 2 shows a functional diagram of the measuring equipment. The software and hardware combined made it possible to register AE activity versus temperature and time, to count the total number of AE pulses, to select the threshold of detectability, to increase the number of measurements within the selected temperature interval and simultaneously to plot differential thermal analysis (DTA) data for different heating rates.

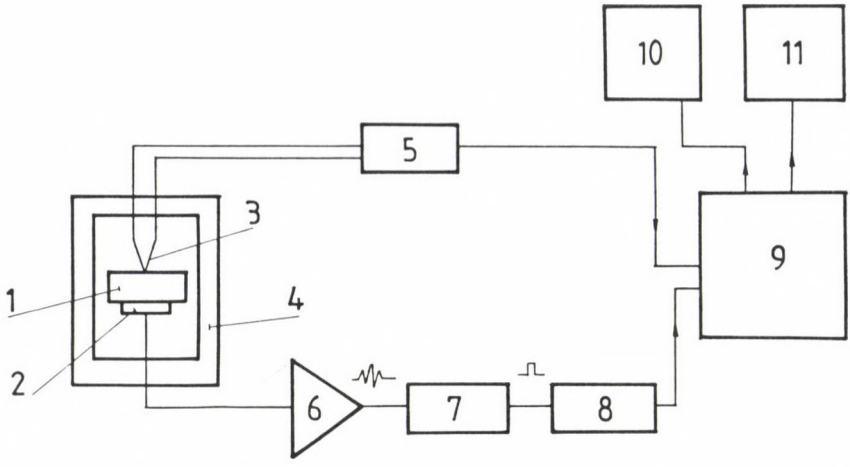


Fig. 2. Functional diagram of the measuring equipment. 1 - sample, 2 - transducer, 3 - thermocouple, 4 - furnace, 5 - A/D converter, 6 - amplifier, 7 - detector, 8 - counter, 9 - micro-computer, 10 - X-Y recorder, 11 - line printer

Samples along with piezotransducers and thermocouples were confined inside a steel vessel which was placed in a cylindrical furnace. Both heating and cooling were conducted at rates within 0.5^{o}C to 3^{o}C per minute. Samples of sodium nitrate and potassium nitrate were pressed from powder and those of gypsum were either cut out from monolith or pressed from crushed gypsum in the form of tablets of 20 mm diameter and 6 mm thick.

Special experiments were made to identify a difference of temperature at the edge and in the centre of the sample. Under the aforementioned heating rates the difference of temperature did not exceed 1-2^{o}C outside the range of temperature of solid-state transformation. Within the range of transformation temperature the difference in temperature could increase up to 5^{o}C.

Studies of AE have yielded the following results:

1. Sodium nitrate AE was characterized by weak background intensity without any bursts whatsoever in the process of both heating and cooling.

2. In potassium nitrate we observed AE bursts at

temperatures of polymorphous transformation, the AE intensity in heating being some hundred times less than that in cooling (Fig. 3). In a repeated heating the AE intensity during solid-state transformation increased nearly by an order, wheres in a repeated cooling in transition its intensity was nearly twice as low as in the first cooling. In subsequent heatings the AE intensity was observed slightly to increase, whilst in cooling it practically did not change.

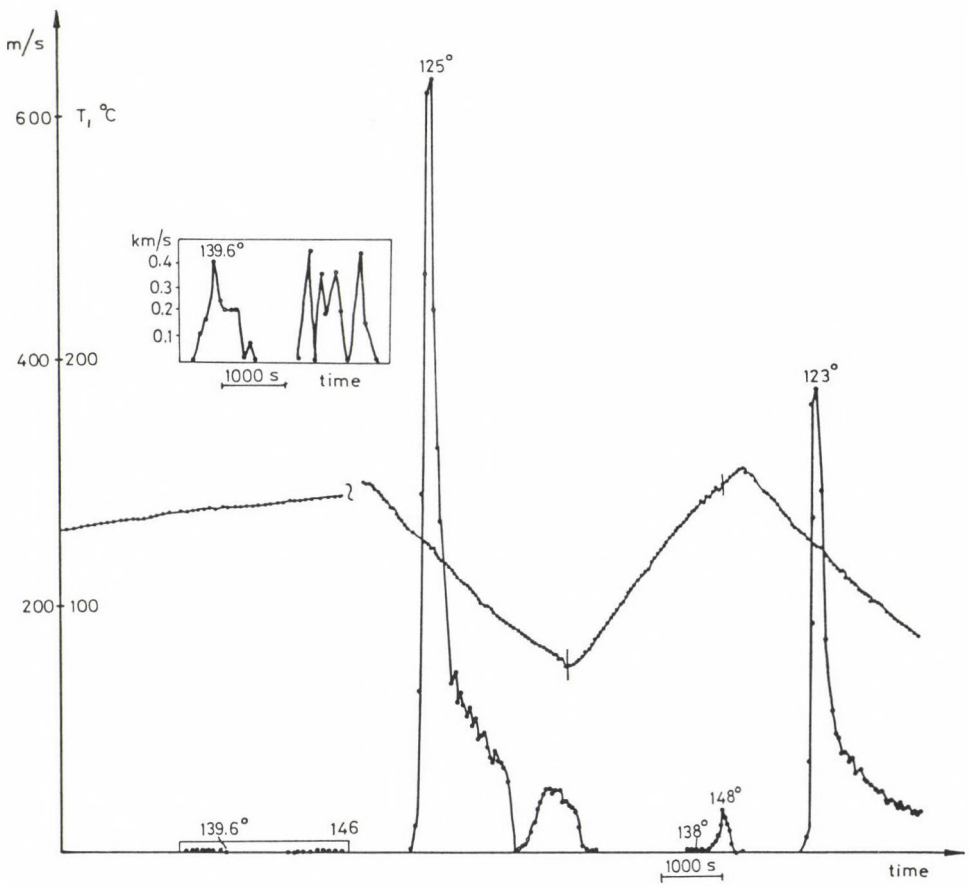


Fig. 3. Typical acoustic emission intensity versus time in the process of heating and cooling of a potassium nitrate sample - temperature variation, o - AE intensity

3. In heating, the gypsum samples over the entire temperature range are producing AE of weak intensity (Fig. 4). During a dehydration process the AE intensity rises but insignificantly.

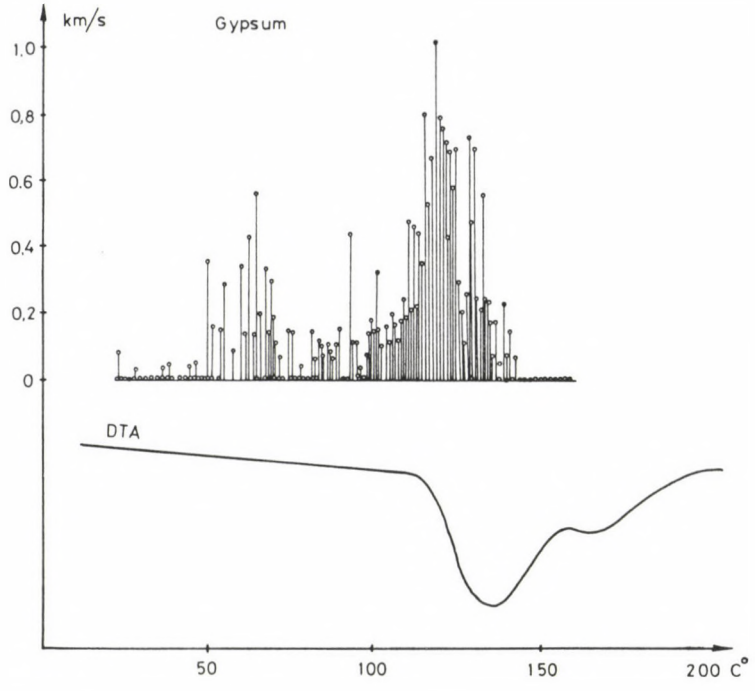


Fig. 4. AE intensity versus temperature in heating gypsum samples

In studying AE intensity records we observed certain specifics: in potassium nitrate under a temperature decrease at 132-135°C temperatures there arose acoustic pulses with a rapidly increasing intensity; an intensity maximum in cooling was observed at 123-125°C with a subsequent abrupt decline. In further cooling the radiation intensity of acoustic pulses continued to be sufficiently high down to room temperature, whereas DTA plotted in temperature decrease indicates a completion of the process at 112-114°C. It appears that in cooling, when transformation is proceeding with a volume decrease, an irregular stress field arises in the specimen and

its subsequent relaxation is accompanied by AE (Fig. 5).

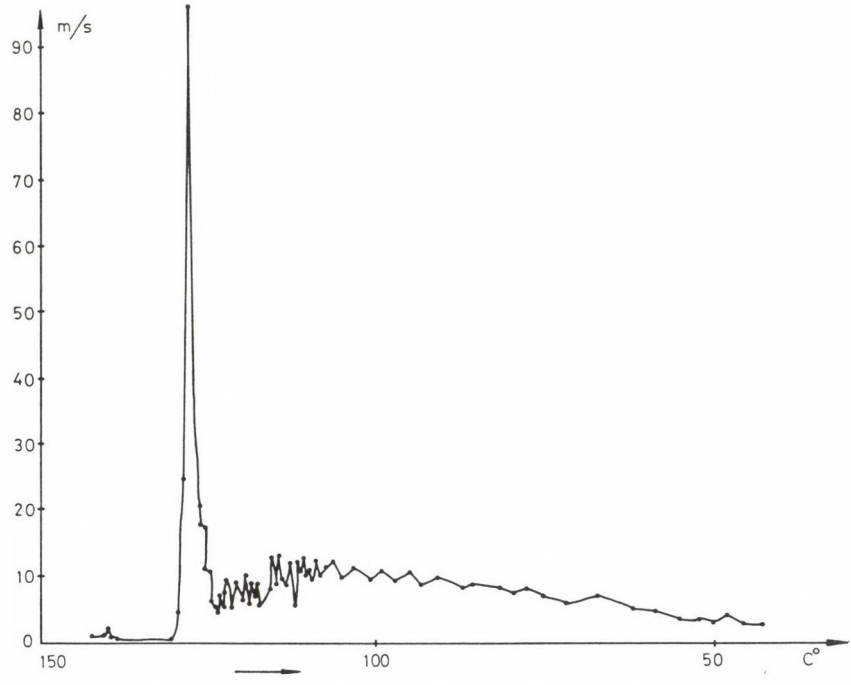


Fig. 5. An example of AE attenuation following a polymorphous transition in a potassium nitrate sample in cooling

In view of the fact that in seismically active areas there is a possibility of the temperature regime undergoing a change, we carried out special experiments in studying the AE intensity when in the process of polymorphous transformation in potassium nitrate the trend of temperature variation changes its sign to the opposite one. Figure 6 presents some results: the AE intensity behaves as an analogue of the Kaiser effect, i.e., if temperature in the specimen is reduced following the beginning of transition and the initial build-up of AE intensity, the AE intensity is observed sharply to decrease, and practically the AE intensity is equal to zero until temperature exceeds the one at which AE intensity started to fall off. On the right is shown the variation pattern of the AE intensity at the "tail" when a temperature decrease is followed by a local increase.

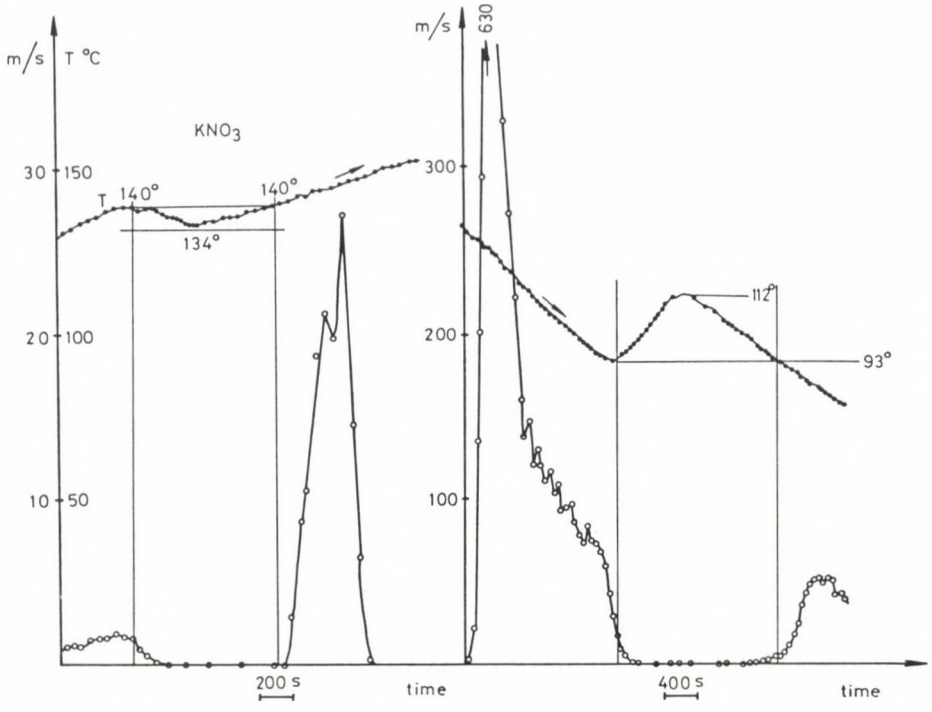


Fig. 6. Manifestation of the Kaiser effect's analog at the start of direct transition (a) and reverse transition (b)

And once again we observe a practical disappearance of AE until temperature decline attains a level at which AE discontinued.

Relief or disappearance of the Kaiser effect's analogue following transformation points to a kind of "annealing" (relief of inner stresses) in the process of transformation. And, as a result of this, we observe no drop of AE in a multiple passage through the transition (Fig. 3).

In conclusion we would like again to point out that the present work had for its purpose, above all, the elaboration of the experimental procedure and is, in effect, only the first stage in the study of the kinetics of solid-state transformations with the aid of AE. The results of the work indicate that in the setup concerned and with the experimental procedure described the AE intensity due to thermal stresses in the specimen and in the setup proper is sufficiently low, has

the character of a uniform noise without bursts, and by setting in the system a specific sensitivity level of registration it can be eliminated from analysis. This follows both from experiments with sodium nitrate experiencing no phase transitions within the investigated temperature range and in the heating of potassium nitrate up to the transition temperature. Under transformation in potassium nitrate, particularly in the process of cooling, there is observed a sharp rise in AE intensity. Furthermore, the AE intensity starts perceptibly to increase even before the DTA diagram registering the integral effect of transformation does show such a transformation. Consequently, AE is a more sensitive "instrument" (characteristic) recording the beginning of the process than is DTA.

REFERENCES

Beattie A G 1973: IEEE transactions of sonics and ultrasonics, SU 20, No. 1, p. 52.

Kalinin V A, Tomashevskaya I S 1985: In: Physical Properties of the Mineral System of the Earth's Interior. KAPG project 3, Czechoslovak Academy of Sciences, Geophysical Institute, Prague, 95-98.

Kalitenko V A, Perga V M, Salivonov I N 1980: Fizika tverdogo tela (Solid-State Physics), 22, No. 6, 1838-1839.

Livshits L D, Ryabinin Yu N, Zverev A S 1973: In: Propagation of Elastic and Elastic-Plastic Waves (in Russian). Nauka Publishers/Kazakh SSR, Alma-Ata, 236-244.

Lyubimova E A, Lyuboshits V M, Parfenyuk O I 1983: Numerical models of the Earth's thermal fields (in Russian). Nauka Publishers

Murrell S A F, Ismail I A H 1976: Tectonophysics, 31, 207-258.

Raleigh C B, Paterson M S 1965: J. Geophys. Res., 70, 3965-3985.

Rapoport E 1966: J. Phys. Chem. Solids, 27, 1349-1363.

Reference physical constants of the rocks 1969. Edited by Clark junior (in Russian). Moscow, MIK Publishers

Tomashevskaya I S, Kalinin V A 1985: In: Physical Properties of Rocks under High Pressures and Temperatures (in Russian). Yerevan, 117-118.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 99-109 (1991)

WAVE PHENOMENA NEAR THE PLASMAPAUSE

N G Kleimenova[1] and O M Raspopov[2]

[1]Institute of the Earth Physics, Moscow, 123810, USSR
[2]Polar Geophysical Institute, Apatity, 184200, USSR

[Manuscript received January 3, 1989]

The cold plasma density gradients in the vicinity of the plasmapause play an exceptional role in generation of different type waves in the magnetosphere near L=4. Hydromagnetic waves (ULF-waves) in the Pc1 and Pc3-4 ranges are very commonly seen in the area near the plasmapause. Ground and satellite data are analysed. The intensity maximum of the Pc4 peak (T=100 s) is always located at a lower latitude than the intensity maximum of the Pc3 peak (T=30 s). This was interpreted as produced by a sharp decrease of the phase velocity of the Alfven waves near the plasmapause. The plasmapause region is a very important area for the generation and ducting of the ELF-VLF whistler mode, too. The chorus emissions are typically observed outside the plasmapause, hiss emissions inside the plasmapause. There is also a very strong interaction between ELF-VLF whistler mode waves and hydromagnetic ones (ULF-waves). Pc5 and Pi2 geomagnetic pulsations frequently modulate the amplitude of VLF-waves observed on the ground.

Keywords: ELF; Pc3; Pi2; plasmapause; pulsations; VLF

INTRODUCTION

Experimental results reported in the last years have yielded important evidence that cold plasma density gradients in the vicinity of the plasmapause play an important role in determining the localization of different type waves in the magnetosphere near L=4. Substantial theoretical work has also been concentrated on the importance of plasma density gradients in the generation of hydromagnetic waves (ULF-waves) and ELF-VLF emissions. The stepwise increase of the Alfven velocity in that region from about 400-800 km/s to 2000-3000 km/s leads to a variation of the period of the fundamental toroidal oscillations with the L-parameter (Webb and Orr 1975). The

presence of the energetic trapped particles near cold plasma gradients produces the condition for cyclotron instability of the whistler mode and of high frequency hydromagnetic waves (Pc1) guided by the plasmapause to the ionosphere. So the area in the vicinity of the plasmapause will probably be rich in different type waves.

Waves with periods of about hundred seconds can be detected both at ground stations and on geostationary satellites. So, geomagnetic observations on the ground at subauroral latitudes give information about plasma instability and wave generation near the plasmapause. Depending on the geomagnetic activity, these stations are projected inside or outside the plasmasphere.

HYDROMAGNETIC WAVES (ULF WAVES)

Most studies of geomagnetic pulsations near the plasmapause are based on ground observations. Waves with periods in the Pc3-4 range (15-150 s) are very commonly seen in the area near the projection of the plasmapause. Using data of stations established around the $L \sim 4$ shell, the Bell Laboratories group (Fukushima and Lanzerotti 1974) found the dominant period of the pulsations to change sharply near the plasmapause. The intensity maximum of the Pc4 peak (T=100 s) near L=4 was always located at a lower latitude than the intensity maximum of the Pc3 peak (T=30 s). That result was obtained using megnetic field measurements made at stations ranging L=3.2-4.4. This was interpreted to be produced by a sharp decrease of the phase velocity of the Alfven waves with decreasing latitude. Such a decrease occurs at the plasmapause which was presumably located at a latitude between the intensity maximum of the Pc4 and Pc3 peaks.

Results obtained by a Soviet group (Kopytenko et al. 1975, Pudovkin et al. 1976) from a dense chain of 11 ground stations along the magnetic meridian over a wide range of latitudes indicated that Pc3 amplitudes have a minimum value in the vicinity of the expected plasmapause location (Fig. 1). For Pc3

pulsations one maximum exists inside the plasmasphere and another at higher latitudes, outside the plasmasphere. This suggests that the plasmapause is a hydromagnetic discontinuity for Pc3 waves because of the jump in the Alfven velocity in this region (Southwood 1975).

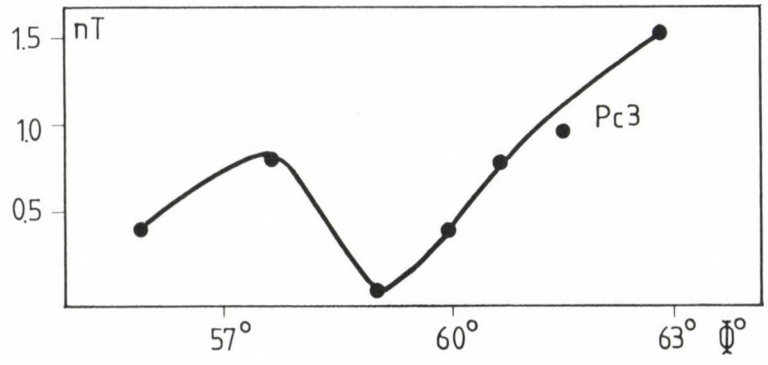


Fig. 1. Amplitude profile of Pc3 pulsations (from Pudovkin et al. 1976)

Orr and Mathew (1971) showed that the dominant periods of the hydromagnetic waves just outside of the plasmasphere are in the Pc3 range (T=30 s), while the dominant periods just inside the plasmasphere are in the Pc4 range (60-100 s). So, the plasmapause density gradients play a very important role in the localization of the origin of the hydromagnetic waves in the Pc3-4 range.

The wave polarization in the H-D plane as measured on the ground (Lanzerotti and Fukunishi 1975, Fukunishi and Lanzerotti 1974) changes sign across the rather narrow region near the plasmapause projection. A reversal in the sense of wave polarization occurs at latitudes where the ISI-2 satellite observed gradients in the electron density vs. L-shell.

The orientation of the elliptical polarizations on both sides of the plasmapause projection and the linear polarization at the plasmapause occur because the surface wave is coupled to the resonant field lines. According to Chen and Hasegawa (1974)

this happens in a region of magnetospheric density discontinuity.

Satellite data have also shown that Pc3-4 pulsations are common near the plasmapause. Electric field measurements by ISEE-1 (Moe et al. 1980) confirmed that Pc3 waves with periods 20-40 s are more often observed near L=4 during local day. The polarization for the electric field vector was mostly left handed in the morning sector and right handed in the afternoon sector.

Data from the Dynamic Explorer-1 satellite (Lin et al. 1986) at L ~ 3-6 and at magnetic latitudes $\pm 35^{o}$ of the magnetic equator showed that in general oscillations are complex. Transverse and compressional waves of different frequencies often occur at the same time (Lin et al. 1986). Their period range was 20 to 500 s, the common wave structure was complicated with short period transverse fluctuations of 20-40 s superimposed on long period waves with T ~ 100 s. They considered that the long period waves were most likely fundamental mode field line resonances while the short period waves were at least second harmonic mode resonances. A decrease of the eigenperiods of the resonant field lines with L was observed very clearly in some cases.

For short period pulsations, as IPDP it was found (Maltseva et al. 1981, Horita et al. 1979) that initially the source of IPDP waves is located near the plasmapause and it drifts later on inside the plasmasphere. The principal generation mechanism is cyclotron instability of the trapped particles.

Pc1 pulsations with periods of 0.2-5.0 s of a cyclotron nature are generated near the plasmapause (Feygin et al. 1979, Webster and Fraser 1985, Kikuchi and Taylor). Kikuchi and Taylor suggested that in the plasmapause region the shear Alfven waves are capable of producing short period pulsations Pc1: they are frequently observed on the ground near the same latitude and local time as OGO3 observed a structured plasmapause with a sharp density gradient. These results indicate that Pc1 excitation is associated with plasma

irregularities near the plasmapause and the situation is favorable in the region of post-storm recovery and in the region of plasma bulge of the afternoon - dusk sector. This mechanism, however takes place only after very strong magnetic storms. A more resonable source of Pc1 is the cyclotron resonance near the plasmapause (Feygin et al. 1979).

Long period surface waves could be excited at plasmapause gradients (Pc5 and Pi2 range).

Chen and Wolf (1972) showed that rapid changes with a time scale of hours in the crosstail electric field could produce a fine structure at the plasmapause which was observed by many satellites as by OGO-1, 3, 5; ISEE-1 at all local times, but especially in the dusk region. More recently Daly and Hughes (1985) using a simple computer model with a stagnation point on the dusk meridian have found that the hydromagnetic wave electric fields cause fluctuations in the cold plasma number density both at large (around 1Re) and small ($\sim$ 0.1 Re) distance scales. The effect primarily depends on the polarization of the wave electric field. The linearly polarized wave electric field located on the noon meridian caused small scale perturbations ($\sim$ 600 km) in the shape of the plasmapause boundary, while a circularly polarized wave electric field caused the boundary to have much larger deformations in addition of the line scale structure. The deformations of the plasmapause boundary by hydromagnetic waves are most noticeable in the dusk region.

ELF-VLF EMISSIONS

The plasmapause region is very rich in different ELF and VLF emissions. Kinter and Gurnett (1978) reported the existence of electrostatic ELF-waves found only at the plasmapause (Fig. 2) and demonstrated that the polarization and the spatial relationship to the plasmapause density gradient are consistent with a drift wave origin. The Hawkeye-1 spacecraft crossed the plasmapause at high altitude. A band of elecric field noise is detected often in the frequency range 1.7 to 175 Hz. No

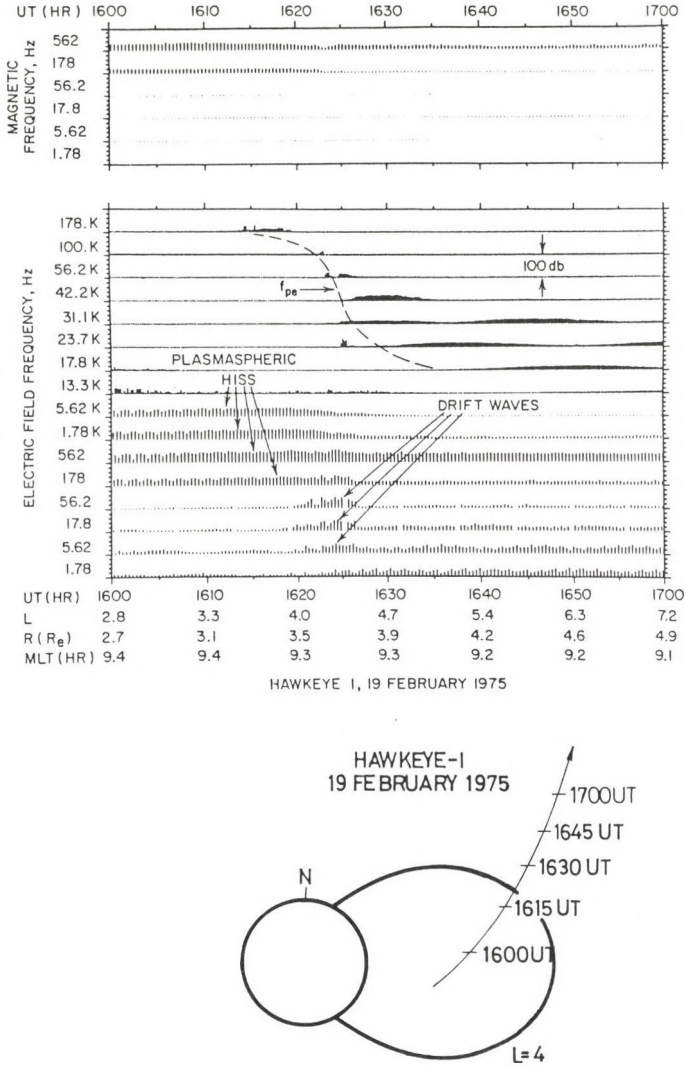


Fig. 2. Electric field spectrograms showing the output of the 16 electric field channels from 1.78 to 178 kHz and 6 magnetic field channels and trajectorie of Hawkeye 1 near the plasmapause (from Kinter and Gurnett 1978)

corresponding magnetic field noise was detected. That means that the ELF noise is electrostatic and the electric field is polarized perpendicularly to the plasma density gradient. The noise is only detected when the scale length of the plasmapause was 0.1 Re or less so a large density gradient is required to produce the ELF noise.

The plasmapause region is a very important area for generation and ducting ELF-VLF whistler waves. Many satellites, as OGO-3, 5, S-3-A, and others which crossed the plasmapause near equator show that outside the plasmasphere the typical phenomena are ELF-VLF chorus with maximum occurrence from 3 to 15 hours local magnetic time. The observation of OGO-5 showed (Daly and Hughes 1985) that outside the plasmasphere in the area where the Alfven velocity dropped to less than 3000 km/s, the ELF chorus was detected (Fig. 3), while inside the plasmasphere in the region with large cold plasma density there is an ELF hiss at the same frequencies. The intensities of ELF emissions between 0.1 and 1 kHz observed aboard OGO 5 are shown in the top part of Fig. 3. Bottom the profile of the Alfven velocity is shown by solid line.

The emissions can be understood in terms of an electron cyclotron resonance instability between energetic electrons and

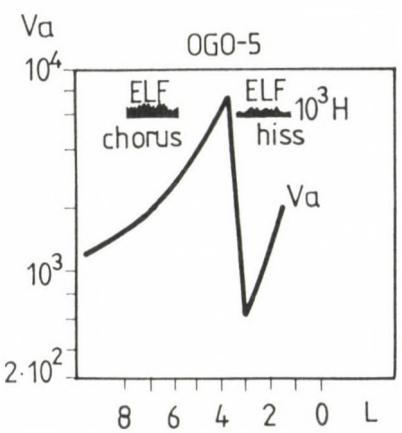


Fig. 3. ELF observations aboard OGO-5 together with the derived profile of Alfven velocity (from Barton et al. 1970)

whistler mode waves moving in opposite direction along the geomagnetic field lines. This mechanism is energetically favoured where the critical energy per particle is small, so it is just inside the plasmapause.

The same result was obtained on OGO-3 (Kinter and Gurnett 1978), the peak of the ELF-chorus intensity has been observed just outside of the sharp gradient of the cold plasma density. Low altitude satellites as Ariel 3, Ariel 4, Injun 3 detected ELF-VLF chorus outside the plasmapause and the same frequency hiss emissions inside the plasmasphere (Kaiser and Bullough 1975).

Many authors have shown that the plasmapause is an excellent guide for ELF-VLF whistler mode waves because of the sharp gradient of cold plasma density. The whistler waves could be ducted by both the inner and outward borders of the plasmapause (Fig. 4). With increasing gradients of the plasma density, the effect of ducting increases as has been found by

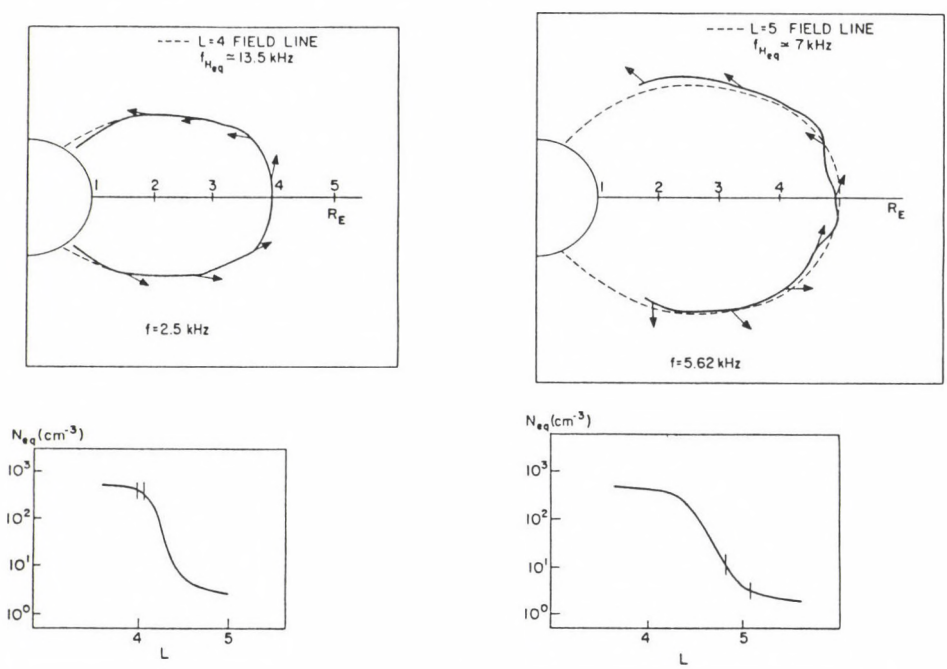


Fig. 4. The scheme of VLF wave guiding by inner and outward border of the plasmapause (from Inan and Bell 1977)

Inan and Bell (1977). If the plasmapause is located near L=4, the ducted waves have a highfrequency cuttoff near 6-7 kHz. This cuttoff explanes the high frequency limit of ground subauroral VLF emissions. The reflection of the wave inward by the plasmapause is alternatively countered by the curvature of the geomagnetic field which deflects the ray outwards.

Ground subauroral VLF observations at the conjugate stations Sogra-Kergelen (Kleimenova 1985) confirmed that the plasmapause ducted VLF waves from the magnetosphere to the ground. There was a very good conjugacy of VLF emission at the stations Sogra-Kergelen when the location of the plasmapause was near the latitude of these stations and it was very poor when the stations were projected inside the plasmasphere during periods of quiet magnetic activity.

The statistical result of the ground data confirmed that the maximum of chorus intensity was everytimes located outside the statistical plasmapause position, and shifted to lower latitudes with increasing magnetic activity (Kleimenova 1985, Raspopov and Kleimenova 1977).

There is also a very strong interaction between VLF waves and hydromagnetic ones in the vicinity of the plasmapause. The long period surface waves on the plasmapause modulate the growth ratio of the cyclotron instability by the strong magnetic wave component along the field line. Such a modulation can be observed only on the ground because the satellites cross this region in the time less than period of the modulating pulsations. The data from Sogra-Kergelen showed (Raspopov and Kleimenova 1977) that the modulation of the ELF-VLF intensity by geomagnetic pulsations Pc5 and Pi2 is typical for subauroral latitudes (Fig. 5). The modulation of the ELF-VLF hiss emissions by Pi2 geomagnetic pulsations is shown on Fig. 5. The very strong interaction between long period hydromagnetic waves and whistler mode electromagnetic waves often occurs in the vicinity of the plasmapause.

The modulation of the ELF-VLF emissions by Pc3 pulsations near the plasmapause latitude is probably due to the fact that the fundamental model of field line (Pc3) oscillations has a

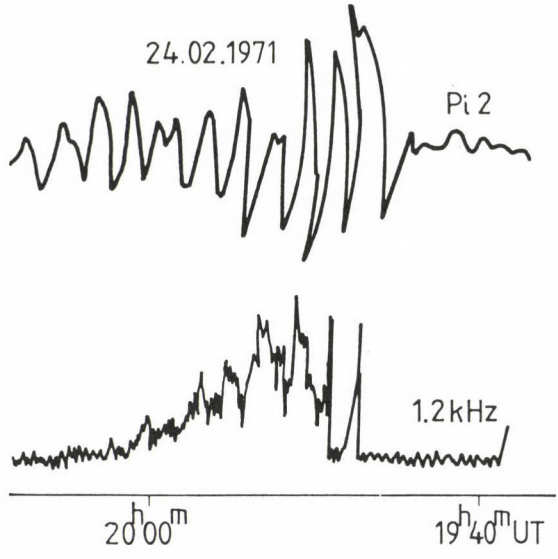


Fig. 5. An example of modulation of the VLF emissions by Pi2 geomagnetic pulsations at Sogra

node near the equator where the generation of electron-cyclotron whistler mode waves is most effective.

REFERENCES

Barton R K, Russell C T, Chapel C R 1970: J. Geophys. Res., 75, 5582-5586.

Chen L, Hasegawa A 1974: J. Geophys. Res., 79, 1024-1032.

Chen A J, Wolf R A 1972: Planet. Space Sci., 20, 483-509.

Daly R A, Hughes W J 1985: J. Geophys. Res., 90, A1, 537-542.

Feygin F Z, Kurchashov Yu P, Troitskaya V A, Fligel D S, Dobes K 1979: Plan. Space Sci., 27, 151-158.

Fukunishi H, Lanzerotti L J 1974: J. Geophys. Res., 79, 4632-4647.

Fukushima H, Lanzerotti L J 1974: J. Geoph. Res., 79, 142-158.

Horita R E, Batfield L N, Heacock R R, Kangas J 1979: J. Atm. Terr. Phys., 41, 293-307.

Inan U S, Bell T E 1977: J. Geophys. Res., 82, 2819-2827.

Kaiser T, Bullough K 1975: Ann. Geophys., 31, 137-141.

Kikuchi H, Taylor H A: Irregular structure of thermal ion plasma

Kinter P M, Gurnett D A 1978: J. Geophys. Res., 83, 39-44.

Kleimenova N G 1985: Geomagnetism and Aeronomy, 25, 266-271.

Kopytenko Yu A, Raspopov O M, Dmitrieva L A 1975: Planet. Space Sci., 23, 1195-1203.

Lanzerotti L J, Fukunishi H 1975: J. Geophys. Res., 88, 4627-4634.

Lin N, Chahill L I, Engebretson M J, Sugiura M, Arnoldy R L 1986: Plan. Space Sci., 34, 155-181.

McPherron D A, Koons H C 1970: J. Geophys. Res., 75, 5559-5564.

Maltseva N F, Troitskaya V A, Gerasimovitch E A, Baransky L N, Asheim S, Holtet J, Aasen K, Egeland A, Kangas J 1981: J. Atm. Terr. Phys., 43, 1175-1188.

Moe T E, Maynard N C, Heppner J P 1980: J. Geophys. Res., 85, AS 2099-2107.

Orr D, Matthew J A D 1971: Planet. Space Sci., 19, 897-905.

Pudovkin M I, Raspopov O M, Kleimenova N G 1976: The Earth electromagnetic field disturbances. Part 2. Short period geomagnetic pulsations. D. Leningrad University, Leningrad

Raspopov O M, Kleimenova N G 1977: The Earth electromagnetic field disturbances. Part 3. VLF emissions. D. Leningrad University, Leningrad

Southwood D J 1975: Ann. Geophys., 31, 101-104.

Webb D, Orr D 1975: Planet. Space Sci., 23, 1551-1561.

Webster D J, Fraser B J 1985: Planet. Space Sci., 33, 777-793.

Acta Geod. Geoph. Mont. Hung., Vol. (1-4), pp. 111-122 (1991)

RESULTS OF 120 KM LONG A DEEP MAGNETOTELLURIC EAST-WEST PROFILE IN THE VILLARRICA ACTIVE VOLCANO ZONE, SOUTH OF CHILE

H G Fournier[1], M Munoz[2], M J Mamani[3], J M Febrer[4], E Borzotta[3], A N Maidana[3]

[1]Institut de Physique du Globe de Paris, Universite Pierre et Marie Curie, 75230 Paris, Place Jussieu 4, France
[2]Departamento de Geologia y Geofisica, Universidad de Chile, Casilla 2777, Santiago, Chile
[3]Centro Regional de Investigaciones Cientificas y Tecnologicas, Casilla de Correo 131, Mendoza 5500, Argentina
[4]Centro Espacial de San Miguel, Av. Mitre, 3100, San Miguel, 1663, Argentina

[Manuscript received July 27, 1989]

Seven deep magnetotelluric (MT) soundings were carried out in the Villarrica active volcano zone in Chile during January 1986 and December 1987. Their interpretation shows an intercalated conductive layer (ICL) at an average depth of between 23 and 40 km, and a sharp decrease of the resistivity at a depth of about 530 km.

Keywords: Chile; intermediate conductive layer; magnetotellurics; ultimate conductive layer; Villarrica volcano

INTRODUCTION

Two deep MT sounding campaigns were done in January 1986 and December 1987 in the Villarrica volcano region, South of Chile (Fig. 1). The point of this study is the alignment of the three volcanoes Lanin, Quetrupillan and Villarrica, considered as the result of a movement from E-S-E to W-N-W of one eruptive center during the end of Tertiary and Quaternary eras. The regional tectonic is characterised by the Liquine-Ofqui fault (FL on Fig. 1), known over a distance of 1000 km, generally NS.

INSTRUMENTATION IN THE FIELD

Two „barres-magnetic-variometers" given by the Geophysical Institute of Paris to the MT Mendoza laboratory and a fluxgate

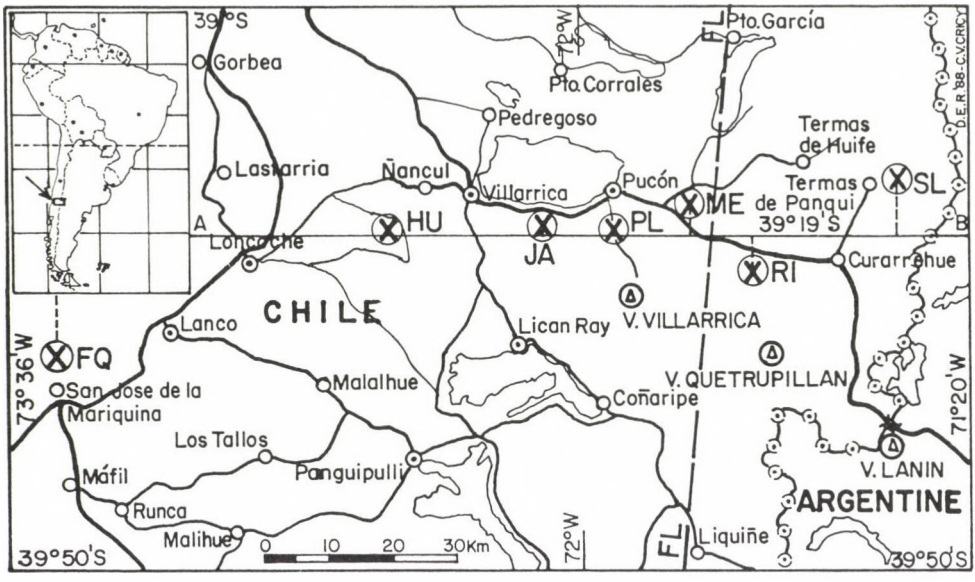


Fig. 1. Map showing the three volcanoes and the seven MT sounding sites: SL: San Luis, RI: Rinconada, ME: Metrenehue, PL: El Playon, JA: El Jaramillo, HU: Huiscapi, FQ: Fundo Quechuco

magnetometer were used. The telluric electrodes were sheets of lead situated in forest humus or in peat: previous records have shown that there was a negligible drift with this procedure. The recording time was six days for the ME, PL and FQ sites, four days for RI and JA and two days for the SL and HU sites.

DATA TREATMENT AND RESULTS

The digital data obtained on the field were analysed in the VAX-VMS/1180 computer of the CRICYT of Mendoza using a chain of four completely automatic programs based on the Fast Fourier Transform. The MT tensorial analysis, a part of this chain is from Rankin's laboratory, University of Edmonton, Canada. This chain gives a complete study of each site including the one-dimensional interpretation sections, using the one-dimensional automatic inversion program by Vozoff (see Fig. 7).

It remains to choose which component is parallel with the local tectonic axis supposed to be known by the geology. Figure 2

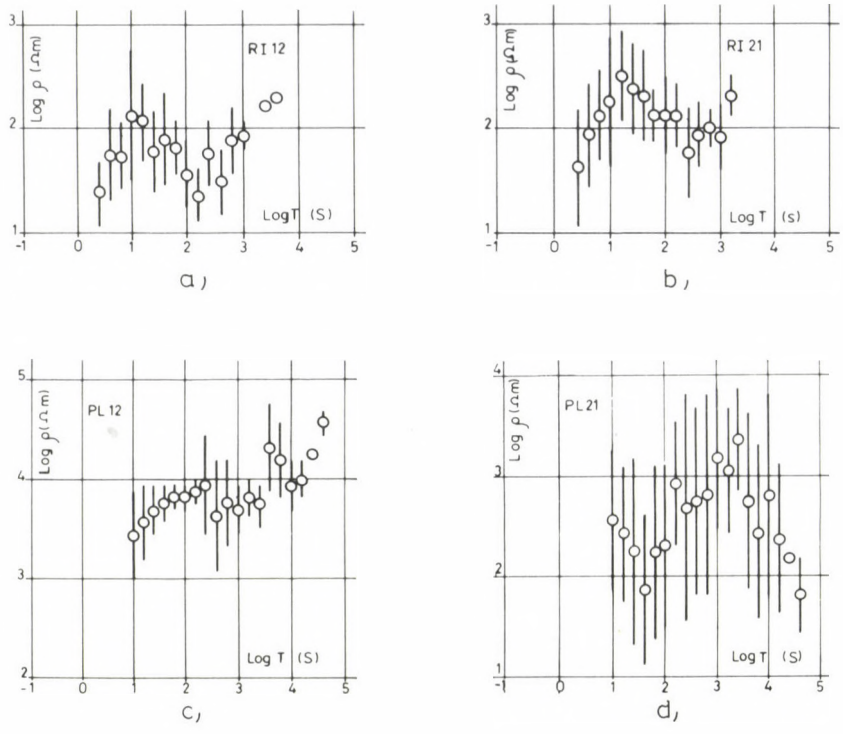


Fig. 2. Examples of Rho values of MT soundings at the first stage showing the error bars, thus before averaging: case of normal dispersion: a) RI12, b) RI21, case of high dispersion: c) PL12, d) PL21, IJ12 means tensorial direction located in the magnetic North-East quadrant, and in the same way, IJ21 in the magnetic North-West quadrant

gives two examples of Rho values with error-bars: sites RI and PL (before the mean average). Concerning the PL site showing very great error-bars, we have reviewed the analysis, using a 0.8 coherency coefficient limitation instead of 0.4 normally used for this MT study. The new result is almost the same for the length of the error-bars and for the Rho values. We therefore do not present the corresponding graph. This means

that the acting phenomenon is the direction of the geological structure axis versus that of the axis of the incoming electromagnetic waves and not the quality of the data recording in the field. Concerning the RI site, having more or less normal error-bars, we give in the three dimensional representation of Fig. 6 the distribution of the number of log Rho values in some intervals versus log T intervals for RI21 (Fournier et al. 1975, Kaufman and Keller 1981, p. 490).The Gaussian distribution seems to be acceptable. - See Fig. 2 (a) for the expression of the corresponding dispersion. Figure 3 shows the tensorial curves used for the interpretation of each site with the corresponding theoretical curve (account is taken of the indetermination principle concerning ICL). Table I gives the

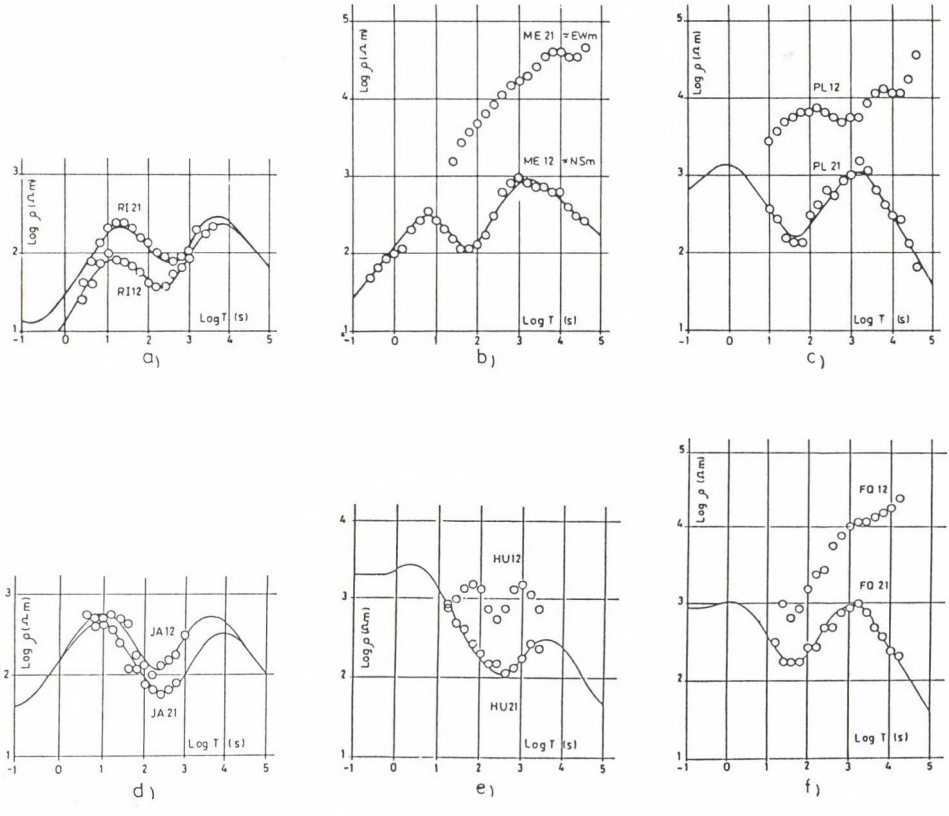


Fig. 3. Tensorial MT sounding curves, after averaging, with the theoretical curves in solid lines: a) RI, b) ME, c) PL, d) JA, e) HU, f) FQ

Table I. Theoretical sections obtained by the use of two successive one-dimensional MT inversion programs

Thickness in km	Resistivity in ohm.m	Thickness in km	Resistivity in ohm.m
RI12		RI21	
0.5	5.2	0.95	15.0
14.0	800.0	30.0	800.0
23.0	20.0	18.0	18.0
470.0	4000.0	470.0	4000.0
	17.0		17.0
ME12		PL21	
0.8	30.0	0.12	92.0
17.0	1000.0	16.0	1500.0
18.0	63.0	15.0	65.0
526.0	4000.0	520.0	4000.0
	93.0		1.0
JA12		JA21	
0.8	25.0	0.95	26.0
25.0	800.0	18.0	800.0
15.0	17.0	20.0	13.0
470.0	4000.0	470.0	4000.0
	17.0		17.0
HU21		FQ21	
38.0	2000.0	18.0	800.0
15.0	17.0	10.5	46.0
445.0	4500.0	470.0	5000.0
	10.0		3.0

corresponding theoretical sections obtained after a revised study of the calculated results given by the Vozoff program, using the non-automatic one-dimensional inversion program proposed by Fournier et al. (1963).

This means that we are using in this study the „cylindrical hypothesis" for the structure and that we are determining it using one-dimensional inversion programs. It is provisionally a first working hypothesis for this study.

INTERPRETATION

The interpretation of Table II shows two following main positions of conductive layers:

1) one ICL in the lower crust, between 23 and 40 km depth in average (Fig. 5). The Moho has been proposed at 37 km depth in the studied region by Diez and Introcaso (1986). In the Andes Chain, the geological structures are not in accord with Cagniard's theoretical conditions (Cagniard 1953), but the results of the RI and JA sites seem to be relatively less disturbed. The ME, PL and FQ results show a distortion with dispersion (Rokityansky 1982). The ICL-s of the RI, JA and HU sites have an integrated conductivity about 1000 to 1500 S which we are allowed to attribute to a melted fraction of the rocks, a phenomenon described by Feldman (1976). According to the geology, the ME and PL structures should be the results of a resistive granitic anticline of the basement, having an axis parallel to the general trend of the Andes Chain (Kaufman and Keller 1981). This could explain the distortion with dispersion presented by the MT curves of the two sites.

On the other hand, the skew of ME, 0.68, is relatively low, indicating a two dimensional structure: effectively the ME site is situated a little to the West of the Ofqui-Liquine great fault and consequently, the conductive effect of the NS-fault explains the interpretation direction of the sounding result, i.e. nearly NS. The site FQ also gives a low value for the skew: 0.54; the geology of this region is less known.

2) the top of one conductive layer situated at a depth of about 530 km, in average (Fig. 5). We suppose that it is the

Table II

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1986	Rinconada	RI	39°23'S	71°41'W	6	R12	58°Em	72°Eg	0.76	1.06	14-37	23	20	1150	3	
							R21	32°Wm	18°Wg	0.75		30-48	18	18	1000	3	
2	1986	Metreñehue	ME	39°16'S	71°53'W	6	NSm	2°Wm	12°Eg	0.82	0.68	18-36	18	63	285	3	560
							EWm	2°Sm	12°Wg	0.82							
3	1986	El Playón	PL	39°19'S	71°59'W	13	R12	38°Em	52°Eg	0.70	1.37						
							R21	52°Wm	38°Wg	0.85		16-31	15	65	230	3	550
4	1986	El Jaramillo	JA	39°18'S	72°08'W	4	R12	50°Em	64°Eg	0.70	0.85	33-53	20	30	670	3	
							R21	40°Wm	26°Wg	0.60		24-51	27	23	1170	3	
5	1987	San Luis	SL	39°16'S	71°36'W	7	R12	45°Em	59°Eg	0.79	1.14	13-18	5	0.25	$2 \cdot 10^4$	3	
							R21	45°Wm	31°Wg	0.81							
6	1987	Huiscapi	HU	39°19'S	72°23'W	4	R12	40°Em	54°Eg	0.76	1.34						
							R21	50°Wm	36°Wg	0.78		38-53	15	17	880	3	
7	1987	Fundo Quechuco	FQ	39°29'S	72°55'W	6	R12	65°Em	79°Eg	0.71	0.54						
							R21	25°Wm	11°Wg	0.95		18-28.5	10.5	46	230	3	500

1: Site number, 2: Year of field work, 3: Site name, 4: Site name abbreviated, 5: Latitude in degrees, 6: Longitude in degrees, 7: Number of analyses whose results are averaged, 8: Tensorial component interpreted, 9: Mean tensorial direction relative to magnetic North, 10: Mean tensorial direction relative to geographic North - the local declination is 14°Eg for the study epoch, 11: Mean coherency for the results of the considered component, 12: Mean skew for the site, 13: Depth of the top and bottom of the ICL in km, 14: Thickness of the ICL in km, 15: Resistivity of the ICL in ohm.m, 16: Integrated conductivity of the ICL in S, 17: Log T interval of moving average calculation in log Rho versus log T, 18: Depth of the top of the ultimate conductive layer in km (UCL)

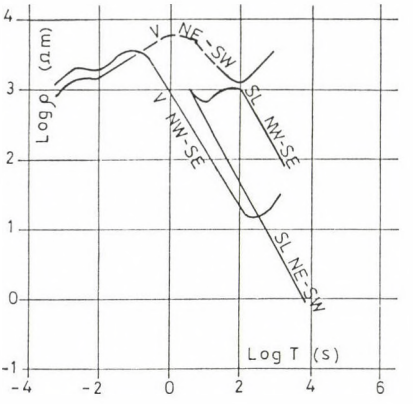


Fig. 4. MT curves of San Luis (SL, NE-SW, 2 10^4 S), Chile and of Vendée (V, NW-SE, 10^3 S), France

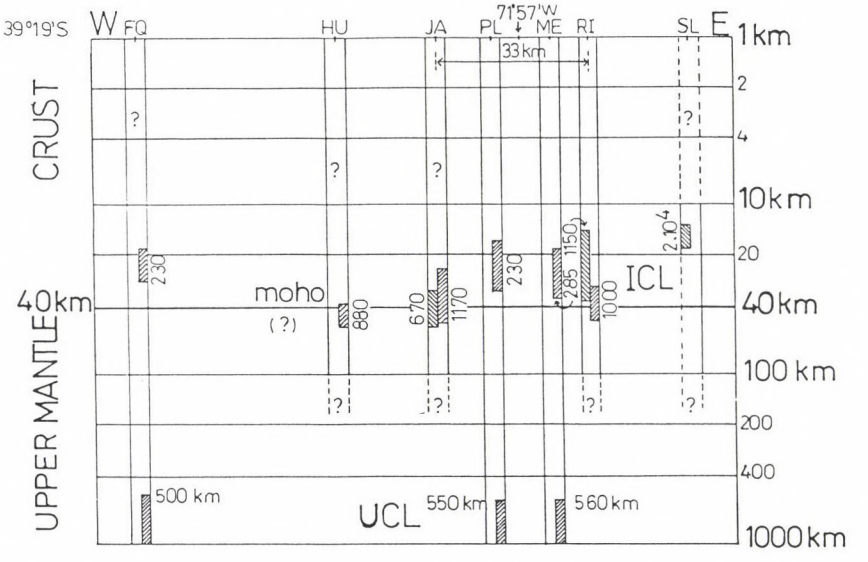


Fig. 5. Schematic cross section with the results of the one-dimensional interpretations. The vertical scale is logarithmic. The cross-section shows the position of the intercalated conductive layer (ICL) and ultimate conductive layer (UCL)

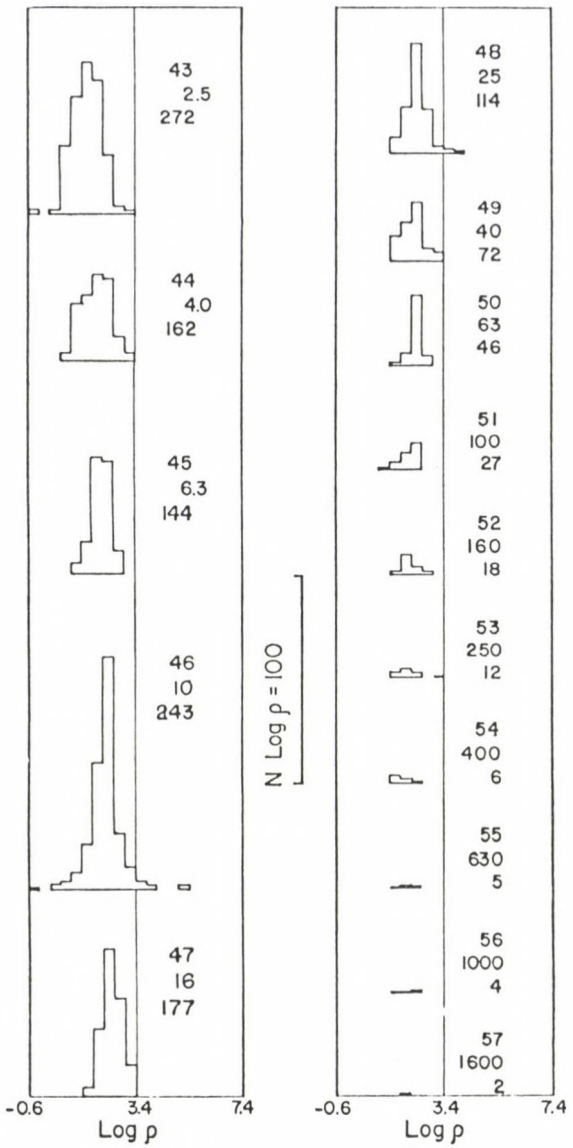


Fig. 6. Three dimensional representation giving the distribution of the number of log Rho values in intervals versus intervals of log T for RI21. The Gaussian distribution is acceptable. The order number, the central period T of the log T interval and the number of log Rho values for each interval of log Rho are indicated. This representation is given automatically by the chain of four programs (9000 FORTRAN lines without comments)

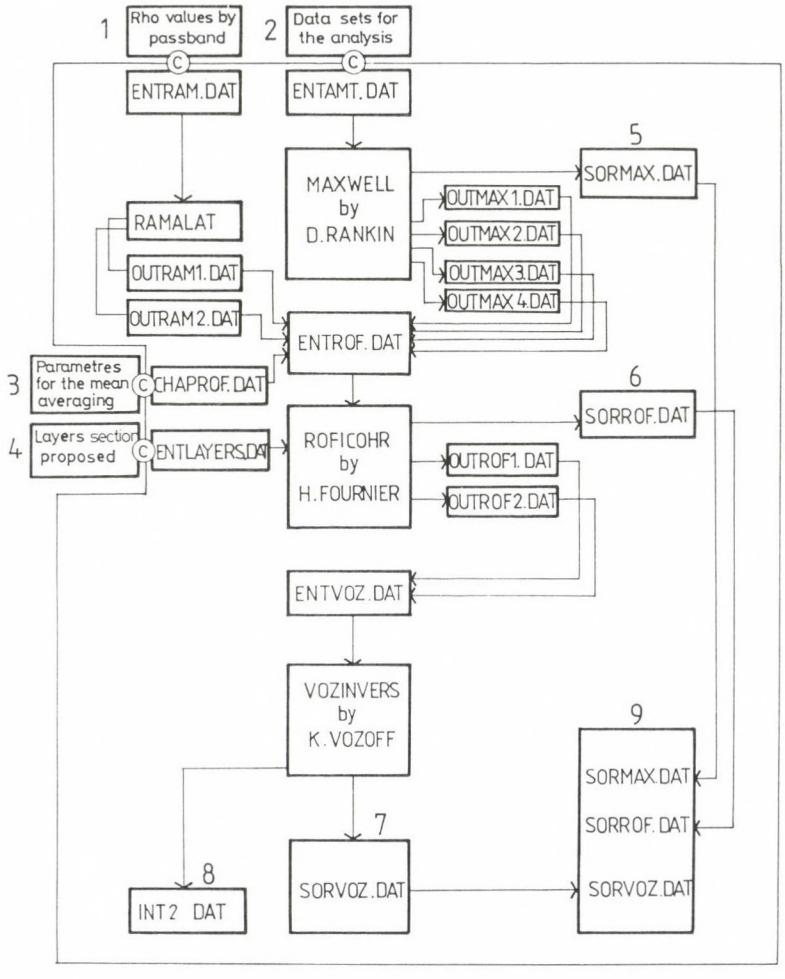


Fig. 7. Flow chart of the Maxwell tensorial analysis chain of programs containing 11000 FORTRAN lines (with comments).
A) Entries of the chain: C means entry in the chain by COPY operation. 1: Direct measurements of Rho or E/H values for the chosen direction of final Vozinvers interpretation. 2: Data sets ready for tensorial analysis having different sensitivity and interval of digitalisation. 3: Values of the parameters for the mean averaging operation, particularly the coherency limit - 0.40 for this study - and the value of the moving smoothing interval -3 for this study. 4: Proposed model of layers section for the Vozinvers one-dimensional inversion program - possible after a first run of the chain giving the tensorial curves. This model corresponds to the chosen direction of interpretation. Three runs of the program chain are needed to study the two tensorial directions of one MT site. The time for one mean run is 4 minutes with 2 minutes CPU time for our VAX-11/780 when there is a low

computer load. Six minutes CPU time give the complete study of one site. In the present study, Vozinvers gave the interpretation section using two iterations with 10 % error when the first interpretation is used as the proposed section.
B) Exits of the chain: 5: The Rho curves for each individual analysis - following the telluric measurement directions or tensorial directions - with the mean coherency, strike and skew. 6: The Rho log distribution, the mean Rho curves and the error-bar values and graphs for all the group of individual analyses, also the mean strike versus log T, etc. (75 pages of listing for a complete output). 7: The layers section given by the Vozoff one-dimensional inversion program, with comments. 8: The same without comments. 9: Exits 5, 6 and 7, together

transition layer, - C in the Bullen structure - transition from olivine and pyroxenes to spinels and stishovite (Feldman 1976, Lacam 1982). It is why we name it the ultimate conductive layer (UCL) (Fournier 1962, Fournier et al. 1963).

PARTICULAR CASE

The result from the SL site is very different from the others; it is interesting to note that it is nearly the same as the result of the MT site of Vendée in France (Fournier et al. 1975). In the two cases (Fig. 4) the MT soundings is situated over a fault containing at depth very conductive rocks. In Vendée the conductive zone was found at the same site by the deep electrical sounding made by the Compagnie Générale de Géophysique in the International Geophysical Year while at the SL site we do not know at the surface what may be the cause of a conductive fault at this depth. There are hot springs in this region - see Fig. 1: Termas de Huife and Termas de Panqui.

ACKNOWLEDGEMENTS

We are indebted to the „Centro Regional de Investigaciones Cientificas y Tecnologicas" of Mendoza for the use of its VAX-11/780 computer and to Mrs Elaine Jones from Chinley (Cheshire) for its English correction.

REFERENCES

Cagniard L 1953: Geophysics, 18, 605-635.

Diez A, Introcaso A 1986: Geoacta, Buenos Aires, 2, 179-201.

Feldman IS 1976: In: A Ádám (ed.) Geoelectric and Geothermal Studies. KAPG Geophysical Monograph, Akadémiai Kiadó, Budapest, 721-730.

Fournier H G 1962: Bul. d'Infor. de L'Assoc. des Ing. de la Fac. Polyt. de Mons, Belgique, 5-12.

Fournier H G, Ward S H, Morrison F H 1963: Magnetotelluric evidence for the low velocity layer. Space Sc. Lab. Univ. of Calif., Berkeley 4, Techn. Rep. on Nonr 222(89), Ser. n. 4, Issue n. 76, 7.

Fournier H G, Benderitter Y, Febrer J M 1975: C.R.Ac.Sc., Paris, 280, Série B, 141-143.

Fournier H G, Febrer J M 1976: Phys. Earth. Plan. Int., 12, 359-364.

Kaufman A, Keller G 1981: The magnetotelluric method. Elsevier, Amsterdam, 595.

Lacam A 1982: C.R.Ac.Sc., Paris, 295, Série II, 795-798.

Rokityansky I I 1982: Geoelectromagetic investigation of the Earth's crust and Mantle. Springer Verlag, Berlin, 381

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 123-133 (1991)

COMPUTER PROCEDURE FOR THE DETERMINATION OF TURBULENT PARAMETERS BASED ON IONOSPHERIC SPORADIC E

P Bencze, K Kovács, J Szendrői

Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H-9401 Sopron, POB 5, Hungary

[Manuscript received August 28, 1989]

It has been shown in earlier papers that the turbulent parameters in the lower thermosphere can be determined at mid-latitudes by means of the parameters of sporadic E (Es) layers and ionospheric, as well as atmospheric models. In this paper the steps of the computation and the structure of the computer program are outlined.

Keywords: computer program; ionospheric sporadic E; lower thermosphere; turbulence

INTRODUCTION

Considering the basic physical processes, the lower thermosphere is a highly important region of the upper atmosphere. The dynamical processes here determine, namely, the distribution of the constituents of the gas mixture not only in this height range, but by controlling the turbopause height (the boundary between homosphere and heterosphere) also at greater altitudes. From this point of view, the turbulent diffusion plays a significant role giving rise to the vertical transport of the constituents.

For the investigation of the dynamical processes ionospheric parameters related directly to dynamic phenomena can also be used (Thrane and Grandal 1981, Thrane et al. 1985, 1987). Thus, at mid-latitudes the parameters of sporadic E (Es) layers due to wind-shears offer a unique possibility to determine the turbulent parameters, since the windshear is

connected with turbulence by the gradient Richardson number (Bencze 1984, 1987). The Es parameters are regularly measured at many ionospheric stations all over the world. Where the wind speed is determined at a nearby station, too, the turbulent parameters can be computed. For the calculation of additional atmospheric and ionospheric parameters necessary in the procedure, atmospheric and ionospheric models can be used.

METHOD

A widely accepted measure of the turbulence in horizontally stratified flows is the gradient Richardson number

$$Ri = \frac{g \frac{\partial \Theta}{\partial Z}}{\Theta \left(\frac{\partial U}{\partial Z} \right)^2} \qquad (1)$$

where the numerator represents the force stabilizing the laminar flow by the vertical gradient of the potential temperature Θ, while the denominator stands for the force making the flow turbulent by the vertical shear of the horizontal wind U. The gradient Richardson number is preferred to the other measures, since it includes directly measurable quantities.

If the gradient Richardson number is known, the turbulent parameters (turbulent intensity, turbulent dissipation and turbulent diffusion coefficient) can be determined by means of empirical relations giving the turbulent intensity as a function of the gradient Richardson number and the horizontal wind velocity (Deacon 1959, Lumley and Panofsky 1964, Monin and Yaglom 1971, Zimmerman and Murphy 1977). In the study of transport processes the vertical turbulent diffusion coefficient K_Z is mostly used, which can be expressed applying these relations by the formula

$$K_z = 0.32 < w^2 > /N = 0.32[\mp 0.15\ Ri^{1/2} + 0.08]^2 U^2/N \qquad (2)$$

where $N = \frac{g}{\Theta}\ \frac{\partial \Theta}{\partial z}$ is the Brunt-Vaisala frequency (Weinstock 1978, 1981).

The determination of the vertical shear of the horizontal wind is based on the wind-shear theory of mid-latitude sporadic E (Whitehead 1961, MacLeod 1966). In the lower thermosphere ions linked to the neutral gas by frequent collisions are forced to move along geomagnetic field lines. This means that at mid-latitudes charged particles move up or down depending on the direction of the practically horizontal neutral wind. If the direction of the horizontal wind changes with height, in the northern hemisphere where an eastward wind turns to westward wind with increasing height, an accumulation of charged particles occurs; that is, due to the shear of the horizontal wind a stratification in the ion density profile is formed by this redistribution of ions, which is called a sporadic E (Es) layer. Motions showing such variation of wind direction with height occur in atmospheric gravity waves. The processes counteracting the effect of wind-shear and the survival of the stratification are recombination of the ions, as well as turbulence. The latter depends on the strength of the wind-shear.

These processes are included in the basic relations of the wind-shear theory of midlatitude sporadic E (e.g. Axford et al. 1966). For the determination of the resultant wind-shear, the vertical shears of both components of the horizontal wind are needed. Thus, two equations are necessary and they are obtained from the basic relations written for both ions and electrons and neglecting the effect of electric fields (Bencze 1989):

$$\frac{dU_y}{dz} = \frac{Q_i}{Q_i - Q_e}\left\{\left[\frac{1+Q_i^2}{Q_i X} - \frac{1+Q_e^2}{Q_i X}\right]\alpha_s n_m\left(\frac{\alpha_o n_o^2}{\alpha_s n_m^2} - 1\right) + a_1 U_y + \right.$$
$$\left. + b_1 U_x + c_1\frac{dT}{dz} + d_1\frac{d^2T}{dz^2} + e_1 g + f_1\frac{d^2n}{dz^2}\right\} \qquad (3)$$

$$\frac{dU_x}{dz} = \frac{Q_i}{Q_i - Q_e}\left\{\left[\frac{(1+Q_i^2)Q_e}{Q_i XZ} - \frac{1+Q_e^2}{XZ}\right] \alpha_s n_m \left(\frac{\alpha_o n_o^2}{\alpha_s n_m^2} - 1\right) + a_2 U_y + \right.$$

$$\left. + b_2 U_x + c_2 \frac{dT}{dz} + d_2 \frac{d^2T}{dz^2} + e_2 g + f_2 \frac{d^2n}{dz^2}\right\} \qquad (4)$$

are used, where

$$a_1 = \left(\frac{2Q_i}{1+Q_i^2} - \frac{1}{Q_i}\right)\frac{dQ_i}{dz} - \left(\frac{2Q_e^2}{(1+Q_e^2)Q_i} - \frac{1}{Q_i}\right)\frac{dQ_e}{dz}$$

$$b_1 = \left(\frac{Q_e}{(1+Q_e^2)Q_i}\frac{dQ_e}{dz} - \frac{1}{1+Q_i^2}\frac{dQ_i}{dz}\right)2Z$$

$$c_1 = \left[\left(\frac{Q_i^2-Z^2}{Q_i^2} - \frac{2(Q_i^2+Z^2)}{1+Q_i^2}\right)\frac{dQ_i}{dz} + \left(\frac{Q_e^2-Z^2}{Q_e^2} - \frac{2(Q_e^2+Z^2)}{1+Q_e^2}\right)\frac{dQ_e}{dz}\right]\frac{k}{Q_i eXB}$$

$$d_1 = \left[\frac{Q_i^2+Z^2}{Q_i} + \frac{Q_e^2+Z^2}{Q_e}\right]\frac{k}{Q_i eXB}$$

$$e_1 = \left[\left(\frac{Q_i^2-Z^2}{Q_i^2} - \frac{2(Q_i^2+Z^2)}{1+Q_i^2}\right) m_i \frac{dQ_i}{dz} + \left(\frac{Q_e^2-Z^2}{Q_e^2} - \frac{2(Q_e^2+Z^2)}{1+Q_e^2}\right) m_e \frac{dQ_e}{dz}\right]\frac{1}{Q_i eXB}$$

$$f_1 = \left[\frac{Q_i^2+Z^2}{Q_i} + \frac{Q_e^2+Z^2}{Q_e}\right]\frac{kT}{nQ_i eXB}$$

$$a_2 = \left[\left(\frac{2Q_eQ_i}{1+Q_i^2} - \frac{Q_e}{Q_i}\right)\frac{dQ_i}{dz} - \left(\frac{2Q_e^2}{1+Q_e^2} - 1\right)\frac{dQ_e}{dz}\right]\frac{1}{Z}$$

$$b_2 = \left(\frac{1}{1+Q_e^2}\frac{dQ_e}{dz} - \frac{1}{1+Q_i^2}\frac{dQ_i}{dz}\right)2Q_e$$

$$c_2 = \left[\left(\frac{Q_e(Q_i^2-Z^2)}{Q_i^3} - \frac{2Q_e(Q_i^2+Z^2)}{(1+Q_i^2)Q_i}\right)\frac{dQ_i}{dz} + \left(\frac{Q_e^2-Z^2}{Q_e^2} - \frac{2(Q_e^2+Z^2)}{1+Q_e^2}\right)\frac{dQ_e}{dz}\right]\frac{k}{eXZB}$$

$$d_2 = \left[\frac{(Q_i^2+Z^2)Q_e}{Q_i^2} + \frac{Q_e^2+Z^2}{Q_e}\right]\frac{k}{eXZB}$$

$$e_2 = \left[\left(\frac{Q_e(Q_i^2-Z^2)}{Q_i^3} - \frac{2Q_e(Q_i^2+Z^2)}{(1+Q_i^2)Q_i} \right) m_i \frac{dQ_i}{dz} + \left(\frac{Q_e^2-Z^2}{Q_e^2} - \frac{2(Q_e^2+Z^2)}{1+Q_e^2} \right) m_e \frac{dQ_e}{dz} \right] \frac{1}{eXZB}$$

$$f_2 = \left[\frac{(Q_i^2+Z^2)Q_e}{Q_i^2} + \frac{Q_e^2+Z^2}{Q_e} \right] \frac{kT}{neXZB} \quad .$$

In these expressions U_y and U_x are the E-W (westward) and S-N (northward) components of the horizontal wind, respectively. $Q_i = \nu_{in}/\omega_i$ and $Q_e = \nu_{en}/\omega_e$, where ν_{in} stands for the ion-neutral collision frequency and ω_i is the ion gyrofrequency. Similarly, ν_{en} is the electron-neutral collision frequency and ω_e represents the electron gyrofrequency. X and -Z are the ratios of the northward and downward components of the geomagnetic field to the total field intensity B, respectively. n is the ion (electron) density, n_0 being the background ion (electron) density (in the absence of wind-shear) and n_m the maximum ion (electron) density in the layer. α_s stands for the effective recombination coefficient inside the Es layer, α_0 being the effective recombination coefficient outside the stratification. e is the unit charge and k Boltzmann's constant. m_i and m_e are the ion mass and the electron mass, respectively.

It is to be noted that in these equations the first terms determine practically the magnitude of the wind-shear. The remaining terms represent rather a correction being orders of magnitude smaller than the first terms. The magnitude of the terms decreases going from left to right on the right side of the equations.

If the wind-shear components are computed from these equations and the vertical gradient of the temperature is taken from atmospheric models, the gradient Richardson number can be determined by means of Eq. (1). The turbulent diffusion coefficient can be obtained from Eq. (2).

COMPUTER PROCEDURE

For the computer procedure the steps of the computation are given by the above equations. These can be followed in the flow diagram presented in Fig. 1. First, the quantities necessary for the determination of the first term are computed. The total intensity (B) and the dip angle (YI = arc sin (B_z/B) = arc cos (B_x/B)) are determined by calculating the field components of the geomagnetic field on the basis of a Legendre model used in IRI 86 (FIELG). Then, the ion and the electron gyrofrequencies, as well as the ion-neutral and the electron-neutral collision frequencies are calculated using atmospheric models (CIRA 1972, MSIS-86). Thus, the ratio of the ion-neutral collision frequency to the ion gyrofrequency Q_i and that of the ion-electron collision frequency to the electron gyrofrequency Q_e can be computed.

The background electron density is obtained by means of ionospheric models (IRI 9) using only those parts of the model, which are necessary for these computations.

In Eqs (3) and (4) α_s and α_o are unknown. As a first approximation it is assumed that α_s is equal to α_o. The effective recombination coefficient outside the Es layer α_o is determined by means of a formula giving α_o as a function of the relative percentage density of the main ions O_2^+, NO^+ and the corresponding dissociative recombination rate coefficients (Torr et al. 1976, Walls and Dunn 1974).

The maximum electron density in the Es layer is calculated from the blanketing frequency fbEs of the Es layer read from ionograms. Then, with the above assumption and using only the first term in Eq. (3) a zero order approximation of the vertical shear of the E-W wind component is obtained.

The components of the horizontal wind velocity are also introduced in the main program. If the data of meteor radar measurements (D2) are used, then the components are directly given and the resultant speed is calculated. If the data of drift measurements carried out by the spaced receiver method (D1) are available, a harmonic analysis is carried out by a

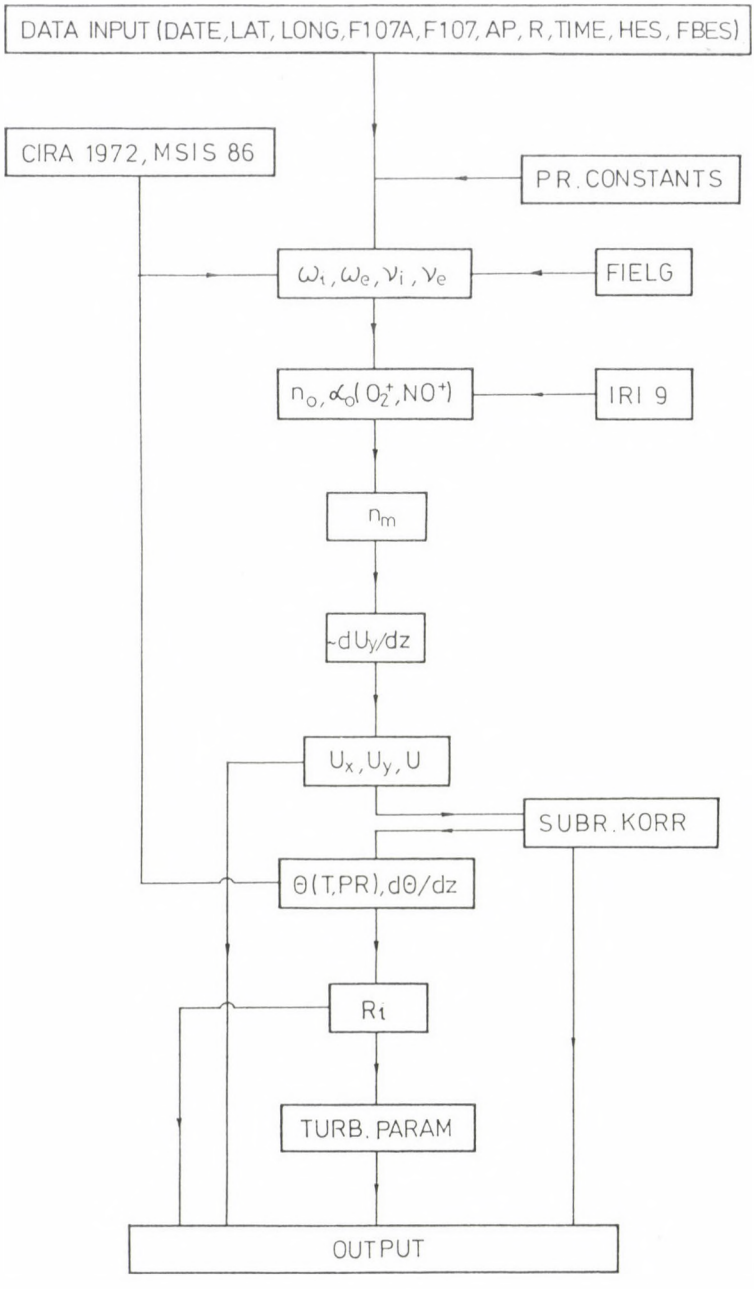


Fig. 1. Flow diagram of the computer procedure for the determination of turbulent parameters

separate program and the amplitude of the mean (prevailing) wind, as well as the amplitudes and phases of the diurnal and semi-diurnal components are determined. In this case the amplitude of the wind is computed for the given hour using the results of the harmonic analysis.

The other terms in Eqs (3) and (4) consist of products in which the parameters (U_x, U_y, dT/dz, d^2n/dz^2, d^2T/dz^2, g) are multiplied by coefficients. In the subroutine KORR first the elements of these coefficients as well as the first and second derivatives of the temperature are determined. The term with the second derivative of the electron density is neglected. Taking the zero order approximation of the vertical shear of the E-W wind component, the terms of Eq. (3) are calculated and a more accurate zero order approximation of the E-W wind-shear is obtained (still with the assumption $\alpha_s = \alpha_0$). In the following step the effective recombination coefficient inside the Es layer α_s is computed substituting this approximate value of the E-W wind-shear into Eq. (3) and expressing α_s from it. Thus, a first approximation for α_s is found. A more accurate value of α_s can be obtained by repeating the above procedure several times. Then the first term of Eq. (3) and all terms of Eq. (4) are determined and the vertical shear of both the E-W and the S-N wind components are calculated.

Returning to the main program, the potential temperature and its vertical gradient are computed, where the temperature and pressure are obtained again by means of atmospheric models. Thus, the gradient Richardson number and the Brunt-Vaisala frequency can be determined, too. Finally, the turbulent parameters, the turbulent intensity, the turbulent dissipation and the turbulent diffusion coefficient are calculated. The formulas used for these computations, however, are valid only for the inertial subrange of the turbulence spectrum (Zimmerman and Murphy 1977). Thus, it is necessary to determine the vertical scale of the motion producing the stratification and to establish whether the computed scale falls into the inertial subrange. (The limits of the inertial subrange are given at the larger scales by the outer scale L, at small scales by the inner scale, which is a multiple of the Kolmogorov microscale

(Lumley and Panofsky 1964, Hocking 1985)). The former is achieved by taking the ratio of the vertical wind amplitude w to the wind-shear.

Monthly averages of these turbulent parameters, but also that of the gradient Richardson number and the wind-shear components are determined. For the study of the height of the turbopause, the molecular diffusion coefficient can also be computed (U.S. Standard Atmosphere 1976). In Fig. 2 a sample output is shown where all parameters utilizable in further investigations are given.

JULIUSRUH JUNE 1982

MONTH [MON]: 6 NUMBER OF DAYS [LDAY]: 11
GEOGR. COORDINATES [LATI, LONGI]: 54.60 13.40

ALT	DVIK	DUZ	DUZE	RI	VT	W	EPS	TK
105.	.8740E-01	3.089	.2595	.6797E-04	38.21	9.057	.1531	357.1
105.	.6728E-01	2.383	.1997	.1128E-03	27.02	4.488	.7543E-01	178.0
110.	.2698E-01	.4277	.8008E-01	.3534E-02	26.25	3.481	.5773E-01	140.0
105.	.2792E-01	.9791	.8287E-01	.6703E-03	14.76	1.263	.2124E-01	50.08
105.	.1599	5.548	.4747	.2095E-04	36.62	8.436	.1423	333.4
105.	.8858E-01	3.065	.2629	.6779E-04	24.33	3.673	.6160E-01	146.0
150.	.2139E-01	.7369	.6351E-01	.1178E-02	34.71	6.751	.1132	268.4

Fig. 2. Sample output of the computer program

As regards the input data, in addition to the geographic latitude and longitude of the place, the altitude (given by the virtual height of the Es layer h'Es), the time (given by the time of its measurement), the Es parameter fbEs and the wind velocity are needed. Furthermore, the 12 month running mean of the sunspot number and if CIRA 1972 is used, the date and time in modified Julian days and the fraction of it, the right ascension and declination of the sun, the point's right ascension and declination, the 10.7 cm solar radio flux for a time of 1.71 days earlier, the 10.7 cm solar radio flux

averaged over four solar rotations (centered on the time in question) and the geomagnetic activity indices Kp for a time 0.279 days earlier, if MSIS-86 is selected, the 3 month average of the 10.7 cm solar radio flux, the 10.7 cm solar radio flux for the previous day and the geomagnetic activity index Ap are prepared.

ACKNOWLEDGEMENT

The authors should like to acknowledge the National Space Science Data Center through the World Data Center-A for Rockets and Satellites for sending the floppy disks containing IRI 9 and MSIS-86.

REFERENCES

Axford W I, Cunnold D M, Gleeson L J 1966: Planet. Space Sci., 909-919.

Bencze P 1984: Handbook for MAP, 10, 179-182.

Bencze P 1987: Acta Geod. Geoph. Mont. Hung., 22, 251-274.

Bencze P 1989: An improved method for the determination of turbulent parameters based on sporadic E (in press)

CIRA 1972, COSPAR International Reference Atmosphere. Akademie-Verlag, Berlin

Deacon E L 1959: Intern. J. Air Pollution, 2, 92.

Hocking W K 1985: Handbook for MAP, 16, 290-304.

IRI 9, International Reference Ionosphere. National Space Science Data Center, World Data Center-A, Greenbelt, U.S.A.

Lumley J L, Panofsky H A 1964: The Structure of Atmospheric Turbulence. Interscience Publishers, New York.

MacLeod M A 1966: J. Atmos. Sci., 23, 96.

Monin A S, Yaglom A M 1971: Statistical Fluid Mechanics: Mechanics of Turbulence. The MIT Press, Cambridge Mass.

MSIS-86 Thermosphere Model. National Space Science Data Center, World Data Center-A, Greenbelt, U.S.A.

Thrane E V and Grandal B 1981: J. Atmos. Terr. Phys., 43, 179-189.

Thrane E V, Andreassen O, Blix T, Grandal B, Brekke A, Philbrick C R, Schmidlin F J, Widdel H U, von Zahn U and Lübken F I 1985: J. Atmos. Terr. Phys., 47, 243-264.

Thrane E V, Blix T A, Hall C, Hansen T L, von Zahn U, Meyer W, Chechowsky P, Schmidt G, Widdel H-U and Neumann A 1987: J. Atmos. Terr. Phys., 49, 751-762.

Torr D G, Torr M R, Walker J C G, Nier A O, Brace L H, Brinton H C 1976: J. Geophys. Res., 81, 5578-5580.

U.S. Standard Atmosphere, 1976, NOAA NASA USAF, Washington D.C.

Walls F L and Dunn G H 1974: J. Geophys. Res., 79, 1911-1915.

Weinstock J 1978: J.Atmos. Sci., 35, 634-649.

Weinstock J 1981: J. Atmos. Sci., 38, 880-883.

Whitehead J D 1961: J. Atmos. Terr. Phys., 20, 49-58.

Zimmerman S P, Murphy E A 1977: In: Dynamical and Chemical Coupling Between the Neutral and Ionized Atmosphere. D. Reidel, Dordrecht-Holland, 35-47.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 135-151 (1991)

AVERAGE EFFICIENCY OF STATISTICAL PROCEDURES

F Steiner

Geophysical Department, University for Heavy Industry, H-3515 Miskolc, Egyetemváros, Hungary

[Manuscript received October 15, 1989]

The type-model of the probability densities may be different for different data sets. Certain ranges of the types occur more or less frequently in the practice, therefore the occurrence frequencies of the different types can be characterized by their density functions.

As error distribution types of the $f_a(x)$ supermodel are predominant in geosciences, the density function $f_t(p)$ is used in the present paper to define the average efficiency of different statistical algorithms ($p = 1/(a-1)$). The density function f_t is given numerically, too. The average efficiency is a global parameter which enables a simple comparison of statistical procedures. Moreover, the values of the average efficiency show most clearly the importance of the robustness for the practice (an overemphasizing of the insensitivity for outliers, i.e. of the resistance often veils in many practical papers the real sense of robustness, and thus the importance of this notion remains unclear, too).

Keywords: efficiency; error distribution types; statistical procedures

1.

Robustness has been treated recently in two monographs (Huber 1981, Hampel et al. 1986). Hajagos and Steiner (1991a) proposed a method to express numerically the robustness for practical purposes. This characterization can be used if a one-parameter supermodel is applied as basis which contains (at least as a limit distribution) the Gaussian distribution, too. In generally conceived geostatistics the supermodel $f_a(x)$ (see e.g. Steiner 1988) can be versatilly used and it fulfills the mentioned condition, too (the Gaussian distribution is obtained here at $a \to \infty$); else the Cauchy-distribution (at $a = 2$), the geostatistical distribution (at $a = 5$) and the Jeffreys-distribution (at $a = 9$) belong to this family, too.

If instead of a, $p = 1/(a-1)$ is introduced as type parameter, then the type distance of two distributions characterized by the type parameters p_1 and p_2 ($p_1 < p_2$) can be expressed by the product $(p_2-p_1) \cdot 0.064$ (for details see Hajagos and Steiner 1991a). In the following any constant multiple of the type distance may be used without influencing the numerical values, consequently the difference p_2-p_1 is considered as type distance.

The standard form of the density function of the distribution types in the supermodel $f_a(x)$ is expressed with the parameter p as follows:

$$f_a(x) = n(p) \quad (1+x^2)^{-\frac{1}{2p} - \frac{1}{2}} \tag{1a}$$

where the norming factor n(p) is to be calculated as

$$n(p) = \frac{\Gamma\left(\frac{1}{2p} + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2p}\right) \cdot \sqrt{\pi}} \; . \tag{1b}$$

(It would be perhaps disturbing to use instead of a, p in the index of $f_a(x)$ as the character of the supermodel does not evidently change if on the right hand side p is used instead of a.)

The minimum asymptotic variance of the estimated values of the location parameter is (Hajagos 1985 or Eq. (116) in Steiner 1988):

$$A^2_{min}(p) = p \, \frac{1+3p}{1+p} \; . \tag{2}$$

That means that if a different algorithm – being an optimum one from the point of view of p – is used for each p, then $A^2_{min}(p)/\sqrt{n}$ according to Eq. (2) would yield the variance of the

estimations of great sets of data, n.

The actual type of the error-distribution is, however, unknown for practical data sets. Some ranges of the type parameter occur with higher probability then the others – that is one hopes that the used method for estimation of the location parameter is not much less accurate than the optimum one. This is expressed by the asymptotic variance $A^2(p)$ of the actually used procedure so that the efficiency

$$e(p) = \frac{A^2_{min}(p)}{A^2(p)} \tag{3}$$

should be made as near to 1 (or to 100 percent) as only possible for the p-s occurring in the given situation.

As $A^2(p)$ can be very easily given for some algorithms in the case of the supermodel $f_a(x)$, e(p) itself can be given in these cases explicitely, too.

For all unimodal distributions which are symmetrical for the origin, the asymptotic scatter (i.e., the square root of the asymptotic variance) of the sample median is

$$A_{med} = \frac{1}{2f(0)} \tag{4}$$

thus by taking Eqs (1b), (2) and (3) into account the efficiency is for this estimation:

$$e_{med}(p) = \frac{4n^2(p) \cdot p \cdot (1+3p)}{1+p} \quad . \tag{5}$$

In the case of the Hodges-Lehmann estimation one has:

$$1/A_{HL} = \sqrt{12} \cdot \int_{-\infty}^{\infty} f^2(x)\, dx \qquad (6)$$

(Huber 1981) which results by substituting $f_a(x)$ in the expression

$$1/A_{HL}(p) = \sqrt{12}\, \frac{n^2(p)}{n(2/p+2)} \quad . \qquad (7)$$

Taking Eq. (2) into account one gets as the formula of the estimation of the efficiency:

$$e_{HL}(p) = \frac{12p \cdot n^4(p) \cdot (1+3p)}{n^2\ (2/p+2) \cdot (1+p)} \quad . \qquad (8)$$

For the efficiency of the arithmetic averaging in function of p one gets the following expression:

$$e_E(p) \begin{cases} \dfrac{(1-2p)\cdot(1+3p)}{1+p} & \text{if} \quad p < 0.5 \\ 0 & \text{if} \quad p \geqslant 0.5 \ . \end{cases} \qquad (9)$$

The variance is namely $1/(a-3)$ for the case $a > 3$, else it is infinite (e.g. Eq. (89) in Steiner 1988), and the „variance" equals statistically the asymptotic variance of the arithmetic means.

Practical and theoretical problems are treated in details by Steiner (1988): by choosing a coefficient k, different algorithms are obtained for the estimation of the location parameter. The general formula for the asymptotic variance are

given there in Eq. (124). Usually k is chosen as k = 1, 2 or 3, and the standard algorithm of the calculation of the most frequent value is that at k = 2. If k = 1 or k = 3 is used, the most frequent value is denoted by M_1 (or M_C) and M_3 (or M_J); in the case of the standard algorithm, the notation M (without index) is used.

The efficiency formulas of the M_1-, M- and M_3-computations are not repeated in this paper, they are only shown graphically (top curves in Fig. 1 taken from Steiner 1990a). In the type range presented there which covers types from the Cauchy type to the Gaussian one, all three curves show high efficiencies. The $e_E(p)$ curve according the Eq. (9), however, decreases rapidly from 100 percent to zero.

Figure 1 gives a basis for the introduction of a qualitative notion of the robustness for users (the mathematical definition is given by Huber 1981, Hampel et al. 1986 and Kerékfi 1978, but these definitions do not cover the present one in all cases). According to Fig. 1 there are significant differences between the different procedures in the rate of the decrease of the statistical efficiency when one gets farther and farther away from the distribution type characterized by p_{max}, corresponding to a maximum efficiency e_{max}. If the decrease of the efficiency is slow, the procedure is called a robust one.

The robustness has ment a boundary toward traditional statistical methods ever since the appearance of „robust statistics". It is a curious coincidence that the notion „robust" was introduced one decade before the appearance of the robust methods by Box for the algorithms being optimum ones for the Gaussian distribution to indicate that these methods can be used not only in the case of Gaussian distributions, but in the case of other distributions, too. In the possession of the notion „type distance" it is evident, however, that this statement is only true for distributions lying very near to the Gaussian one. That is why robustness got later a distinguishing attribute.

Hajagos and Steiner (1991a) proposed the following procedure to get a numerical measure of the robustness: to fit to

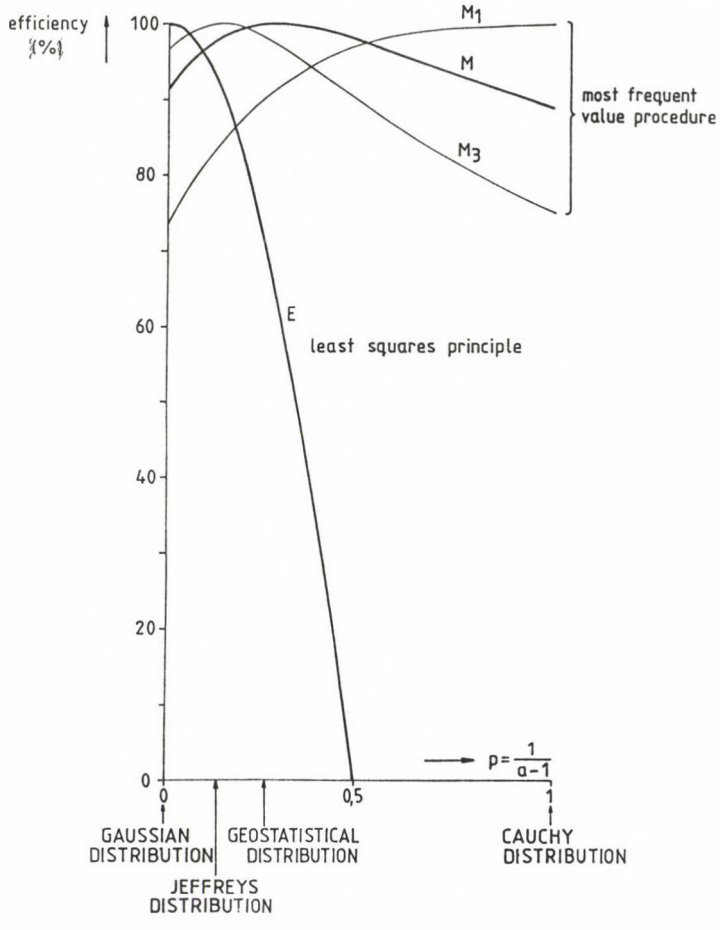


Fig. 1. Efficiency curves for different statistical algorithms

the location of the maximum a (second-order fitting) parabola and to determine from it by

$$e(p) = e_{max} - c \cdot (p_{max} - p)^2 \qquad (10)$$

the value c characterizing the more or less quick rate of decrease (it is the second order term in the Taylor-polynomial

at the maximum multiplied by (-1)). If this coefficient is denoted by C for the arithmetical averaging then

$$R = C/c \qquad (11)$$

is the measure of robustness. The traditional case is characterized by R = 1. This case is no more considered as robust, i.e. R has to be much greater than 1 to belong to a robust method in the present sense of the word.

Figure 2 (from Hajagos and Steiner 1991a) shows the curves $e_{M_3}(p)$ and $e_E(p)$ in a magnified form for the range of distribution types Gaussian-geostatistical. The fitting parabolas are not drawn as the value 3.7 belonging to M_3 can be easily verified. Figure 3 (also from Hajagos and Steiner 1991a)

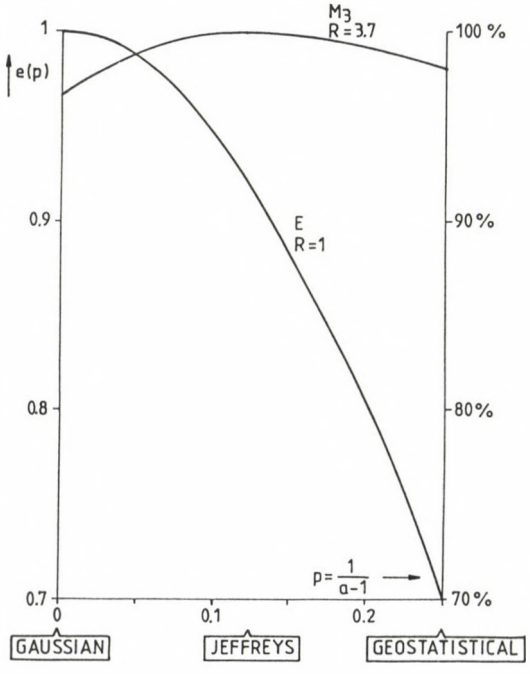


Fig. 2. A magnification of the maxima of the efficiency curves indicates the differences in the flatness of these curves, i.e. in the robustness

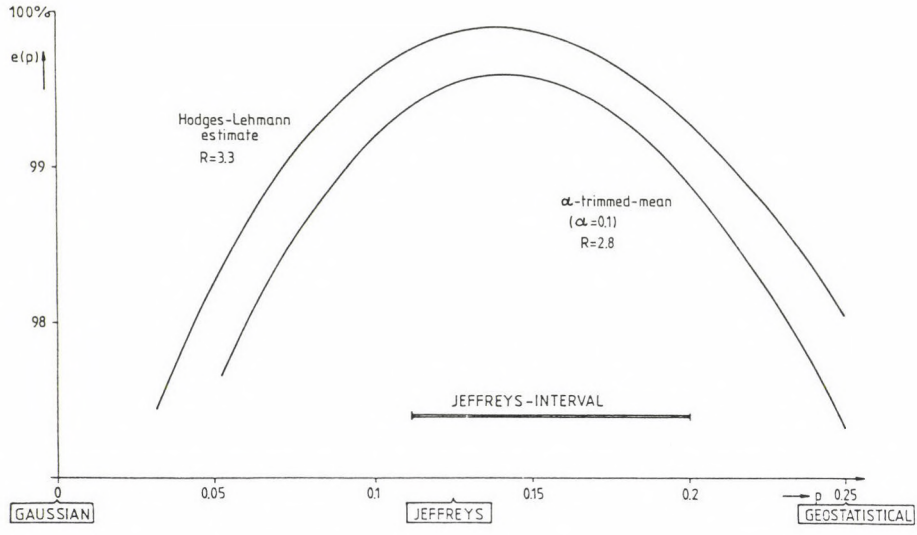


Fig. 3. The range of the maxima of the efficiency curves of the α-trimmed mean and of the Hodges-Lehmann-estimation

shows in the vicinity of the maximum in an even greater magnification for the Hodges-Lehmann-estimation, further for the α-trimmed mean (for α = 0.1) the efficiencies of the region around their maxima; the measure of the robustness is somewhat less, R = 3.3 for the Hodges-Lehmann-estimation, and even less, R = 2.8 for the α-trimmed mean.

2.

In the cases treated in the previous chapter, e_{max} has been exactly 100 percent or less by a few tenth percents. In such cases the measure of the robustness, R, can be considered in a certain sense directly as a measure for the qualification of the statistical method (in the vicinity of the distribution type p_{max}).

If e_{max} lies far away from 100 percent, the R-s computed from Eqs (10) and (11) are not applicable to qualify the method. Let us consider e.g. the efficiency of the median determination, as presented in Fig. 4 with abscissae differing significantly from the previous one. The R-values would be much greater than previously (in accordance with Hampel (1986) who qualified the median as a „most robust" method), but such a high degree of robustness is accompanied by a maximum efficiency of less than 84 percent. Thus the value R cannot be accepted as a unique qualifying measure in the case if e_{max} significantly differs from 100 percent. We have to continue the

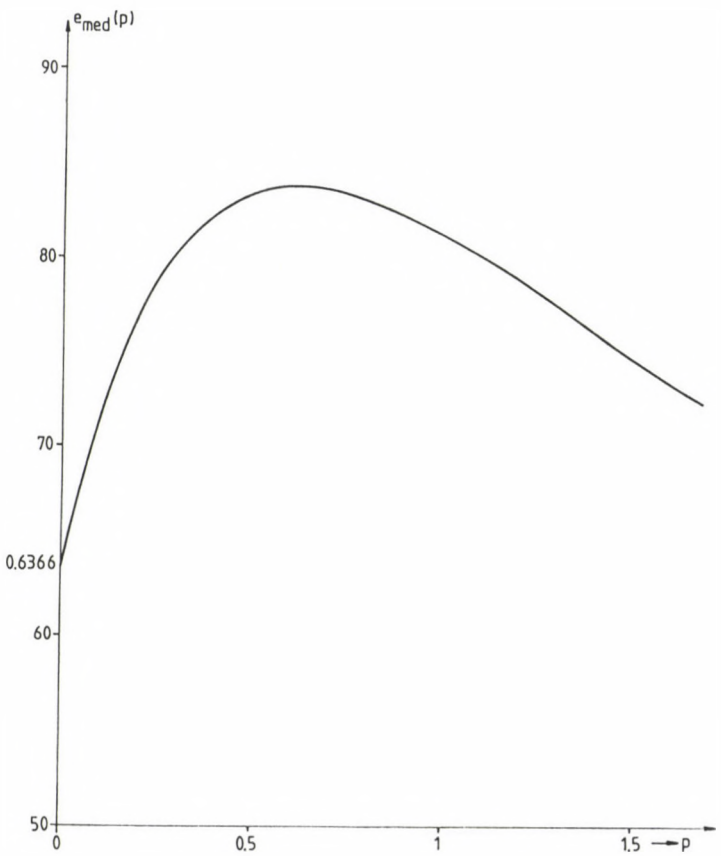


Fig. 4. Efficiency curve of the sample median

search for a global measure of robustness which enables a simple comparison of statistical methods.

3.

The efficiency is of crucial importance in the qualification of statistical methods, but a single value e(p) refers to a single type of distribution characterized by (p) (which rigorously has a mathematical probability of occurrence zero) and a range of types has only a finite probability of occurrence. Even if the same wide intervals are taken into account around different values of p, the probabilities of occurrence are different for them, thus the occurrence probability of the types should be also characterized by their density functions.

For the determination of a density function $f_t(p)$ the necessary quantity of type determinations may not be available even in geosciences (index t refers to the density function of types of distributions). Nevertheless, Steiner (1990b) has shown that the available data and the requirement of simplicity are fulfilled by the following density function

$$f_t(p) = 16pe^{-4p} \quad . \tag{12}$$

Let us give some facts which motivate the selection of Eq. (12) as density function as long as no more information is available about distribution types.

Both a derivation or Fig. 5 show that the maximum of $f_t(p)$ is at $p = 0.25$. Dutter (1987, p. 22) has presented the density function of a mother distribution considered as typical in geostatistics. This is a density function $f_a(p)$ with the parameter $a = 5$. Newcomb's (1886) classical astronomical density function is very near to a density function with the parameter $a = 5.3$ (Fig. 38 in Steiner 1988). A similar value of a characterizes the error distribution presented by Romanowski (1964) resulting from automatic angle measurements in a laboratory - these data show that such a type of distribution may have an eminent role outside of geosciences, too. A much smaller, but not negligible (five times less) density of probability is attributed to the

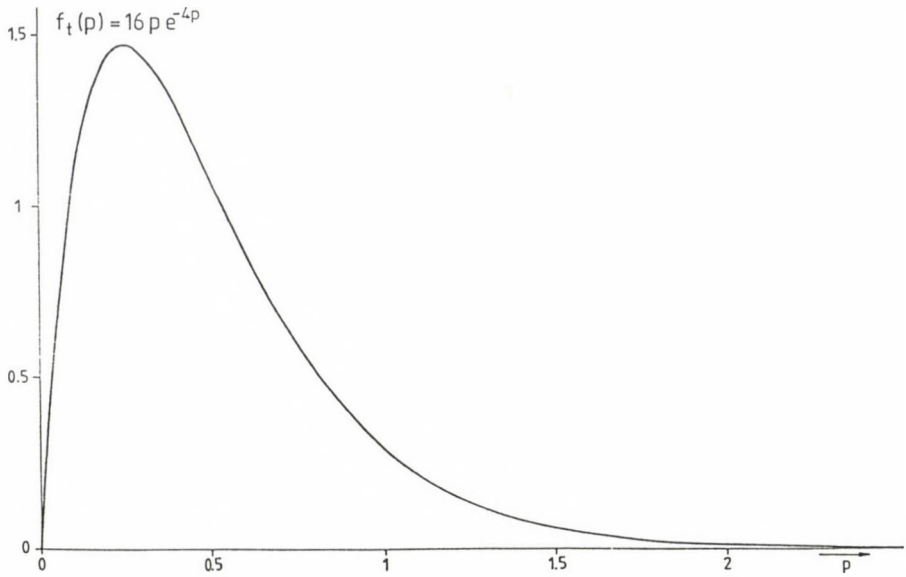


Fig. 5. Density function $f_t(p)$ of the occurrence of error distribution types

Cauchy-distribution by Eq. (12); the occurrence of this distribution is not negligible in geosciences as shown by Landy and Lantos (1982) by a practical example and by Steiner (1982) indirectly, from the analysis of a data system. The density function $f_t(p)$ allows in a rate of about 9 percent distributions with $p > 1$. (Such distributions may occur seldom in practice, but in the difficult case of the detection of „symmetrical and near outliers" the application of such a type could be actual, and therefore it would be a mistake to neglect such distributions; some meteorological data sets may have such a character, too.) The occurrence frequency of 9 percent of these types $p > 1$ is much less than the probability of the types $0 < p < 0.5$: the latter have an occurrence probability of about 60 percent.

The probability density of the Gaussian distribution is according to Eq. (12) zero. Steiner (1988) summarized in a table some statements about the occurrence of this type.

Jeffreys wrote that experiments carried out in most carefully controlled conditions result in distributions characterized by $6 < a < 10$ $(0.11 < p < 0.20)$. This interval is called „Jeffreys' interval" (within it, the type $a = 9$ is called Jeffreys' distribution). The $f_t(p)$ curve in Fig. 5 shows that the Jeffreys-interval is characterized by a great average $f_t(p)$ value being about 90 percent of the maximum. χ^2-tests carried out for the distribution types of this interval having high occurrence frequency show in a significant percentage of the cases that the „zero hypothesis according to which the distribution is a Gaussian one, can be accepted" (Hajagos 1988). This misleading result corresponding to the usual levels of the χ^2-test, however, cannot yet be accepted.

As geosciences do not yield Gaussian distributions (seeming exceptions can be produced by incorrect tests), some examples from classical statistical textbooks should be investigated. Our investigations in this respect led to similar results, thus we refer here only to Hajagos and Steiner (1991c) where two basic examples of Kendall and Stuart (1958) have been investigated and the result was in the first case a distribution with $a = 10$, in the second a distribution with about 50 ($a = 47$). Thus the classical examples in textbooks are perhaps nearer to the Gaussian distribution than Jeffreys' interval, but they are no Gaussian ones. The density function $f_a(x)$ for $a = 50$ is of course very near to the Gaussian distribution (the type distance is only 0.02) a comparison of the density functions lying near to each other is given in Fig. 6. Hajagos and Steiner (1991c) have shown that such near types of distributions can be distinguished only on the basis of data sets consisting of $n = 3 \cdot 0.6/0.02^2 = 8100$ data and this distribution being so near to the Gaussian one is characterized by the density function $f_t(p)$ with the same probability density of occurrence, i.e. 20 percent of the maximum, as the Cauchy-distribution. The distribution at $a = 10$ has a density of 76 percent of the maximum. We conclude therefore that the occurrence of distributions near to the Gaussian one is described by the density function $f_t(p)$ (Eq. 12) that no objections can be made based on our present knowledge against it (if only that the

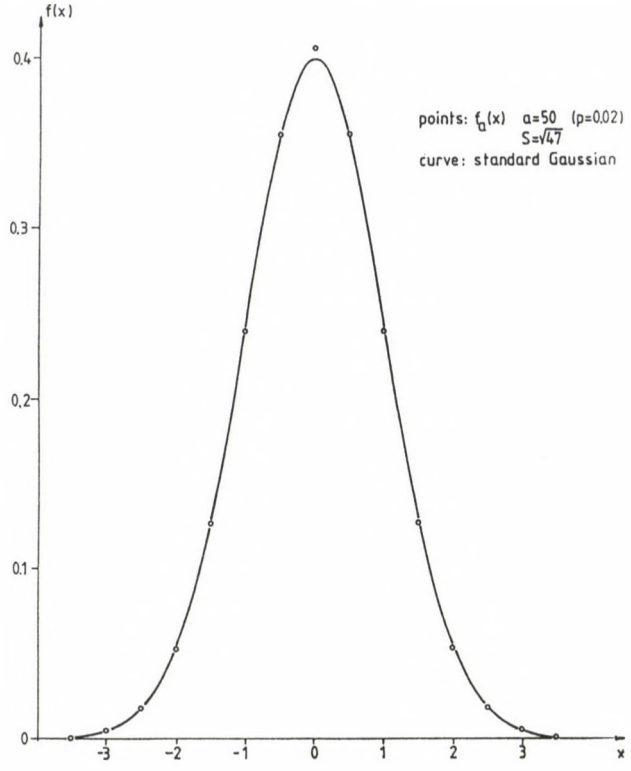


Fig. 6. Density function of the standard Gaussian distribution (curve) and the values of the density function of the $f_a(x)$ distribution for a = 50 with the same scatter (circles). The latter has according to $f_t(p)$ in Fig. 5 a significant probability; the two near types can be only distinguished on the basis of a sample of about 8000 elements

probability of the occurrence of the type set characterized by $p < 0.2$ is too high, since this probability is almost exactly 1/5, or that the probability of somewhat more than 0.1 is too high which belongs to distributions qualified according to Csernyák (1991) as sterile ones if the density function of Eq. (12) is applied).

4.

Having in mind the previous chapters, it is easy to find a

global measure for the comparison of statistical algorithms: if the function e(p) gives the efficiency of a method, then the average efficiency defined as

$$\bar{e} = \int_0^{\infty} e(p) \cdot f_t(p)\, dp \qquad (13)$$

is such a measure as in it the efficiencies, e(p), are multiplied with the corresponding probability of occurrence $f_t(p)$ of the error distribution functions characterized by the type parameter p, i.e. the averaging of the efficiencies is made using them as weights. (Equation (13) can be easily generalized for type-sets with more parameters, or even for those defined more generally, where a measure of probability is defined; this generalization is not dealt with here.)

The curves in Fig. 7 represent the integrands of Eq. (13) for different algorithms of the determination of the location

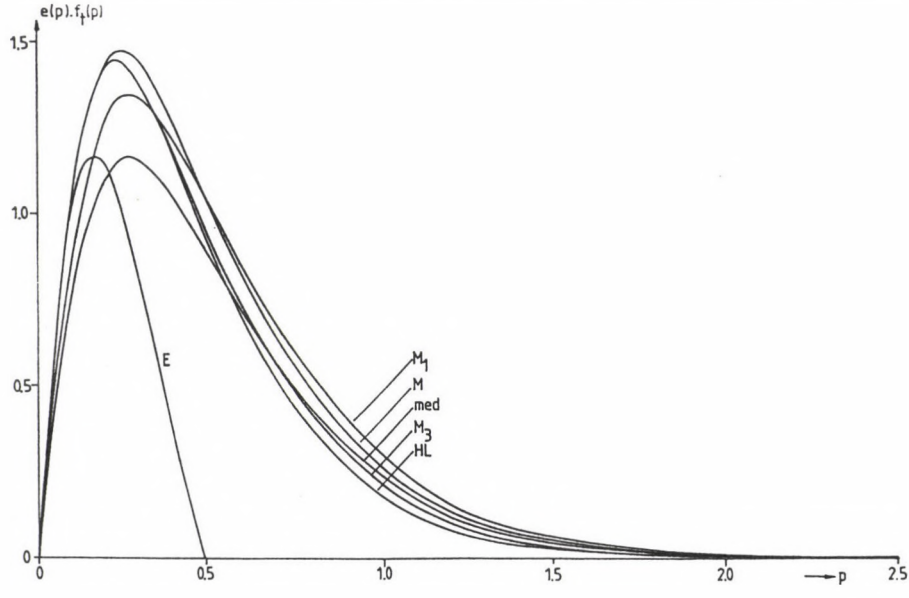


Fig. 7. The curves of the integrands in the formula of the average efficiency $\bar{e}$ (Eq. 13) for different statistical methods

parameter. The results of the calculation of the average efficiency are given in Table I.

Table I

Name of the statistical method	Notation	Average efficiency e percent
Calculation of the arithmetic mean	E	36
Determination of the sample median	med	80
Hodges-Lehman estimation	HL	85
Most frequent value k=3	M_3	90
" " " k=1	M_1	94
" " " (standard) k=2	M	96

It is interesting to compare different degrees of robustness for methods were e_{max} is nearly or exactly 100 percent by the values of the average efficiency. The maximum efficiencies of the HL-estimation and of the M_3-calculation are nearly at the same location p_{max} (see the corresponding curves in Figs 2 and 3) but due to the higher robustness of M_3 (R = 3.7), the average efficiency of this estimation is significantly higher ($\bar{e}_M$ = 90 percent) than that of the Hodges-Lehmann-estimation where the robustness is somewhat less, R = 3.3, but the average efficiency is only slightly over 85 percent.

The table suggests that the maximum average efficiency of the standard method of the determination of the most frequent value is connected to the fact that its maximum lies at the same place as the maximum occurrence frequency of the types, characterized by the density function $f_t(p)$. This fact is partly self-evident (as it is clear that the standard method

should be defined so that its maximum be there where that of $f_t(p)$ is), partly it does not show the whole background of the situation. This becomes evident if the average efficiency of an algorithm having a location of p_{max} far from p = 0.25, namely that belonging to the M_1-calculation, $\bar{e}$ = 94 percent is compared with the average efficiency 36 percent of the E-calculation. In latter case the maximum locations of $e_E(p)$ and of $f_t(p)$ are in vane near to each other: in a distance of 0.25. In other words: the distance between the maximum locations of $e_E(p)$ and $f_t(p)$ is three times less than the distance between e_M (p) (p_{max} = 1) and $e_E(p)$, the decisive factor in the difference of the average efficiencies (36 and 94 percent, respectively) is the consequence of the difference in the robustness; let us remember, the E-calculation was only characterized by R = 1.

The basic importance of a sufficiently large robustness appears for the practice in the high average efficiencies. It should be considered that this measure could be (also) used as a measure of the robustness itself. The integral in Eq. (13) takes namely the whole efficiency e(p) into consideration while R has a point-like character (nevertheless the point-like information can be used in a rather wide range of types, as e.g. a comparison of the efficiency curves of Fig. 3 shows with Fig. 58 by Steiner (1988), where the curves $e_{HL}(p)$ and e(p) can be compared in a four times wider range). On the other hand while R (or any other parameter which indicates exclusively the flatness of the e(p)-curve) is a practically useless measure of the robustness due to the rather low value of e_{max}, the average efficiency $\bar{e}$ takes this property into account, it influences in a negative sense the determination of the robustness (let us remind to the fact that $\bar{e}_{med}$ = 80 percent in spite of a high R due to the flatness of the efficiency curve). That means that we propose the use of the average efficiency also as measure of the robustness in a wide range of types; the average efficiency should, however, used primarily as a general parameter characterizing the usefulness of statistical algorithms.

REFERENCES

Csernyák L 1991: Publications of the Technical University for Heavy Industry, Series D, Natural Sciences, 35, No. 4.

Dutter R 1987: Mathematische Methoden in der Montangeologie. Vorlesungsnotizen 1986/87. Leoben

Hajagos B 1988: Publications of the Technical University for Heavy Industry, Series A, Mining, 44, 217-230.

Hajagos B 1985: Publications of the Technical University for Heavy Industry, Series A, Mining, 40, 225-238.

Hajagos B, Steiner F 1991a: Publications of the Technical University for Heavy Industry, Series D, Natural Sciences, 35, No. 4.

Hajagos B, Steiner F 1991b: Publications of the Technical University for Heavy Industry, Series D, Natural Sciences, 35, No. 4.

Hajagos B, Steiner F 1991c: Publications of the Technical University for Heavy Industry, Series D, Natural Sciences, 35, No. 4.

Hampel F R, Ronchetti E M, Rousseeuw P J, Stahel W A 1986: Robust Statistics. Wiley, New York

Huber P J 1981: Robust Statistics. Wiley, New York

Kendall M G, Stuart A 1958, 1961, 1966: The Advanced Theory of Statistics. Griffin, London, Vol. I, II, III

Kerékfi P 1978: Alkalmazott Matematikai Lapok, 4, 327-357.

Landy I, Lantos M 1982: Publications of the Technical University for Heavy Industry, Series A, Mining, 37, 29-32.

Newcomb S 1886: American Journal of Mathematics, 8, 343-366.

Romanowski M 1964: Bulletin Géodésique, 73.

Steiner F 1982: Publications of the Technical University for Heavy Industry, Series A, Mining, 37, 17-28.

Steiner F 1988: A special double number of the Journal Geophysical Transactions, 34, 139-260.

Steiner F ed. 1990a: The Most Frequent Value. Akadémiai Kiadó, Budapest

Steiner F 1990b: Introduction to Geostatistics (in Hungarian). Tankönyvkiadó, Budapest

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 153-182 (1991)

THE P^*-NORM

B Hajagos and F Steiner

Department of Geophysics, University for Heavy Industry, H-3515 Miskolc, Egyetemváros, Hungary

[Manuscript received February 23, 1990]

The inversion means mathematically a search for model parameters which yield computed values, the deviations of which from the measured ones are characterized by a minimum norm. For a long time exclusively the L_2-norm of the deviations has been used (due to computational advantages); from the generalized L_p-norms the application of the L_1-norm is only spreading nowadays more and more. (This norm is optimum at the Laplace distribution of errors, contrary to the L_2-norm which yields optimum at the Gaussian distribution.)

The present paper suggests the use of the P_k^*-norm in three different realizations as a variant of the recently defined P_C-, P- and P_J-norms with greater resistance. The P_C^*-norm belonging to $k = \sqrt{3}$ is most advantageously used in the neighbourhood of the Cauchy-distribution, the P-norm belonging to $k = 2\sqrt{3}$ in the neighbourhood of the geostatistical distribution, and the P_J^*-norm belonging to $k = 3\sqrt{3}$ is optimum in the vicinity of the Jeffreys distribution. The comparisons are promoted by tables and figures (e.g. curves of the efficiencies).

Keywords: inversion; most frequent value; norm; statistical efficiency; type of the error distribution

1. MOST IMPORTANT NOTIONS AND FACTS IN CONNECTION WITH THE M_k-ESTIMATION

For the sake of easier understanding of the new proposals and statements in this paper, let us first give a short outline of the statistical estimation method called „M_k-estimation".

If the value of T for which holds:

$$T = \frac{\sum_{i=1}^{n} \varphi(x_i - T) \cdot x_i}{\sum_{i=1}^{n} \varphi(x_i - T)} \qquad (1)$$

is determined on the basis of a sample x_1, x_2, ... x_i, ... x_n with the help of the function $\varphi(\cdot)$ which is symmetrical to the origin and has only one maximum then this value characterizes the location of the gathering of the data, i.e. it is - in the most general sense of the word - the most frequent value; this fact was shown in details by Steiner (1988a). (The determination of T is made most advantageously by iterations.) If one has a sufficiently large number of data to characterize the distribution of the data x by their probability density function f(x), then Eq. (1) is substituted by the equation

$$T = \frac{\int x \cdot \varphi(x-T) \cdot f(x)dx}{\int \varphi(x-T) \cdot f(x)dx} \qquad (2)$$

which can also be read as a prescription for iteration.

The analytical form of the weight function φ naturally defines the process for the definition of T, i.e. that of the estimation of the location parameter, but within this process the width of the interval around the origin has eminent importance where the function has values differring significantly from zero. This interval should be fitted to the „most dense" interval of the data set x_1, ... x_n what is most easily realized by determining the dihesion of the data set, ε :

$$\varepsilon^2 = \frac{3 \sum_{i=1}^{n} \frac{(x_i-T)^2}{[\varepsilon^2+(x_i-T)^2]^2}}{\sum_{i=1}^{n} \frac{1}{[\varepsilon^2+(x_i-T)^2]^2}} \qquad (3)$$

where Eq. (3) can be again considered as a prescription for an iteration (see Eq. 28 in Steiner 1988b), and the product of this dihesion with a number k is used as scale parameter in the weight function $\varphi(\cdot)$. If the density function f(x) is given, then Eq. (3) is substituted by

$$\varepsilon^2 = \frac{3\int_{-\infty}^{\infty} \frac{(x-T)^2}{[\varepsilon^2+(x-T)^2]^2} f(x)dx}{\int_{-\infty}^{\infty} \frac{1}{[\varepsilon^2+(x-T)^2]^2} f(x)dx} . \qquad (4)$$

If a function φ is to be used for which the necessary number of operations is a minimum, then the weight function

$$\varphi(x-T) = \frac{1}{(k\varepsilon)^2+(x-T)^2} \qquad (5)$$

should be applied. Using this function φ in Eq. (1), the final result T of the double iterations according to Eqs (1) and (3) is called in a more restricted, but correctly defined sense as the most frequent value. This fact is emphasized by denoting this value by M instead of T in the standard case when k = 2, else the notation M_k is used.

The density function f(x) of the probability distribution from which the sample is taken, determines the most advantageous value k. Error distributions occurring in realty can be well modelled by means of the so-called „supermodel $f_a(x)$":

$$f_a(x) = n(a) \cdot (1+x^2)^{-a/2} \qquad (a > 1) \qquad (6)$$

(the norming factor n(a) is computed as

$$n(a) = \frac{\Gamma\left(\frac{a}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{a-1}{2}\right)}) \cdot \qquad (6a)$$

Equation (6) gives the density functions $f_a(x)$ in a standard form i.e. the location parameter T is substituted by 0 and the scale parameter S by the unity. The general form of the density function is obtained, as usual, by substituting x by (x-t)/S and by dividing the expression by S.

Some important special cases (characterized by different values of the type determining parameter a) are given in Table I.

Table I

value of the type parameter $\underline{a}$	name of the random distribution-type	density function in the standard case
2	Cauchy	$\frac{1}{\pi}\frac{1}{1+x^2}$
5	geostatistical	$\frac{3}{4}\frac{1}{(1+x^2)^2\cdot\sqrt{1+x^2}}$
9	Jeffreys	$\frac{35}{32}\frac{1}{(1+x^2)^4\cdot\sqrt{1+x^2}}$

If the type parameter a is increased above all bounds under the condition $S = \sqrt{a-3}$ then one obtains the density function of the standard Gaussian distribution:

$$f_G(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \quad . \tag{7}$$

Using the notation

$$s_i(k\varepsilon) = \int_{-\infty}^{\infty} \left[\frac{(k\varepsilon)^2}{(k\varepsilon)^2+(x-T)^2} \right]^i f(x)\, dx \tag{8}$$

the asymptotic scatter of the estimation of the most frequent value $T = M_k$ can be written for symmetrical distributions as (Hajagos 1985):

$$A(M_k) = k\varepsilon \frac{\sqrt{S_1(k\varepsilon) - S_2(k\varepsilon)}}{2S_2(k\varepsilon) - S_1(k\varepsilon)} \quad . \tag{9}$$

As the minimum asymptotic variance A^2_{min} the so-called (Cramér-Rao-bound) of the determination of the location parameter for

the family $f_a(x)$ is:

$$A^2_{min} = \frac{a+2}{a(a-1)} \tag{10}$$

(Hajagos 1985), the efficiencies are easily computed for any $f(x) = f_a(x)$ and k by means of:

$$e(M_k) = \frac{a+2}{a(a-1)A^2\ (M_k)} \quad . \tag{11}$$

Based on the supermodel $f_a(x)$, so the question can be answered now for which type of distribution has the statistical algorithm determining the most frequent value M_k a maximum efficiency at a given value of k. One gets

$$\begin{aligned} &\text{at } k = 1, \text{ for the Cauchy,} \\ &\text{at } k = 2, \text{ for the geostatistical,} \\ &\text{at } k = 3, \text{ for the Jeffreys} \end{aligned} \tag{12}$$

distribution maximum efficiency, i.e. 100 percent (e.g. Steiner 1990a). Distribution types which differ significantly from that of maximum efficiency have naturally rather lower efficiencies, but the decrease is slower than that of the arithmetic mean described by

$$e(E) = \frac{(a+2)\ (a-3)}{a(a-1)} \tag{13}$$

if one gets further and further away from the Gaussian distribution. (The type distance is with close approximation proportional to the differences of the values $p = 1/(a-1)$ for pairs of elements of the supermodel $f_a(x)$ as shown by Hajagos and Steiner 1991a.) This robust character enables to use only the values k = 1, 2 and 3 (with the exception of very special cases). In order to identify the corresponding distribution types, M_1 and M_3 will be substituted (in the index) by M_C, and M_J, respectively. In the case of the standard value k = 2 no index is used (in this particular case the geostatistical

distribution has the highest, 100 percent efficiency for the computation of the most frequent value). As the occurrence probabilities of the distributions in geosciences are modelled by the density function

$$f(p) = 16pe^{-4p} \tag{14}$$

(where similarly $p = 1/(a-1)$, Steiner 1990b) the average efficiency (expressing also the robustness) is defined by

$$\bar{e} = \int_0^\infty f(p)e(p)dp \quad . \tag{15}$$

Having a value of $\bar{e} = 96$ percent in the case $M_k = M$, really the value $k = 2$ should be used in the cases when one has no information about the type of distribution. The function $e(p)$ in Eq. (15) is naturally the efficiency $e(M_k)$ in Eq. (11) obtained by the values $a = 1+1/p$ and by $A(M_k)$ according to Eq. (9) where $s_i(k\varepsilon)$ (computed according to Eq. (8) are obtained with $f(x) = f_a(x)$).

In the case of the family $f_a(x)$, the scale parameter $S = k\varepsilon$ can also be obtained directly according to the iteration prescription

$$S^2 = (a+1)\, S^2 \left(\frac{S_1(S)}{S_2(S)} - 1 \right) \quad . \tag{16}$$

thus values of $a = 2$; 5 and 9 are to be used in Eq. (16) in turn for the computation of M_C, M and M_J. If $f(x)$ is no element of the supermodel $f_a(x)$, the statistical procedure using Eq. (16) is theoretically not identical with that which uses the dihesion multiplied by k, in the practice, however, both procedures mean the same thing. Consequently, they should be considered as two variants of a statistical method resulting with good approximation in the same results.

The expression

$$n_{eff} = \sum_{i=1}^{n} \frac{(k\varepsilon)^2}{(k\varepsilon)^2+(x_i-T)^2} \tag{17}$$

is the effective number of data (points), as if x_i is very far from T with respect of ($k\varepsilon$), then fractions with the practical value zero belong only to x_i in the sum of Eq. (17); and if x_i is close to T, then the value belonging to x_i has a value of nearly 1. Thus n_{eff} is always less than n: Eq. (17) eliminates the values which do not play a significant role in Eq. (1) due to their low weights, therefore it is correct to consider the sum in Eq. (17) as the effective number of data.

2. CONSIDERATIONS ABOUT THE M_k^*-ESTIMATION

The number of the operations increases only by just one if in the computation of the values in the weight function, the square of the expression in Eq. (5) is considered as weight:

$$\varphi^*(x-T) = \frac{1}{\left[(k^*\varepsilon)^2+(x-T)^2\right]^2} \quad . \tag{18}$$

If this weight function appears in Eq. (1) instead of $\varphi(\cdot)$, then the location parameter T is estimated by M^* (Steiner 1988b, Appendix II); more exactly that means that in such a case that value is estimated by the iteration prescribed in Eq. (1) which is unambiguously attributed by Eq. (2) to a given f(x), and in this case T is substituted in Eq. (2) by M^*, too. As Eq. (18) is an unimodal, symmetric weight function, too, M^* is also a location parameter of a most frequent value-character. If the coefficients of ε in Eqs (5) and (18) are to be distinguished, then the factor in Eq. (18) will be denoted by k^*.

The most advantageous value of k is determined again by the supermodel $f_a(x)$. The decision is made on the basis of the maximum efficiency of the possible distribution types, as Eq. (134) in Steiner (1988b) can be rewritten as

$$A(M_k^*) = k\varepsilon \frac{\sqrt{S_3(k\varepsilon)-S_4(k\varepsilon)}}{4S_3(k\varepsilon) - 3S_2(k\varepsilon)} \tag{19}$$

and the efficiency is obtained simply from the formula

$$.e(M_k^*) = \frac{a+2}{a(a-1) \cdot A^2(M_k^*)} \qquad (20)$$

being analogous with Eq. (11).

Let us try at first $k^* = 1$; the curve of the efficiency vs. a is shown in Fig. 1. The maximum efficiency lies at a = 1.1, this type of distribution is characterized (according to Eq. 6) by heavy flanks and by a strong peak in the centre. Thus this distribution can only be used for modelling real distributions in exceptional cases. If regional parts of a Bouguer map are approximated by a simple analytical formula, then the deviations of the Bouguer anomalies from this

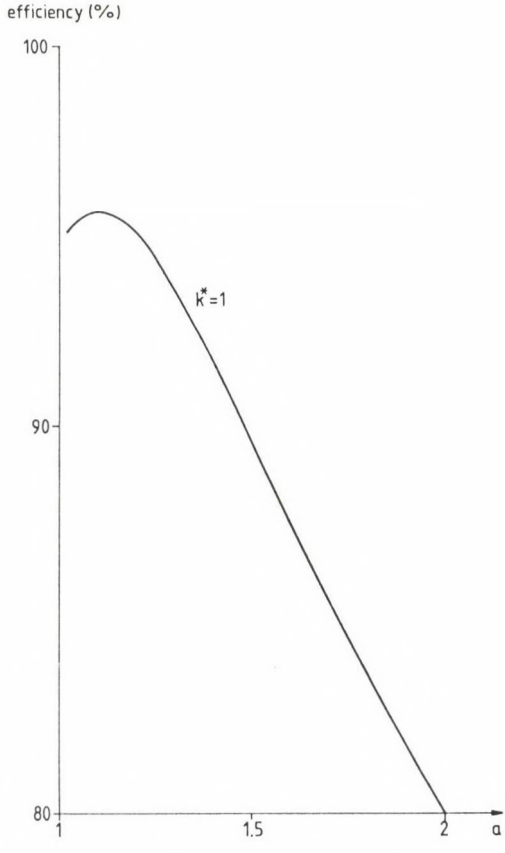


Fig. 1. If k^* has a small value, the maximum place of the efficiency curve lies far from the range of the often occurring type of distributions beginning at about a = 2 (the meaning of the type parameter a is given by Eq. 6)

approximative function can be in the range of the local effects by one order of magnitude greater than the deviations for the pure regional parts of the map, and a distribution with a strong central peak and with heavy flanks is expected for the whole map. It is thus not surprising that Zilahi-Sebes (1987) reported best results exactly with $k^* = 1$ for such a problem.

Let us return to the general statistical problem: to find k^*-values which can be used in the most cases.

In order to use the results in the theory of the M_k-estimations, let us compare the dihesion factors which are optimum ones for an $f_a(x)$ belonging to a value of a in the case of M_k^* and of M_k-estimations. (These k^*- and k-values, respectively, can be obtained as maximum place of the efficiency curve, if the efficiencies are computed for a value of a vs. the corresponding coefficient, or by a special method Steiner 1988a.)

Figure 2 represents the curve of the quotient of these

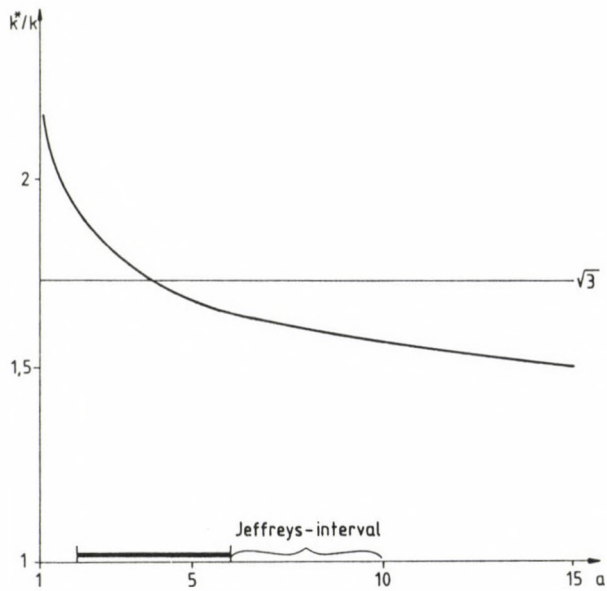


Fig. 2. The optimum value of k in Eq. (5) and of k^* in Eq. (18) can be determined for the distribution types having the type parameter a. The ratio k^*/k is around $\sqrt{3}$ for the practically occurring types of distributions

factors, i.e. the values k^*/k vs. a. The quotient k^*/k is quite well averaged by the value $\sqrt{3}$ over the whole range, but it is even more important that in the interval of the most usually occurring a-s (thick line) - and even in the Jeffreys-interval, too - $\sqrt{3}$ is a good approximation of the values of the quotient k^*/k (at a = 2 this value is only by 10 percent lower, while at a = 5 by 3 percent, at a = 6 by 6 percent, and at a = 10 by it is 10 percent higher).

These considerations show that the value

$$k^* = k\sqrt{3} \tag{21}$$

characterizes the connection between the factors k and k^* being optimum ones for a given value of a. This can be expressed also with the corresponding scale parameters (the latter ones being the products of the dihesion with the corresponding factors):

$$S^* = S\sqrt{3} \ . \tag{22}$$

Be S = 1 (what can be reached in each case by selecting a suitable unit for x) and for sake of simplicity the value of T be 0. The corresponding weight functions $\varphi(.)$ and $\varphi^*(.)$ (Eqs 5 and 18) are (taking into account Eq. 22) simply:

$$\varphi(x) = \frac{1}{1+x^2} \quad \text{and} \quad \varphi^*(x) = \frac{1}{(3+x^2)^2} \ . \tag{23}$$

As any weight function multiplied by a constant yields the same estimation of the location parameter T according to Eq. (1), the visual comparison should be promoted by taking 9-times the value $\varphi^*(x)$ obtained from Eq. (23). Both weight functions have so a maximum with the value 1. Figure 3 shows the weight functions; the small differences between the two curves suggest that M_k and M_k^* do not differ significantly. Since the values of $\varphi^*(x)$ are however in the case of great x-values significantly less than those of the weight function $\varphi(x)$, the estimation M_k^* is a method for the determination of the most frequent value which is much less sensitive for outliers (Appendix II in

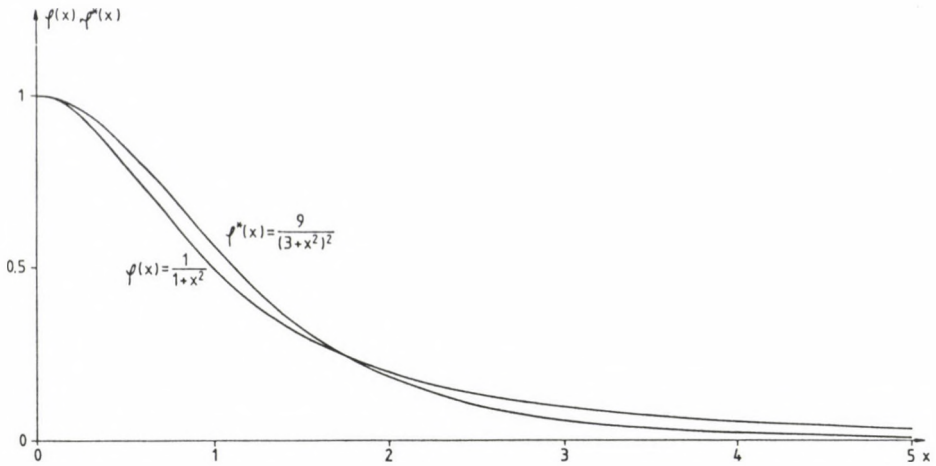


Fig. 3. There are no big differences between the corresponding weight functions φ and φ^*; it is, however, clear that the weight function φ^* defines a much more resistant statistical procedure than weight function φ does (see the sections of the curves at high x-values)

Steiner 1988b gives a more exact proof of this statement by IC-functions). This fact shows the importance of not only the M*-estimation, but of the introduction of the P*-norm, too as a generalization (see in the next chapter).

It has been mentioned that (with the exception of very special cases) it is sufficient to treat the cases k = 1, 2 and 3. Consequently, in the case of the M*-estimations - according to Eq. (21) - it is sufficient to introduce the values $k^* = \sqrt{3}$, $2\cdot\sqrt{3}$, $3\cdot\sqrt{3}$, and it is expected that the maximum efficiency is obtained near to the distribution types

$$\left.\begin{array}{ll} k^* = \sqrt{3} & \text{Cauchy,} \\ k^* = 2\cdot\sqrt{3} & \text{geostatistical,} \\ k^* = 3\cdot\sqrt{3} & \text{Jeffreys} \end{array}\right\} \qquad (24)$$

and this is expressed by the notations M^*_C at $k^* = \sqrt{3}$, and M^*_J at $k^* = 3\cdot\sqrt{3}$ analogously to the M-estimations. The standard character of the case $k^* = 2\cdot\sqrt{3}$ is emphasized by writing simply

M^*, i.e. simply without index.

The curves of the efficiencies determined for the supermodel $f_a(x)$ from Eq. (20) (and naturally from Eq. 19) for the three coefficients in Eq. (24) are presented on Fig. 4. (Here k is used instead of k^* because here no distinction is made between M_k and M_k^*-estimations, and further - according to the next chapter - P^* is used with the corresponding index instead of M^*, as these curves of efficiencies are very instructive not only for the calculation of a single T-value, but also for that of several unknowns.) In the case $k = \sqrt{3}$ the location of the maximum, a_{max} is by only about 0.2 less than the value a = 2 characterizing the Cauchy-distribution, and the maximum efficiency is more than 97.57 percent. It should be remarked about this value that an efficiency of exactly

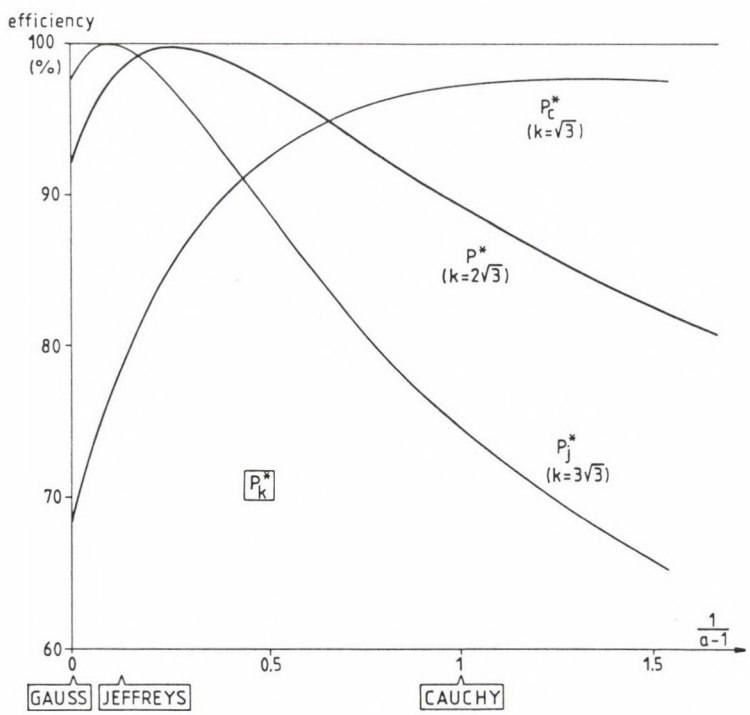


Fig. 4. The efficiency vs. type of distribution for the three values of the factor k defining completely P^* ($k = 2\cdot\sqrt{3}$ is considered as standard case)

100 percent must not be expected for the M^*-estimation and for the elements of the family $f_a(x)$, as the M^*-estimate is for no k value a maximum likelihood estimate of an $f_a(x)$-type distribution. The good approximation of the 100 percent efficiency in the standard case $k = 2\cdot\sqrt{3}$ is shown in the enlarged Fig. 5: the maximum efficiency is 99.7 percent giving a further proof of the affinity of the M_k- and M_k^*-estimations. (The value $a_{max} = 4.95$ is for all practical applications identical with the value a = 5, characterizing the geostatistical distribution, thus confirming the central line of Eq. (24).) The curve P_J^* in Fig. 4 ($k = 3\cdot\sqrt{3}$) has a maximum at a = 10.9, i.e. in a distance from the Jeffreys-distribution which is - expressed in type distance - negligible. The maximum is only by 0.03 percent less than 100 percent. It is characteristic for the robustness that the optimum statistical

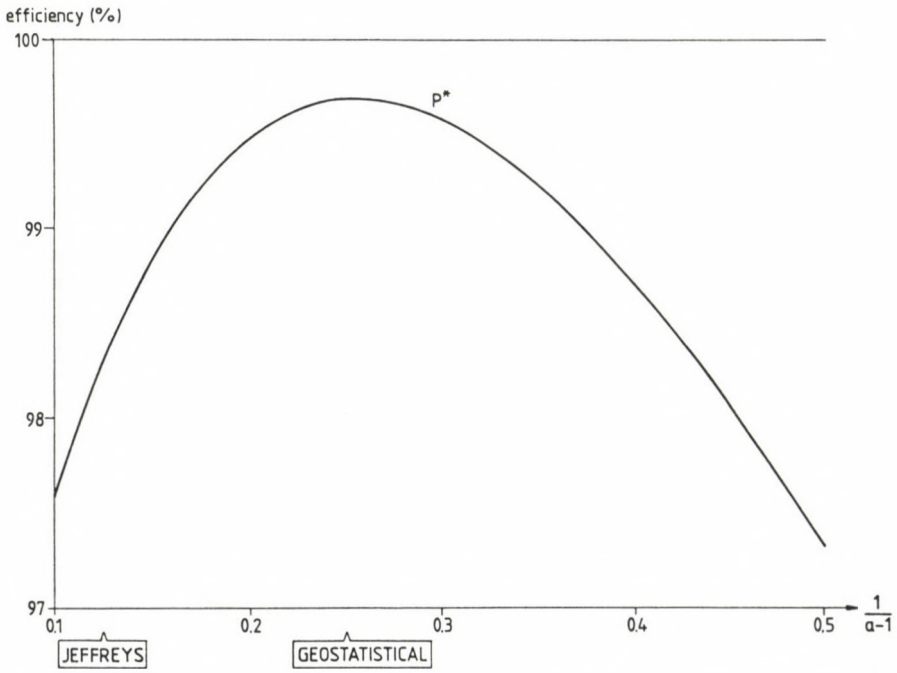


Fig. 5. Central zone of the efficiency curve belonging to P^* (magnification of a part of Fig. 4)

algorithm for distributions similar to the Jeffreys-distribution has an efficiency of about 75 percent for the Cauchy-distribution; the optimum for the Cauchy-distribution M_C^* ($k = \sqrt{3}$) has again an efficiency of 78.6 percent at distributions similar to the Jeffreys one (the efficiency is 68.6 percent at the sterile Gaussian distribution, at the geostatistical distribution it is 85.2 percent).

The robust character of the standard case ($k = 2 \cdot \sqrt{3}$) can be characterized by similar values: the efficiency is more than 90 percent in the complete Gauss-Cauchy-interval, and 98.3 percent at the Jeffreys-distribution (at the Gaussian one, 92.2 percent); at $a = 3$, where the averaging has an asymptotically zero efficiency it is 97.3 percent, etc. but the robust character can be also characterized in this case by a single numerical value, namely by the average efficiency e (Eq. 15) similarly to the characterization of the standard case M (i.e., M_K for $k = 2$). The results shows that the average efficiency is for the standard M^*-estimation 96 percent, i.e. just the same value as for the standard variant of the original most frequent value calculation.

3. DEFINITION OF THE P_k^*-NORMS — COMPARISON WITH THE P_k-NORMS

Both in solid Earth geophysics and in exploration geophysics responses to open problems are mostly given by model parameters which yield computed values best fitting to the measured ones. These model parameters which are accepted as most likely ones are obtained by the so-called inverse methods (recently surveyed by Tarantola 1987).

In case of inverse methods it is evident that the degree of approximation of the measured data system by the actually calculated values is judged on the basis of an arbitrarily chosen norm - at least it seems to be so for a first glance. Steiner (1990a and 1990b) gave the general L_p-norm of the deviations for n measurement data x_i - in accordance with Tarantola (1987) - in the following form:

$$L_p\text{-norm: } \left(\frac{1}{n} \sum_{i=1}^{n} |x_i - T(\bar{m};\bar{y}_i)|^p \right)^{1/p} \tag{25}$$

where the calculated values are obtained to the practically accurately (errorfree) known $\bar{y}_i$ vectors by an analytically a priori known function T to the given model parameter values $\bar{m} = m_o, m_1, m_2, \ldots m_j, \ldots m_{jo}$. The deviations of the measured x_i data from these calculated ones are denoted by capital letters, i.e.:

$$X_i = x_i - T(\bar{m};\bar{y}_i) \ . \tag{26}$$

By means of Eq. (26) the L_p-norm is simply:

$$L_p: \quad \left(\frac{1}{n} \sum_{i=1}^{n} |X_i|^p \right)^{1/p} \tag{27}$$

where the two best known cases belong to p = 1 and to p = 2:

$$L_1: \quad \frac{1}{n} \sum_{i=1}^{n} |X_i| \tag{28}$$

and

$$L_2: \quad \sqrt{\frac{1}{n} \sum_{i=1}^{n} X_i^2} \ . \tag{29}$$

The norms judge the deviations in respect of their totality as big or small, and if some kind of norm has been chosen, then the result of the inversion is the parameter-vector, for which the chosen norm of the deviations X_i has a minimum.

The by far most usual method in inversion is the minimizing of the L_2-norm, i.e. the method of the least squares. The gist and at the same time the limitation and dangers of this choice are illustrated by quotations from

Tarantola (1987): „Least squares are so popular for solving inverse problems because they lead to the easiest computations. Their ... drawback is ... their strong sensitivity to a small number of large errors (outliers) in a data set" (p. 187), "... least-squares theory is justified only when data uncertainties ... can be described using Gaussian statistics" (p. 410). By choosing the presently much less usual L_1-norm for minimization one obtains a resistant (unsensitive for a small number of outliers) inversion method which is an optimum in the case of Laplace error distributions. In the most simple case, if T = const., the minimum of the L_1-norm is at the sample median, and that of the L_2-norm is at the arithmetic mean of the sample elements.

If the P_k-norm is introduced with the following definition:

$$\varepsilon \left\{ \prod_{i=1}^{n} \left[1 + \left(\frac{X_i}{k\varepsilon} \right)^2 \right] \right\}^{\frac{1}{2n}} \tag{30}$$

where ε is the dihesion of the deviations X_i, i.e. the equation

$$\varepsilon^2 = \frac{3 \sum_{i=1}^{n} \frac{X_i^2}{(\varepsilon^2 + X_i^2)^2}}{\sum_{i=1}^{n} \frac{1}{(\varepsilon^2 + X_i^2)^2}} \tag{31}$$

is fulfilled (compare Eq. 3), then in the simplest T = const. case the minimum of the P_k-norm vs. T (under the condition Eq. 31) is to be found exactly as the determination of the most frequent value M_k is made.

This fact is proven as follows:

1. the expression is multiplied by k,
2. $(k\varepsilon)$ is included into the product by multiplying each factor by $(k\varepsilon)^2$,
3. the (2n)-th power of the expression is calculated,
4. the expression is logarithmed.

The expression

$$\sum_{i=1}^{n} \left[\ln(k\varepsilon)^2 + X_i^2 \right] \tag{32}$$

obtained so has obviously its minimum at the same place where the P_k-norm has (see Eq. 30), as in the transformations no steps occur which would influence the minimum place: Equation (32) is therefore from the point of view of the location of the minimum equivalent with the P_k-norm given by Eq. (30). So far X_i was used in a most general sense according to Eq. (26). By using now

$$X_i = x_i - T$$

- as supposed originally, when the statement about M_k has been considered - the minimum after T of Eq. (32) follows from the zero value of the derivate after T of

$$\sum_{i=1}^{n} \ln \left[(k\varepsilon)^2 + (x_i - T)^2 \right] \tag{33}$$

i.e. from the equation:

$$\sum_{i=1}^{n} \frac{x_i - T}{(k\varepsilon)^2 + (x_i - T)^2} = 0 \tag{34}$$

which is after some rearrangement

$$T = \frac{\sum_{i=1}^{n} \dfrac{x_i}{(k\varepsilon)^2 + (x_i - T)^2}}{\sum_{i=1}^{n} \dfrac{1}{(k\varepsilon)^2 + (x_i - T)^2}} . \tag{35}$$

And this is in fact Eq. (1) defining the most frequent value with the weight function of Eq. (5).

An expression can also be easily given which leads in the case T = const. similarly (i.e., determining of the minimum place) to Eq. (1) with the weight function of Eq. (18), i.e. to the M^*-estimation:

$$\frac{2\varepsilon\,(k^2/3+1)}{n}\cdot\sum_{i=1}^{n}\frac{x_i^2}{(k\varepsilon)^2+x_i^2}\,. \tag{36}$$

The constant of the just introduced P_k^*-norm seems at first to be superfluously complicated but not only the minimum place has significance (yielding the solution of the inverse problem), but their minimum value too, which is considered as an error characteristics.

Hajagos and Steiner (1991b) show in detail how the values of the minimum of Eq. (36) characterize the uncertainty in the system of the deviations.

The sum in Eq. (36) can be written as

$$n-\sum_{i=1}^{n}\frac{(k\varepsilon)^2}{(k\varepsilon)^2+x_i^2} \tag{37}$$

i.e. (by taking Eq. 17 also into acount) as

$$n - n_{eff}\,. \tag{37a}$$

That means that by minimizing the P_k^*-norm, the effective data number according to Eq. (17) is maximized. If now T = const, then the maximum place of the expression

$$\frac{n_{eff}}{(k\varepsilon)^2}=\sum_{i=1}^{n}\frac{1}{(k\varepsilon)^2+(x_i-T)^2} \tag{38}$$

could also be looked for instead of the minimum place of the P_k^*-norm; here the derivate of the sum is zero:

$$\sum_{i=1}^{n}\frac{x_i-T}{\left[(k\varepsilon)^2+(x_i-T)^2\right]^2}=0 \tag{39}$$

after rearrangement one gets:

$$T = \frac{\sum_{i=1}^{n} \frac{x_i}{\left[(k\varepsilon)^2 + (x_i - T)^2\right]^2}}{\sum_{i=1}^{n} \frac{1}{\left[(k\varepsilon)^2 + (x_i - T)^2\right]^2}} \qquad (40)$$

and this is Eq. (1) written with the weight function of Eq. (18) yielding really $T = M_k^*$.

The proposed values k = 1, 2 and 3 for the most frequent values M_k correspond to M_C, M and M_J, it is therefore advantageous to use the same indices for the P_k-norm for these k-values, i.e. P_C, P and P_J, referring to the case k = 2 as to the standard one (where no index is written) and to the preference of the Cauchy- and Jeffreys-distributions at k = 1, and at k = 3, respectively. Similarly, the P_k^* norm corresponding to $k = 2\cdot\sqrt{3}$ (considered again as a standard one) is denoted simply by P^*, and those preferring the Cauchy- and Jeffreys-distributions at $k = \sqrt{3}$ and $k = 3\cdot\sqrt{3}$ by P_C^* and P_J^*, respectively. It is advisable to summarize these in a table (Table II) where the integral forms of the norms are also given, being valid if not only a sample with n elements is known from X, but its density function, f(X), too. (ε-s are the dihesions of the deviations X_i which fulfil Eq. 31.) In the case of T = const the relative efficiency of the standard cases was 96 percent both for P^* and for the P-norm (see in the Table II the central column referring to the standard case). If one has no preliminary information about the type of the distribution, the actual decrease of the efficiency is generally very slight using the P- or the P^*-norm.

The scale parameter $S = k\varepsilon$ can be directly obtained for a distribution $f_a(x)$ with a known a immediately from Eq. (16). Alternatively the sum-equivalent of Eq. (16),

$$S^2 = (a+1)\frac{\sum_{i=1}^{n} \frac{1}{S^2+X_i^2}}{\sum_{i=1}^{n} \frac{1}{\left[S^2+X_i^2\right]^2}} - S^2 \qquad (41)$$

can be considered as a prescription for iteration, namely using

Table II.

	Cauchy	Geostatistical	Jeffreys
P_k-norm	$P_C = \varepsilon \cdot \left\{ \prod_{i=1}^{n} \left[1+(X_i/\varepsilon)^2\right] \right\}^{\frac{1}{2n}}$	$P = \varepsilon \cdot \left\{ \prod_{i=1}^{n} \left[1+(X_i/2\varepsilon)^2\right] \right\}^{\frac{1}{2n}}$	$P_J = \varepsilon \cdot \left\{ \prod_{i=1}^{n} \left[1+(X_i/3\varepsilon)^2\right] \right\}^{\frac{1}{2n}}$
	$P_C = \varepsilon \cdot \exp\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln\left[1+(X/\varepsilon)^2\right] f(X)dX \right\}$	$P = \varepsilon \cdot \exp\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln\left[1+(X/2\varepsilon)^2\right] f(X)dX \right\}$	$P_J = \varepsilon \cdot \exp\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln\left[1+(X/3\varepsilon)^2\right] f(X)dX \right\}$
P_k^*-norm	$P_C^* = \frac{4\varepsilon}{n} \cdot \sum_{i=1}^{n} \frac{X_i^2}{3\varepsilon^2+X_i^2}$	$P^* = \frac{10\varepsilon}{n} \cdot \sum_{i=1}^{n} \frac{X_i^2}{12\varepsilon^2+X_i^2}$	$P_J^* = \frac{20\varepsilon}{n} \cdot \sum_{i=1}^{n} \frac{X_i^2}{27\varepsilon^2+X_i^2}$
	$P_C^* = 4\varepsilon \int_{-\infty}^{\infty} \frac{X^2}{3\varepsilon^2+X^2} f(X)dX$	$P^* = 10\varepsilon \cdot \int_{-\infty}^{\infty} \frac{X^2}{12\varepsilon^2+X^2} f(X)dX$	$P_J^* = 20\varepsilon \cdot \int_{-\infty}^{\infty} \frac{X^2}{27\varepsilon^2+X^2} f(X)dX$

$$\text{for } P_C \text{ and for } P_C^*, \quad a = 2,$$
$$\text{for } P \text{ and for } P^*, \quad a = 5,$$
$$\text{for } P_J^* \text{ and for } P_J^*, \quad a = 9$$

(to obtain as results $S^2 = \varepsilon^2$, $S^2 = (2\varepsilon)^2$, and $S^2 = (3\varepsilon)^2$, respectively). This expressions will appear in φ directly, in φ^* multiplied by 3. (This solution has been mentioned at the end of Chapter 1, as a variant of the computation of the most frequent value.) In such a case the P-norms are used in the forms

$$\text{const.} \left\{ \prod_{i=1}^{n} [1+(X_i/S)^2] \right\}^{\frac{1}{2n}} \quad \text{and} \quad \text{const.} \left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln[1+(X/S)^2] f(X)dx \right. \quad (42)$$

respectively, for the search of the minimum place, the P^*-norms in the form

$$\text{const.} \sum_{i=1}^{n} \frac{X_i^2}{3S^2+X_i^2} \quad \text{and} \quad \text{const.} \int_{-\infty}^{\infty} \frac{X^2}{3S^2+X^2} f(X)dx \quad (43)$$

respectively.

A comparison of P_k^*-norms for given data sets is most simple with the corresponding P_k-norms from the point of view of the solution of the inverse problem if we consider the sum-equivalents of the latters given in Eqs (32) and (33). An excellent feature of the P_k^*-norm is that an outlier be at any great distance, the term corresponding to the actual X_i tends to a finite value, namely to unity, while the corresponding term in Eq. (33) increases with the increase of the distance above all limits without limits, too - the increase is, however, slower and slower (as follows from the character of the logarithmic function). (This fact explains the mentioned higher resistance of the M_k^*-estimation in comparison with the M_k-estimation.) The "increase without limits" in connection with the P_k-norm results for given deviation values by far smaller additive terms in the case of far away outliers than in the case of the L_1-norm with linear increase (see Eq. 28) not mentioning the sum in the L_2-norm (Eq. 29), where the additive terms increase parabolically, in an even increasing rate for

outliers in greater and greater distances, i.e., a single value gets a more and more deterministic role in the sum of the L_2-norm what results in the mentioned harmful sensitivity of the L_2-norm for outliers.

From the point of view of resistance and robustness the P_k- and P_k^*-norms differ significantly from the L_2-norm. What the often mentioned computational advantages of the L_2-norm concerns, however - they are valid in an important point of view as far in the minimization of the P^*- and P_k^*-norms, as in that of the L^2-norm in the most frequent by occurring practical applications.

This practical case is when the function $T(\bar{m};\bar{y})$ is in the parameters linear, i.e.

$$T(\bar{m};\bar{y}) = m_0+m_1 \cdot T_1(\bar{y})+\ldots+m_j \cdot T_j(\bar{y})+\ldots m_{jo} \cdot T_{jo}(\bar{y}) \qquad (44)$$

when T_j-s are arbitrary, but a priori known functions. In such a case the derivation of the L_2-norm yields after all unknown m_j-s a simple linear algebraic system of equations, the derivation of the P_k- and P_k^*-norms a system of equations which is also linear but weighted with

$$\varphi(X_i) = \frac{1}{(k\varepsilon)^2 + X_i^2} \qquad (45)$$

and

$$\varphi^*(X_i) = \frac{1}{\left[(k\varepsilon)^2 + X_i^2\right]^2} \qquad (46)$$

respectively. The weights depend through X_i on the parameter vector $\bar{m}$ - and therefore the system of equations can be only solved iteratively (a detailed description and explanation is given by Steiner (1988b at pp. 167 and 174-175) about the P_C-norm and about a given function $T(\bar{m},\bar{y})$; about the P_k^*-norm see Steiner 1988d). It is evident that even an iterative solution of an algebraic system of equations is much simpler than the search for the minimum place in the case of many variables. The L_1-norm does not possess this common peculiarity of the L_2-, P_k- and P_k^*-norms. On the other hand any norm, even the L_2-norm

leads to a many-variable general minimum problem if $T(\bar{m};\bar{y})$ does not possess the form given in Eq. (44), or it cannot be brought to such a form, respectively. In other words, the computational advantages of the L_2-, P_k- and P_k^*-norms are nearly the same, the iteration in the P_k- and P_k^*-norms is in each step the same as a weighted least squares computation; the weights come in the case of the P_k-norm from Eq. (45), in the case of the P_k^*-norm from Eq. (46).

4. THE PERMISSIBLE RATIO OF OUTLIERS IN COMPUTATIONS ACCORDING TO THE P- AND P^*-NORMS

The point of view of resistance has been mentioned already several times. It is justified question now that using different k-s and corresponding a-s - what amount of outliers is permissible in the measured data set without any influence of the final result. The answer is given for two important cases - and the results will help us in the general orientation, too.

4.1 The case of arbitrarily dispersed outliers

In one of the possible cases outliers are dispersed in all possible positions without concentrating in a density comparable to the concentration (around T) of the regularly appearing data (i.e., of the data without blunders). This case is dealt with in many details by Steiner (1988a) with the result that in the case of the variant using Eq. (41) the maximum quantity of permissible outliers is to be computed from

$$1 - \frac{(a+1)^2}{(a+1)^2 + 4(a+2)} \qquad (47)$$

meaning that in the case of $a \rightarrow 1$, their proportion may approximate even 75 percent. More interesting for the present study are the cases a = 2; 5 and 9:

variant a
- in the case of P_C and P_C^* max. 16/25 = 64 percent,
- in the case of P and P^* max. 7/16 = 43.75 percent,
- in the case of P_J and P_J^* max. 44/144 = 30.56 percent

arbitrarily dispersed outliers are permitted. There are no differences from the point of view of the permissible outliers betwen P_k and P_k^*, as according to Steiner (1988a) this maximum depends only on the value of a used in the determination of the scale parameter according to Eq. (41). As in the determinations of the dihesion a = 2, in all the norms P_k and P_k^* with k = 1; 2 and 3 and with $k^* = \sqrt{3}$; $2\sqrt{3}$ and $3\sqrt{3}$, respectively, the maximum ratio of permissible outliers is in all these cases 64 percent. That means that using the k-variant of the P and P*-norms, the maximum permissible ratio of outliers is 64 percent, in contrast to the a-variant where this ratio changes with the value of a.

If $a \rightarrow \infty$, in the limiting case the minimization of the P- and P*-norms is identical with the minimization according to the L_2-norm (as $f_a(x)$ converges to the Gaussian density function), see the summarizing tables in Steiner 1990a and 1990b (where $k \rightarrow \infty$, but this is a consequence of $a \rightarrow \infty$). This fact is, however, shown by the present paper, too: by multiplying the weights in Eq. (45) by the (constant) $(k\varepsilon)^2$, and those in Eq. (46) by the (similarly constant) $(k\varepsilon)^4$, the cases $\varphi \equiv 1$ and $\varphi^* \equiv 1$ are more and more approximated by increasing the value of k. This is equivalent with the use of the L_2-norm. In such a case Eq. (47) yields 0 percent of permissible outliers, i.e. no outliers are permitted in the L_2-norm without an accidental catastrophic deterioration of the results (remember that in the present case the outlier can lie really anywhere). Using the a-variant of P_J or P_J^*, the permissible rate of dispersed outliers is rather high, 30 percent and the efficiency remains at the Gaussian distribution close to the 100 percent (e = 97.5 percent). That means that if in a practical case the Gaussian distribution can be accepted with certainty, even in that case the minimization of the P_J- or of the P_J^*-norm is to be recommended. (It is repeatedly

emphasized in connection with the latter that in the variants using k-times the dihesions the value a = 2 is used in each case, the permissible ratio of the arbitrarily dispersed outliers is 64 percent using the minimization of the P-, P_j-, P^*- and P^*_j-norms.

4.2 Concentrated outliers

It is evident that outliers lying concentrated may modify already in a lower ratio the location of the minimum of the different norms, i.e. the result of the inversion.

In Hajagos's (1988) extremely simplified model the distribution of the data without blunder is characterized by a Dirac- δ in T with a length of the arrow (1-b), where b is the ratio of the outliers. Outliers are now supposed to appear concentrated, i.e. their distribution is characterized by a Dirac- δ of the value b in a point differring from T. It is insignificant from the point of view of the ratio of the permissible outliers, in which point this Dirac- δ is (naturally, this point differs from T), thus let us consider the case when this arrow of the length b is at the point (T+1).

It is easily proven that a ratio of 50 percent outliers (b = 0.5) is not permitted in such a case: in this special case of the symmetrical U-distribution both M_k and M^*_k yield the value (T+0.5) for any value of k.

If b is decreased continuously from 0.5 then till a certain value of b the most frequent value shifts continuously from this position nearer and nearer to T; this shift from the value (T+0.5) is denoted by t, corresponding to a certain value of b. Thus the most frequent value is at a distance from the Dirac- δ in T with the value (1-b) of (0.5-t), from the Dirac- δ at (T+1) with the value b of (0.5+t). These values appear in the following equations which represent the substitution into the integral forms of the condition equation of the location parameter and of the scale parameter (Eqs 2, 4, 16) of this f(x) consisting of two Dirac- δ-s. These equations do not contain integrals due to the known selection characteristics of

the Dirac- δ-s:

$$\frac{0.5-t}{\left[(kS)^2+(0.5-t)^2\right]^q}\cdot(1-b) = \frac{0.5+t}{\left[(kS)^2+(0.5+t)^2\right]^q}\cdot b \qquad (48)$$

$$\frac{(a+1)\cdot(0.5-t)-S^2}{\left[S^2+(0.5-t)^2\right]^2}\cdot(1-b) + \frac{(a+1)\cdot(0.5+t)^2-S^2}{\left[S^2+(0.5+t)^2\right]^2}\cdot b = 0\cdot \qquad (49)$$

These two equations enable the investigation of both forms of the most frequent value in both variants taking into account the following:

a) $q = 1$ at M_k, $q = 2$ at M_k^*;

b) if a is fixed (as an arbitrary value greater than 1 or according to the present convention as 2, 5 or 9), then an S fulfilling Eq. (49) has a factor in Eq. (48) $k = 1$ for the M-estimations and $k = \sqrt{3}$ for the M^*-estimations;

c) if a has the fixed value 2 in Eq. (49), then S fulfilling Eq. (49) is equal with the dihesion ε which may be multiplied with an arbitrary k (according to the conventions in this paper, we prefer $k = 1$, 2 and 3 at M_k and $k = \sqrt{3}$, $2\cdot\sqrt{3}$, and $3\cdot\sqrt{3}$ at M_k^*).

At the beginning of the computations one has to decide about the question which of the estimations M or M_k is to be checked, the values q are thus known. The pair (a,k) is also given, as either a is arbitrary (but greater than 1) and the corresponding values are $k = 1$ and $\sqrt{3}$, respectively, depending on the value of q (1 or 2), - or $a = 2$, accompanied by a value of k greater than 0. The triplet (q, a, k) defines the type of the estimation used, and the value of b defines the asymmetry of the J-distribution (i.e. of the "skew U-distribution") which should be used for testing the estimation. If b is not less than a critical value then Eqs (48) and (49) can be solved for t and S.

Below the mentioned critical b value, the most frequent value is at T, i.e. the computation of the most frequent value is not influenced by the concentrated outliers in a ratio b. From this point of view, a maximum ratio b can be found for any

triplet (q, a, k) at which the concentrated outliers may appear without influencing the result of the computation. It is a secondary question that this ratio called OUT (and given in percent) is determined analytically from Eqs (48) and (49) (as made by Hajagos 1988) or with a program using exclusively computational technics.

The OUT-values are presented for the k-variants of the P- and P*-norms for the full range $k \geq 1$ vs. 1/k in Fig. 6. Figure 7 shows the OUT-curves for the a-variant vs. 1/(a-1) for the

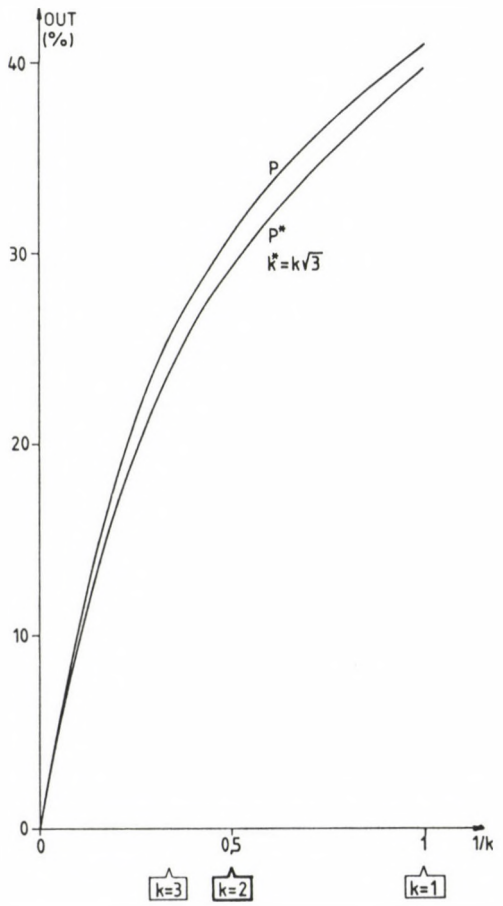


Fig. 6. The maximum permissible ratio of outliers, OUT, for the statistical methods based on the P- and P*-norms (k-variant)

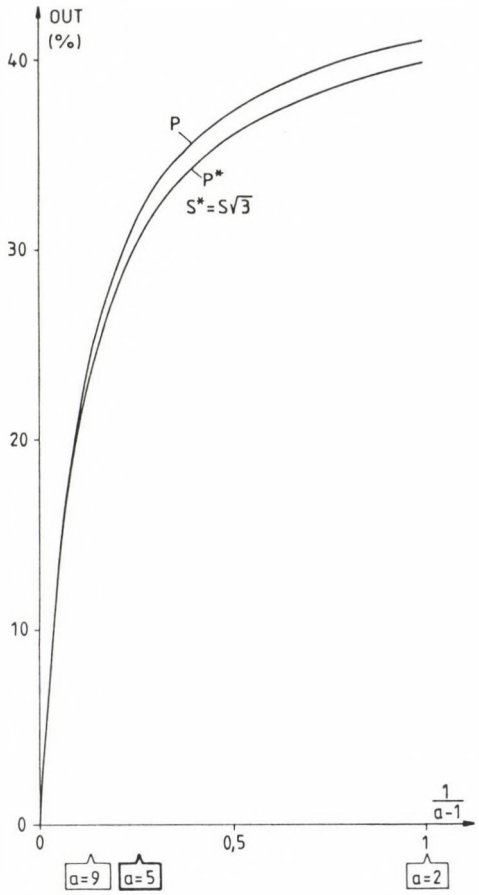


Fig. 7. The maximum permissible ratio of outliers, OUT, for the statistical methods based on the P- and P^*-norms (a-variant)

full range $a \geqslant 2$. (The reader should remember that a defines here the statistical procedure, and not the type of distribution, as e.g. in Figs 4 and 5.)

In spite of the fact that these figures give a complete picture, we give in the following the exact OUT-values for both variants of the P_C, P, P_J and P_C^*, P^*, P_J^* norms, too.

Variant		Percents		
k	P_C: 41.10	P: 31.05	P_J: 25.49	
	P_C^*: 39.83	..		
a	P_C: 41.10			
	P_C^*: 39.83	..		

These numerical values (and Figs 6 and 7) show that

a) there are no great differences between the OUT-values for the k- and a-variants of the P-and P^*-norms (in the case of P_C and P_C^*, there are no differences at all between the variants);

b) the OUT-values for P^* are everywhere less than the corresponding values for P, being in connection with the property of the P^*-norms mentioned in Chapter 3, namely that it maximizes the effective number of points (n_{eff});

c) as the values appearing for the P-norms do not differ significantly from that of the P^*-norms which maximize the value n_{eff}, it seems that the maximization of the effective point number is realized in good approximation by the minimization of the P-norms.

REFERENCES

Hajagos B 1985: Publications of the Technical University for Heavy Industry, Series A, Mining, 40, No. 1-4, 225-238.

Hajagos B 1988: Publications of the Technical University for Heavy Industry, Series A, Mining, 45, No. 1-4, 125-132.

Hajagos B, Steiner F 1991a: Publications of the Technical University for Heavy Industry, Series D, Natural Sciences, 35, No. 4 (in press)

Hajagos B, Steiner F 1991b: Different measures of the uncertainty (present issue)

Steiner F 1985: Robust Estimates (in Hungarian). Lecture notes. Tankönyvkiadó, Budapest

Steiner F 1988a: Publications of the Technical University for Heavy Industry, Series A, Mining, 45, No. 1-4, 133-150.

Steiner F 1988b: Journal Geophysical Transactions, 34, No. 2-3, 139-260.

Steiner F 1988c: Publications of the Technical University for Heavy Industry, Series A, Mining, 45, No. 1-4, 91-100.

Steiner F 1988d: Publications of the Technical University for Heavy Industry, Series A, Mining, 44, No. 4, 241.

Steiner F ed 1990a: The Most Frequent Value. Akadémiai Kiadó, Budapest

Steiner F 1990b: Foundations of Geostatistics (in Hungarian). Tankönyvkiadó, Budapest

Tarantola A 1987: Inverse Problem Theory. Elsevier, Amsterdam

Zilahi-Sebess L 1987: Publications of the Technical University for Heavy Industry, Series A, Mining, 43, No. 1-4, 113-131.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 183-199 (1991)

DIFFERENT MEASURES OF THE UNCERTAINTY

B Hajagos and F Steiner

Geophysical Department, University for Heavy Industry, H-3515 Miskolc, Egyetemváros, Hungary

[Manuscript received March 12, 1990]

The most widely used measure of uncertainty of data sets from measurements is the minimum m value of the L_2-norm, i.e. the standard deviation or scatter (σ). This choice is motivated by the extensive use of the L_2-norm in inversion problems. If the quickly spreading L_1-norm is used, however, the adequate measure of the uncertainty is the mean deviation denoted by d.

In the last years successful experiments have been made for the application of the P-norm in inversions. Most recently, a resistant variant of the P-norm, the so-called P^*-norm has also been defined. The present paper gives the formulas for the uncertainty measures U- and U^* as belonging to both mentioned cases, i.e. to the P- and P^*-norms, further the numerical values are also given for a wide range of error distributions in a supermodel. Comparison of numerical values shows that U and U^* approximate in a wide range of types an interquantile range corresponding to the same probability values as is approximated by σ in a narrow range around the Gaussian distribution.

Keywords: asymptotic variance; interquartile range; intersextile range; inversion; norm; uncertainty

1. INTRODUCTION

The nearly exclusive use of the Gaussian distribution as error distribution type is motivated by the history of science as well as by its computational advantages (see the table with time scale by Steiner 1988). If all error distributions could be accepted as Gaussian ones without outliers, then not only the whole of robust statistics would be superfluous (e.g. Huber 1981) but also Fisher's maximum likelihood principle would not be generally usable: in such a case it would only serve as a possible theoretical basis of traditional statistics.

It is very comfortable to accept in all cases the Gaussian error distribution with the form of the density function

$f_G(x) = (S\cdot\sqrt{2\pi})^{-1} \cdot \exp(-x^2/2S^2)$. In this case the quasiautomatically computed empirical scatter is a good approximation of the scale parameter, S, of the Gaussian distribution. If S (the „scatter") in $f_G(x)$ is known, the length q(p) of all confidence intervals can be easily given (from a table or by another method) to any value of the probability p outside of which, to the left and to the right, errors occur with a probability p. Two data are generally known even without tables or other computation: a) an error is expected outside of the interval (-3S, +3S) with a probability of only 0.27 percent, i.e. the occurrence probability of such errors is practically negligible; b) within the interval (-0.6745 S; +0.6745 S) the probability for the occurrence of an error is the same, namely 0.5 as outside of it.

S yields not only these, e.g. all the values q(p), but any error characteristics is computable from it, thus if the distribution is really a Gaussian one, the scatter characterizes in itself all the error measures. If one expects, however, the occurrence of many types of error distributions, then one has to use a lot of uncertainty measures which are motivated by different points of view and which have certain connections between themselves, and these uncertainty measures should be known for a wide range of error distribution types. This paper has only the moderate task to summarize the properties of some important uncertainty measures together with their interconnections. A final conclusion about the „most realistic" uncertainty measure cannot yet be made and it is even possible that such a conclusion shall never be made.

2. THE INTERQUANTILE RANGES q(p). THE SEMI-INTERQUARTILE RANGE q AND SEMI-INTERSEXTILE RANGE Q. THE UNCERTAINTY U

In the following f(x) denotes the density function of the error distribution, its distribution function is F(x). The most frequently used models of the density functions are simple formulas; in the case of the supermodel $f_a(x)$ e.g. the value of the standard density function is obtained as

$$f_a(x) = \nu(a) \cdot (1+x^2)^{-a/2} \qquad a > 1 \tag{1}$$

for any value of x; in such a case the location parameter, T is zero and the scale parameter has the value 1. In a general case x is to be substituted by (x-T)/S and the expression is to be divided by S. (The norming factor $\nu(a)$ is obtained from

$$\nu(a) = \frac{\Gamma\left(\frac{a}{2}\right)}{\sqrt{\pi} \cdot \Gamma\left(\frac{a-1}{2}\right)} \tag{1a}$$

for a given value of a.) The computation of the distribution functions F(x) is mostly somewhat more complicated, simple formulas for the computation of $F_a(x)$ are only known for the integer values of a (Eq. 88b in Steiner 1988). The inverse of the distribution function, $x = F^{-1}(p)$ is even more complicated, i.e. if one has to know which is the value of x where the distribution function F(x) gives a certain probability p. This value of x is called p-quantile and it is denoted by Q_p, thus

$$F(Q_p) = p \qquad \text{and} \qquad Q_p = F^{-1}(p) \quad \text{resp.} \tag{2}$$

The series of the quantiles gives information about the given distribution. Table I presents such series of values concerning some distributions $f_a(x)$, i.e., for a few values of the type parameter a, including the limiting case $a \to \infty$, the Gaussian-distribution, and naturally the Cauchy-distribution, too, being an element of the supermodel $f_a(x)$ (if a = 2).

The values Q_p of Table I refer to $p \geq 0.5$ as evidently

$$Q_{1-p} = -Q_p \tag{3}$$

holds for all distributions being symmetrical to the origin, i.e. it is sufficient to present the range $p \leq 0.5$ or $p \geq 0.5$ of probabilities, and here the latter has been chosen as in that case the quantiles Q_p are positive. But the selection of the probability range $p \leq 0.5$ is more advantageous also from another, more important point of view: these Q_p-values are

Table I

Values of $x = F_a^{-1}(p) = Q_p$ belonging to some probability values $p(\geqslant 0.5)$
(The values x refer to standard density models $f_a(x)$ for a = 2; 2.2; 2.4; 2.8; ... ; 100. The limit case $a \rightarrow \infty$ i.e. the Gaussian distribution is also the standard one.)

p \ a=	2	2.2	2.4	2.8	3	3.2	4	5	6	7	8	10	20	40	100	
	Cauchy									Jeffreys intervall						Gauss
0.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.55	0.1583	0.1397	0.1260	0.1073	0.1005	0.0948	0.7886	0.0669	0.0591	0.0535	0.0492	0.0430	0.0292	0.0202	0.0126	0.1256
0 60	0.3249	0.2855	0.2571	0.2181	0.2041	0.1924	0.1597	0.1353	0.1194	0.1081	0.0994	0.0869	0.0589	0.0408	0.0255	0.2533
0.65	0.5095	0.4449	0.3988	0.3366	0.3144	0.2960	0.2449	0.2070	0.1825	0.1650	0.1517	0.1326	0.0897	0.0621	0.0388	0.3853
0.70	0.7265	0.6282	0.5595	0.4684	0.4364	0.4099	0.3374	0.2843	0.2501	0.2259	0.2075	0.1811	0.1223	0.0846	0.0528	0.5244
0.75	q=1.0000	q=0.8522	q=0.7517	q=0.6221	q=0.5773	q=0.5407	q=0.4416	q=0.3703	q=0.3249	q=0.2929	q=0.2687	q=0.2342	q=0.1577	q=0.1090	q=0.0680	q=0.6744
0.80	1.3763	1.1484	0.9989	0.8126	0.7500	0.6993	0.5649	0.4704	0.4112	0.3697	0.3386	0.2944	0.1975	0.1362	0.0849	0.8416
0.83	Q=1.7320	Q=1.4169	Q=1.2166	Q=0.9741	Q=0.8944	Q=0.8306	Q=0.6642	Q=0.5496	Q=0.4787	Q=0.4294	Q=0.3927	Q=0.3407	Q=0.2277	Q=0.1568	Q=0.0977	Q=0.9674
0.85	1.9626	1.5860	1.3508	1.0710	0.9802	0.9079	0.7215	0.5947	0.5168	0.4630	0.4230	0.3665	0.2444	0.1682	0.1047	1.0364
0.90	3.0776	2.3613	1.9429	1.4769	1.3333	1.2219	0.9455	0.7666	0.6600	0.5877	0.5347	0.4610	0.3046	0.2087	0.1296	1.2815
0.95	6.3137	4.3780	3.3646	2.3506	2.0647	1.8518	1.3587	1.0659	0.9011	0.7933	0.7160	0.6110	0.3966	0.2698	0.1668	1.6448
0.975	12.706	7.8955	5.6261	3.5738	3.0424	2.6622	1.8373	1.3882	1.1496	0.9989	0.8937	0.7540	0.4801	0.3238	0.1994	1.9599
0.99	31.820	17.014	10.913	6.0532	4.9246	4.1565	2.6215	1.8734	1.5048	1.2829	1.1331	0.9404	0.5825	0.3884	0.2376	2.3263
0.995	63.656	30.344	17.937	8.9445	7.0179	5.7544	3.3722	2.3020	1.8032	1.5135	1.3226	1.0832	0.6563	0.4336	0.2639	2.5758
0.9975	127.32	54.075	29.456	13.180	9.9624	7.9274	4.3031	2.7987	2.1347	1.7623	1.5229	1.2298	0.7281	0.4764	0.2885	2.8070
0.999	318.30	116.06	56.714	21.957	15.787	12.061	5.8973	3.5865	2.6356	2.1260	1.8086	1.4322	0.8211	0.5304	0.3190	3.0902
0.9995	636.61	206.83	93.066	32.256	22.343	16.553	7.4616	4.3051	3.0718	2.4326	2.0439	1.5936	0.8909	0.5697	0.3408	3.2905

identical with the semi-interquantile ranges (see below).

When for the characterization of the fluctuation of the x-values a range q(p) is to be given, where the x-values occur with a probability of (1-2p) and x-values less than the lower limit of the range and more than the upper limit, have the same probability p, obviously the corresponding formula is to be used

$$q(p) = Q_{1-p} - Q_p \ . \tag{4}$$

If $p = 1/4$, $Q_p = Q_{1/4}$ is the lower quartile, $Q_{1-p} = Q_{3/4}$ is the upper quartile, and the length of the interval. $q(0.25) = Q_{3/4} - Q_{1/4}$ is called interquartile range; if $p = 1/6$, then $Q_p = Q_{1/6}$ is the lower, $Q_{5/6}$ is the upper sextile, the difference $Q(1/6) = Q_{5/6} - Q_{1/6}$ is the intersextile range.

In the case of symmetrical distribution it is obvious to use the half of the range q(p) as a measure of uncertainty, but its use can also be suggested for asymmetrical distributions. Namely, by adding and subtracting this value from the correct value (from T), below and above this interval the values of the random variable occur with the same probability p. The semi-interquartile range belonging to $p = 1/4$ and denoted by q, is:

$$q = q(0.25)/2 = [Q_{3/4} - Q_{1/4}]/2 \ . \tag{5}$$

Bessel proposed the semi-interquartile range in 1815 as a measure of uncertainty; if T is the symmetry point of a symmetrical distribution, the probability for the occurrence of values within the interval (T-q; T+q) is the same as outside of it. If a more rigorous measure is wanted, then the occurrence ratio of 2:1 is to be used. In this case the semi-intersextile range

$$Q = q(1/6)/2 = [Q_{5/6} - Q_{1/6}]/2 \tag{6}$$

is the measure of uncertainty: the occurrence of values within the interval (T-Q; T+Q) is really twice as much probable as outside of it. Table I shows that this quantity is very near to the standard deviation in the case of a Gaussian distribution.

Thus in the case of a Gaussian distribution, the sample average plus and minus the scatter σ gives an interval within which about twice as much values occur as outside of it.

If any value of p or any ratio of occurrence probabilities is chosen as appropriate measure of the uncertainty, be it either q (the semi-interquartile range), or Q (the semi-intersextile range), or any semi-interquantile range belonging to a value $p \geq 0.5$ (for error distributions symmetrical to the origin, thus also for the $f_a(x)$ distribution where the values in Table I are valid, the upper quantile equals with the semi-interquantile range), — one may doubt, however, about the correctness of the selection of a single value of p and of a single semi-interquantile range, respectively, for this purpose. It would be namely better to take several q(p)-s weighted with an appropriate weight function into account. Steiner (1985) proposed to use the following integral which yields in all cases a finite value in contrast to weighting functions which do not contain q(p):

$$\bar{U} = 48q\,(0.25) \int_0^{0.5} q(p)\, \frac{(0.5-p)^2}{q(p) + q(0.25)}\, dp \,. \qquad (7)$$

Steiner (1985) presented examples for a given q for the increase of $\bar{U}$ with increasing flanks of the distributions, — at the same time Eq. (7) does not overestimate the role of the flanks: even in the case of a distribution with infinite entropy $\bar{U}$ is less than twice q. The entropy, as uncertainty measure is not dealt with here as its value may be infinite for certain types of distributions. The same refers to the average deviation d, the scatter σ is even less applicable. Nevertheless, the theory of statistics is so full of these quantities that they shall be treated in the next chapter.

3. UNCERTAINTY MEASURES DEFINED AS MINIMUM NORMS

The solution of the inversion problem in geophysics means to fit the measured values x_i to an a priori analytically known

function $T(\bar{m}; \bar{y}_i)$ by minimizing the deviations

$$X_i = x_i - T(\bar{m}; \bar{y}_i) \tag{8}$$

between these values according to a chosen norm, more exactly it means the determination of a set of parameters $\bar{m} = m_o, m_1 .. m_j, \ldots m_{jo}$ where the corresponding norm is a minimum (Tarantola 1987). As norm most frequently the expression

$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} X_i^2} \tag{9}$$

called L_2-norm is used, but the L_1-norm being

$$\frac{1}{n} \sum_{i=1}^{n} \left| X_i \right| \tag{10}$$

comes more and more into use due to its robustness.

Further norms also exist as the expression

$$\left[\frac{1}{n} \sum_{i=1}^{n} \left| X_i \right|^p \right]^{1/p} \tag{11}$$

of the L_p-norm is valid not only for p = 1 and 2 (giving the L_1 and L_2-norms) but also for any positive number (Tarantola 1985), and therefore Eq. (11) defines an infinite number of norms. In the practice, however, values other than 1 or 2 are hardly used for p. This is partly due to computational advantages, partly due to the fact that the elements of the supermodel defined by the density function

$$f_p(x) = \frac{p^{1-1/p}}{2\,\Gamma(1/p)}\, e^{-|x|^p/p} \qquad (p > 0) \tag{12}$$

are in the practice quite seldom adequate models of error distributions. As e.g. Tarantola (1985) has shown the L_p norm minimization algorithm is an optimum statistical algorithm in the case of $f_p(x)$-type error distributions.

As the elements of the supermodel $f_a(x)$ (Eq. 1) prove quite often to be acceptable models for the distributions in statistical problems, the P- and P*-norms - being optimum ones for the $f_a(x)$ distributions - should be taken seriously into account (for definitions see Hajagos and Steiner 1991). According to these, P_C- and P_C^* can be used with the highest statistical efficiency in the vicinity of the Cauchy- (a=2), P_J and P_J^* in the vicinity Jeffrey's type (a=9), while the P and P*-norm are optimum ones in the vicinity of the so-called geostatistical distribution (a=5) and have an efficiency of over 90 percent in the complete Gaussian-Cauchy-range. (As this case is considered as a standard one for both series of norms, no index is used here.)

The resulting model parameters differ when different norms are used, and these vectors are denoted by $\bar{m}(L_1)$, $\bar{m}(L_2)$, $\bar{m}(P_C)$, $\bar{m}(P)$, $\bar{m}(P_J)$, $\bar{m}(P_C^*)$, $\bar{m}(P^*)$, $\bar{m}(P_J^*)$. Using anyone of these $\bar{m}(.)$-s, a different set of deviations is obtained from Eq. (8), and therefore these deviations are denoted per analogiam by $X(L_1)$, $X(L_2)$, $X(P_C)$, ... $X(P_J^*)$ for the sets of deviations obtained with the corresponding m(.)-s.

These sets of deviations are those which minimize the norms L_1, ... P_J^*, and just these minimum norms can be cosidered as measures of uncertainty from the point of view of the corresponding norm. (Some heuristic trains of thought are given in Steiner 1990). Thus for the L_2-norm, i.e. in the conventional statistical methods, the expression

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2(L_2)} \qquad (13)$$

i.e. the scatter, σ, is used to characterize the uncertainty, while in the case of the L_1-norm the average deviation, d is this measure:

$$\frac{1}{n}\sum_{i=1}^{n}\left|X_i(L_1)\right| \quad . \tag{14}$$

For the P- and P*-norms the uncertainties are denoted by U_C, U, U_J, U_C^*, U^*, U_J^*. The formulas for the uncertainty measures were given in Table II both in form of sums and in integral forms. These formulas are naturally the same as the index of the X_i-s hints at that special set of deviations which gives the minimum value of the norm in question. (ε is the dihesion of the actual set of deviations X_i, see e.g. Hajagos and Steiner 1991.) Of course, every special set of deviations is automatically given simultaneously with the results of the inversion made by the corresponding norm.

Figure 1 informs about the different uncertainties for the supermodel $f_a(x)$. Here the quantile values are given at which the uncertainty measures of the two standard cases, U and U*, as well as σ and d are identical. The precise numerical values for all the uncertainty measures studied, i.e. for the three variants of U and U*, for σ and for d, are given in Table III, where - for completeness - the values of the dihesion are also given.

The curves in Fig. 1 show that U is a good approximation of the semi-intersextile-range, i.e. the absolute values of the deviations are with a probability of 2/3 less than this value. A similar conclusion refers to U*, too, but here the deviations from Q are in certain type ranges significantly greater than the deviations of $\bar{U}$. The standard deviation σ is only in the immediate vicinity of the Gaussian distribution practically identical with Q, and the same refers to d in the vicinity of the geostatistical distribution. (The weighted average of the interquantile ranges, $\bar{U}$, according to Eq. (7) is also plotted in Fig. 1, and it is everywhere nearer to Q, then to the semi-interquartile range, q.)

The conclusions from Fig. 1 suggest that the uncertainty measures should be presented for different types of distributions by their ratio to Q, i.e. to the semi-intersextile-range. The curve U/Q, but even the curve U*/Q confirm

Table II

$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2(L_2)}$	$\sigma = \sqrt{\int_{-\infty}^{\infty} X^2(L_2)\, f(X)dX}$
$d = \frac{1}{n}\sum_{i=1}^{n} \lvert X_i(L_1)\rvert$	$d = \int_{-\infty}^{\infty} \lvert X(L_1)\rvert\, f(X)dX$
$U_C = \varepsilon \left\{ \prod_{i=1}^{n} \left[1 + \left(\frac{X_i(P_C)}{\varepsilon} \right)^2 \right] \right\}^{\frac{1}{2n}}$	$U_C = \varepsilon \exp\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln\left[1 + \left(\frac{X(P_C)}{\varepsilon} \right)^2 \right] f(X)dX \right\}$
$U = \varepsilon \left\{ \prod_{i=1}^{n} \left[1 + \left(\frac{X_i(P)}{2\varepsilon} \right)^2 \right] \right\}^{\frac{1}{2n}}$	$U = \varepsilon \exp\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln\left[1 + \left(\frac{X(P)}{2\varepsilon} \right)^2 \right] f(X)dX \right\}$
$U_J = \varepsilon \left\{ \prod_{i=1}^{n} \left[1 + \left(\frac{X_i(P_F)}{3\varepsilon} \right)^2 \right] \right\}^{\frac{1}{2n}}$	$U_J = \varepsilon \exp\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln\left[1 + \left(\frac{X(P_J)}{3\varepsilon} \right)^2 \right] f(X)dX \right\}$
$U_C^* = \frac{4\varepsilon}{n} \sum_{i=1}^{n} \frac{X_i^2(P_C^*)}{3\varepsilon^2 + X_i^2(P_C^*)}$	$U_C^* = 4\varepsilon \int_{-\infty}^{\infty} \frac{X^2(P_C^*)}{3\varepsilon^2 + X^2(P_C^*)} f(X)dX$
$U^* = \frac{10\varepsilon}{n} \sum_{i=1}^{n} \frac{X_i^2(P^*)}{12\varepsilon^2 + X_i^2(P^*)}$	$U^* = 10\varepsilon \int_{-\infty}^{\infty} \frac{X^2(P^*)}{12\varepsilon^2 + X^2(P^*)} f(X)dX$
$U_J^* = \frac{20\varepsilon}{n} \sum_{i=1}^{n} \frac{X_i^2(P_J^*)}{27\varepsilon^2 + X_i^2(P_J^*)}$	$U_J^* = 20\varepsilon \int_{-\infty}^{\infty} \frac{X^2(P_J^*)}{27\varepsilon^2 + X^2(P_J^*)} f(X)dX$

Table III

a=	1.8	2	2.2	2.4	3	4	5	10	40	100
						1.0000	0.7071	0.3780	0.1644	0.1015
d			3.6043	1.9954	1.0000	0.6366	0.5000	0.2910	0.1303	0.0808
U_C	2.5939	2.0000	1.6638	1.4463	1.0880	0.8231	0.6875	0.4327	0.2009	0.1253
U	1.8995	1.5000	1.2693	1.1174	0.8605	0.6628	0.5585	0.3563	0.1667	0.1041
U_J	1.6616	1.3333	1.1411	1.0128	0.7916	0.6166	0.5223	0.3358	0.1577	0.0986
U_C^*	1.7983	1.4641	1.2389	1.0783	0.7936	0.5772	0.4698	0.2812	0.1264	0.0784
U_J^*	2.9507	2.2401	1.7877	1.4809	0.9806	0.6473	0.5005	0.2759	0.1185	0.0730
U^*	4.4772	3.2278	2.4635	1.9639	1.1961	0.7317	0.5444	0.2837	0.1188	0.0728
	1.1155	1.0000	0.9117	0.8418	0.6993	0.5616	0.4818	0.3148	0.1489	0.0932

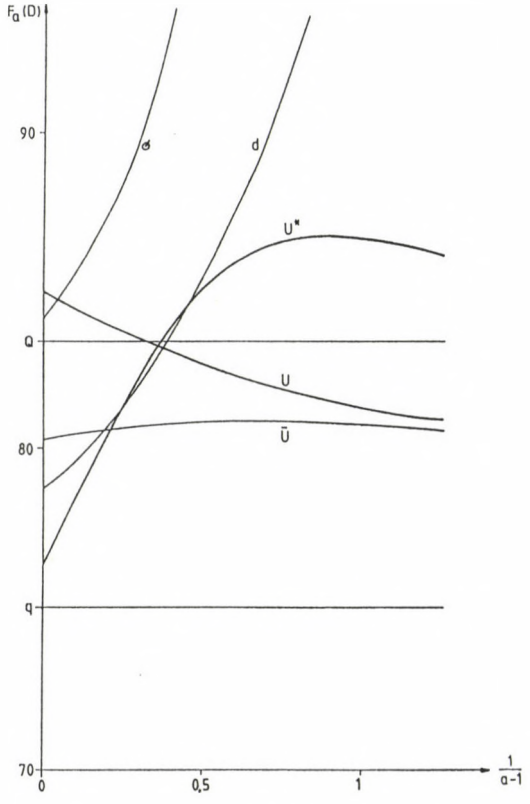


Fig. 1. Minima of different norms, as measures of uncertainty; they approximate more or less the semi-intersextile range Q

that U (and U^*, too) can be considered as a value approximating Q, i.e. they can be interpreted as semi-intersextile ranges in the complete Gaussian-Cauchy-type interval, similarly as the scatter (standard deviation) is interpreted as such but only in the immediate vicinity of the Gaussian distribution. Figure 2 shows that the curve σ/Q is steeply increasing till $1/(a-1) = 0.5$ later at $(a \leqslant 3)\sigma$ is infinite, i.e. it cannot be used as an error characteristics. The value d only gets unapplicable from a = 2 on.

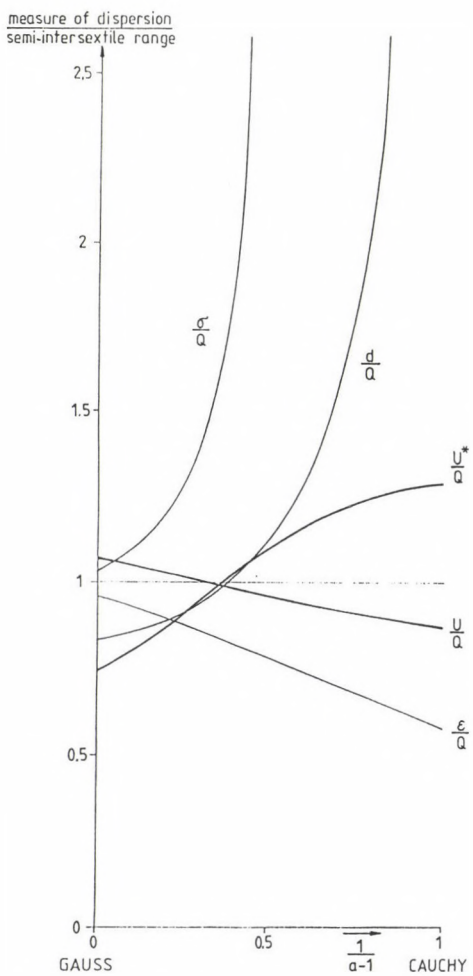


Fig. 2. The norm minimum/Q ratio vs. distribution type

Figure 3 shows the curves U_C/Q, U/Q and U_J/Q, Fig. 4 the corresponding curves denoted by U_C^*/Q, U^*/Q and U_J^*/Q. There are places where these ratios differ significantly from 1, but in the majority these values are near to 1. The actual judgement about the significances of these deviation also depends on the accuracy of the determination of the uncertainty measures; the error of errors is a very wide field which shall not be treated in the present paper. It is to be mentioned, however, that independently of the deviation from Q of an uncertainty measure for a distribution $f_a(x)$ with the type parameter a, a comparison of the data in Tables III and I gives all necessary factors to be multiplied with the uncertainty measure obtained

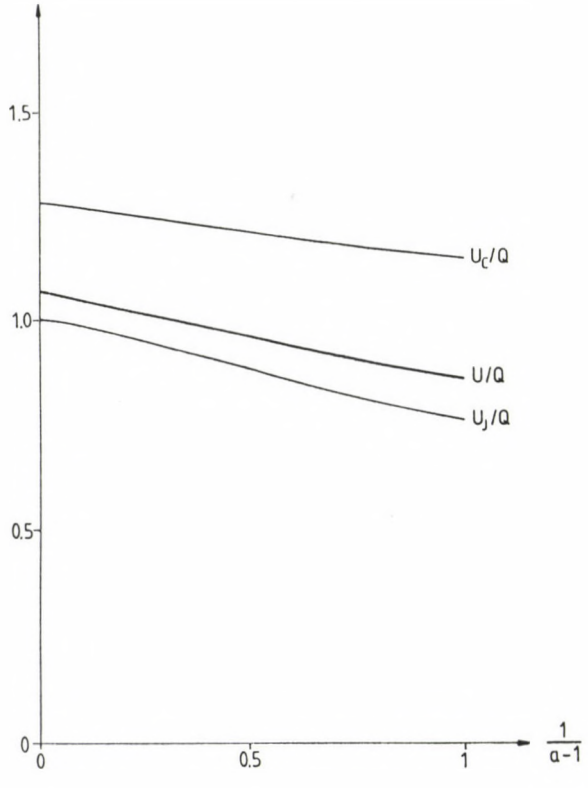


Fig. 3. The dependence of the three kinds of U as referred to Q, on the type of the distribution

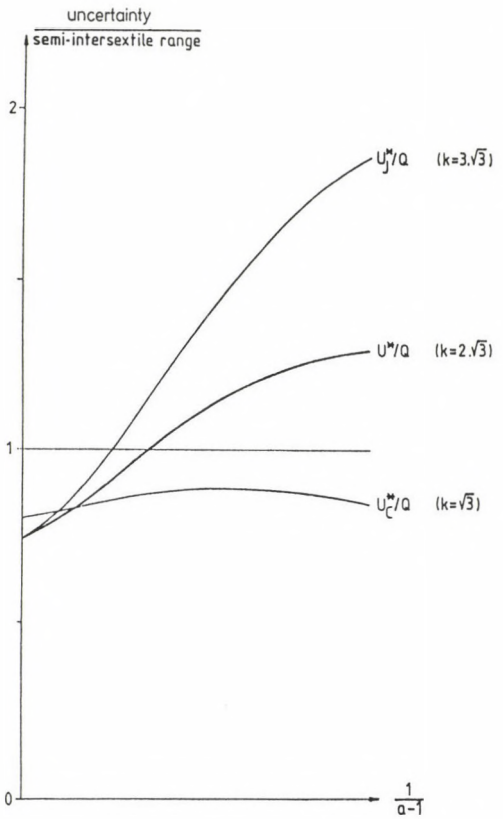


Fig. 4. The dependence of the three kinds of U^* as referred to Q, on the type of distribution

as minimum norm for obtaining any semi-interquantile range (in Table I) and therefore one gets a complete picture on the error distribution. For this purpose, sufficiently precise values of σ, d, U_C, U, ... are needed what necessitates not only a great number of values, but sufficiently resistant error characteristics, too. Such a sufficiently large resistance is characteristic only for the uncertainty measures U_C^*, U^*, and U_J^*, if outliers can occur really anywhere.

4. ABOUT ASYMPTOTIC SCATTERS

If the task is to determine just the most characteristic

values T (e.g. the symmetry point), it is often necessary to estimate the error of this estimate (standard deviation) (the latter is obtained after all as the location of the minimum norm). As the distribution of the estimates approximates at great values of n under loose conditions the Gaussian distribution (Huber 1981), it is sufficient to give the scatter (standard deviation) of the estimates to be computed as $A/\sqrt{n}$ if the asymptotic scatter A is known.

The estimates of T are denoted by M_C, M and M_J, if the corresponding P-norms are used, and by the same notations with stars, if the P^* norms are used. Essentially all the six quantities should be considered as most frequent values.

The formulas for the asymptotic variances are given by Hajagos and Steiner (1991). It is advantageous to denote with star these asymptotic scatters A or to use them without star, and to use the same indices C and J (or leave them unindexed) as it is used for the corresponding most frequent value.

The following table contains not only the asymptotic scatters for these six kinds of most frequent values for different values of the type parameter a, but these of the average, A_E, too (but it is identical with σ, as well known), and the asymptotic scatter of d, denoted by A_{med} which is for unimodal distributions being symmetrical at T:

$$A_{med} = \frac{1}{2f(T)} \qquad (15)$$

(see e.g. Steiner 1988). The quotients A/U and A^*/U^* are also included into this table; the latter curves are plotted in Fig. 5, too.

The curve A/U in Fig. 5 and the last but one row of Table IV shows that A/U = 1 is fulfilled in high approximation. It is analogous to the equality $\sigma = A_E$. The curve A^*/U^* is also approximating in average 1, but significantly less closely. As the thin curve in Fig. 5 shows, the ratio A_{med}/d has values between 1.6 and zero in the Gaussian-Cauchy-range (an exact identity is valid at a = 3), thus it cannot be stated for the complete range that A_{med} and d could be substituted by each other, not even if a great error is permitted.

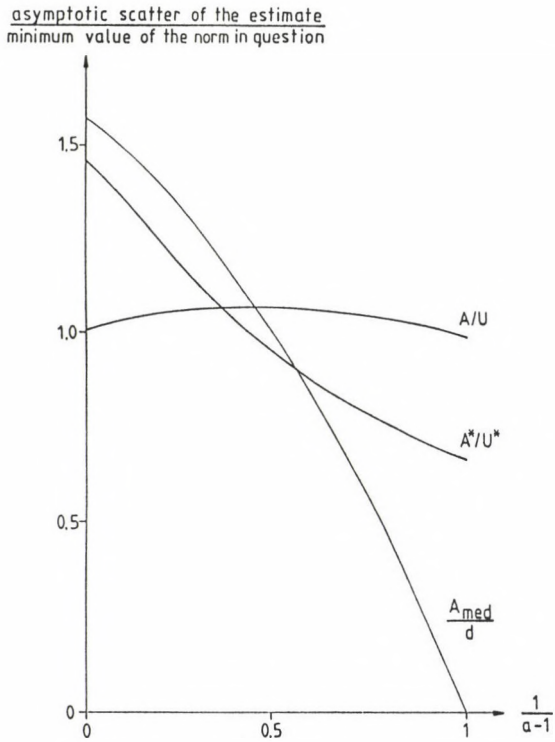


Fig. 5. The connection of the asymptotic scatters and corresponding norm minima vs. distribution type

The comfort being usual for A_E and G (as they are identical for the first half of the Gaussian-Cauchy-range, i.e. if these values are not infinite ones) is not valid for the location of the minimum of the L_1-norm, i.e. for the value d and the asymptotic scatter of the median. We have seen that the values A^* and U^* are near to each other (have approximately the same value), and A and U can be taken as practically equal. In the evaluation of the latter statement the errors of the estimates of these values from given samples are also to be taken into account, as these uncertainties are often uncertain

Table IV

a=	2	2.5	3	4	5	10	40	100	
A_E				1.0000	0.7071	0.3780	0.1644	0.1015	1.0000
A_{med}	1.5708	1.1982	1.0000	0.7854	0.6667	0.4295	0.2020	0.1263	1.2533
A_C	1.4142	1.1011	0.9261	0.7311	0.6214	0.4002	0.1818	0.1174	1.1646
A	1.5000	1.1251	0.9236	0.7080	0.5917	0.3694	0.1699	0.1057	1.046
A_J	1.6330	1.1941	0.9636	0.7234	0.5974	0.3652	0.1658	0.1030	1.0177
A^*	1.4347	1.1241	0.9492	0.7526	0.6411	0.4144	0.1948	0.1218	1.2085
A^*	1.4976	1.1261	0.9253	0.7094	0.5925	0.3690	0.1692	0.1053	1.0413
A^*	1.6375	1.2008	0.9697	0.7275	0.6000	0.3652	0.1652	0.1025	1.0127
A/U	1.000	1.063	1.073	1.068	1.059	1.037	1.019	1.015	1.013
A^*/U^*	0.669	0.826	0.944	1.096	1.184	1.337	1.428	1.442	1.452

to several tens of percents. The deviations of the values A^* and U^* in Fig. 5 have as maximum the just mentioned value while the deviations of A and U are not more than a few percent.

REFERENCES

Hajagos B, Steiner F 1991: The P^*-norma (present issue)

Huber P J 1981: Robust Statistics. John Wiley and Sons, New York

Steiner F 1985: Publ. of the Techn. Univ. for Heavy Industry, Miskolc, Series A, Mining, 38, 139-144.

Steiner F 1988: Geophys. Transactions, 34, 139-260.

Steiner F 1990: Introduction to Geostatistics (in Hungarian). Tankönyvkiadó, Budapest (in press)

Tarantola A 1987: Inverse Problem Theory. Elsevier, Amsterdam

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 201-219 (1991)

RELATIVE MOTIONS BETWEEN THE INNER CORE AND THE EARTH'S MANTLE DUE TO ATTRACTIONS OF THE SUN AND THE MOON

H Jochmann

Zentralinstitut für Physik der Erde, 1500-Potsdam, Telegrafenberg A 17, GDR

[Manuscript received March 22, 1990]

The response of the inner core motion with respect to the mantle due to the attraction of the Sun and the Moon is studied. It could be shown that some periodic constituents of polar motion and the dipole part of the geomagnetic westward drift, which correspond with astronomical nutation periods, are in accordance with the applied hypothesis of a relative precession and nutation of the inner core.

Keywords: geomagnetic secular variation; inner core motion; polar motion

Recently, Smylie et al. (1984) discussed the motion of the inner core and mantle coupled via mutual gravitation. The basic idea of their investigations was the hypothesis that the symmetry axis of the inner core is inclined to that of the mantle. The above mentioned authors showed that this model is capable to produce a retrograde precession relative to the mantle. The amounts of the precession and the inclination are given by the westward drift and the inclination of the geomagnetic dipole. Some authors (Bullard 1949, Schmutzer 1978) showed that the direction of the observable dipole field - which is a result of the induced dipole field of the inner core and the dipole part of the fluid outer core - coincides with the symmetry axis of the inner core. Smylie et al. (1984) proved that the gravitational torque exerted by the mantle and the fluid outer core can explain this precessional motion, because this torque is larger than viscous or electromagnetic torques influencing the motion of the inner core. They rigorously calculated the torques exerted by the mantle and the outer core, and assuming that the westward drift of the geomagnetic dipole represents

the precessional motion, the rotational velocity of the inner core was calculated, supposing regular precession.

The hypothesis of inner core motion, as indicated by the variation of the geomagnetic dipole's westward drift, was confirmed by calculating the polar motion component which is excited by inner core motion (see Jochmann 1989).

In the following, the attraction of the Sun and the Moon within a model consisting of an inner core and mantle will be discussed. Some periodic variations of the dipole's westward drift and polar motion will be explained by these considerations.

1. THE EQUATIONS OF MOTION

The motions of the inner core and the mantle are described by the motions of an inner core fixed co-ordinate system ($\underline{i}_1$, $\underline{i}_2$, $\underline{i}_3$) and a mantle fixed co-ordinate system ($\underline{e}_1$, $\underline{e}_2$, $\underline{e}_3$) with respect to an inertial frame of reference ($\underline{E}_1$, $\underline{E}_2$, $\underline{E}_3$) attached to the plane of the ecliptic. $\underline{E}_3$ is perpendicular to this plane.

The inertia frame is chosen in the above mentioned way, because it allows the connection to the orbital motions of the Earth and the Moon. The motions of the above mentioned co-ordinate systems are described by Euler's angles (see Fig. 1 and 2).

$\underline{e}'$ and $\underline{i}'$ are the unit vectors along the nodal lines between the mantle and the inertial frame and the inner core and the mantle frame.

The motions of the inner core and the mantle are governed by the equations

$$\begin{aligned} \frac{d\underline{H}_m}{dt} + (\underline{\omega}_1 \times \underline{H}_m) &= \underline{L}_m + \underline{\Gamma} \\ \frac{d\underline{H}_i}{dt} + (\underline{\omega}_2 \times \underline{H}_i) &= \underline{L}_i - \underline{\Gamma} \end{aligned} \qquad (1)$$

where $\underline{H}_m$ and $\underline{H}_i$ are the angular momenta of the mantle and the inner core. ω_1 and ω_2 are the angular velocities of the

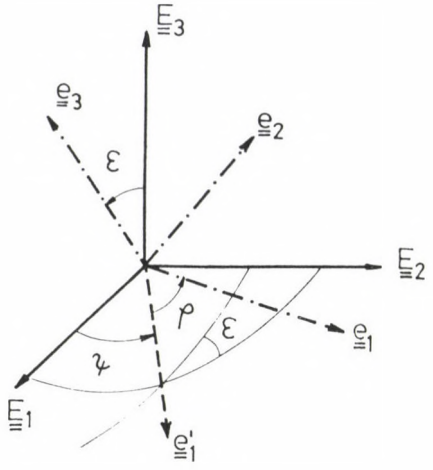


Fig. 1. The mantle fixed and the inertial coordinate systems

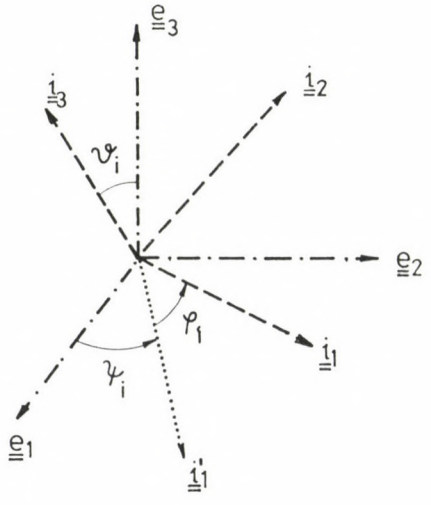


Fig. 2. The inner-core fixed and the mantle fixed co-ordinate systems

co-ordinate systems where the equations (1) are given. $\underline{\Gamma}$ is the torque caused by mutual gravitation between the inner core and the mantle. $\underline{L}_m$ and $\underline{L}_i$ are the torques acting on the mantle and the inner core which are caused by the attractions of the

Sun and the Moon.

According to Smylie et al. (1984) it is convenient to choose co-ordinate systems for the construction of equations (1) with following unit vectors along their axes

$\underline{i}_1'$, $\underline{e}_3 \times \underline{i}_1'$, $\underline{e}_3$ for the first equation (1)

$\underline{i}_1'$, $\underline{i}_3' \times \underline{i}_1'$, $\underline{i}_3$ for the second equation (1).

The angular velocities of these co-ordinate systems are

$$\begin{aligned} \underline{\omega}_1 &= \underline{\omega}_m + \dot{\psi}_i \, \underline{e}_3 \\ \underline{\omega}_2 &= \underline{\omega}_s - \dot{\varphi}_i \, \underline{i}_3 \end{aligned} \tag{2}$$

where $\underline{\omega}_m$ is the spatial angular velocity of the mantle and $\underline{\omega}_s$ is that of the inner core.

According to Fig. 2 the relative rotational velocity of the inner core with respect to the mantle can be written in the notation

$$\underline{\omega}_i = \dot{\psi}_i \, \underline{e}_3 + \dot{\vartheta}_i \, \underline{i}_1' + \dot{\varphi}_i \, \underline{i}_3 \, . \tag{3}$$

Introducing spatial rotational velocity components $\dot{\psi}_s$, $\dot{\vartheta}_s$ and $\dot{\varphi}_s$ along the same vectors as in equation (3) we obtain an expression for the spatial rotational velocity of the inner core

$$\underline{\omega}_s = \dot{\psi}_s \, \underline{e}_3 + \dot{\vartheta}_s \, \underline{i}_1' + \dot{\varphi}_s \, \underline{i}_3 \, . \tag{4}$$

According to Eqs (3) and (4) the spatial angular velocity of the mantle is given by

$$\begin{aligned} \underline{\omega}_m = \underline{\omega}_s - \underline{\omega}_i = (\dot{\psi}_s - \dot{\psi}_i) \, \underline{e}_3 + (\dot{\vartheta}_s - \dot{\vartheta}_i) \, \underline{i}_1' + \\ + (\dot{\varphi}_s - \dot{\varphi}_i) \, \underline{i}_3 \, . \end{aligned} \tag{5}$$

Using the Euler angles between the mantle and the inertial frame, the spatial angular velocity of the mantle reads

$$\underline{\omega}_m = \dot{\psi}\underline{E}_3 + \dot{\varepsilon}\underline{e}' + \dot{\varphi}e_3 \,. \tag{6}$$

Using the tensors of inertia of the mantle $\underline{I}_m$ and the inner core $\underline{I}_i$ the angular momenta of both components of the Earth are obtained according to

$$\underline{H}_m = \underline{I}_m \, \underline{\omega}_m = \underline{I}_m \, (\underline{\omega}_s - \underline{\omega}_i)$$

and (7)

$$\underline{H}_i = \underline{I}_i \, \underline{\omega}_s \,.$$

Introducing (7) in (1) and using the expressions (4) and (5), we obtain after some transformations between different coordinate systems the following differential equations

$$\frac{d}{dt} (\dot{\vartheta}_s - \dot{\vartheta}_i) - (\sigma - \dot{\psi}_i) (\dot{\varphi}_s - \dot{\varphi}_i) \sin \vartheta_i = \frac{1}{A} (\underline{L}_m \underline{i}'_1 + \underline{\Gamma} \underline{i}'_1)$$

$$\frac{d}{dt} (\dot{\varphi}_s - \dot{\varphi}_i) \sin \vartheta_i + (\sigma - \psi_i) (\dot{\vartheta}_s - \dot{\vartheta}_i) = -\frac{1}{A} \underline{L}_m (\underline{e}_3 \times \underline{i}'_1) \tag{8}$$

$$\dot{\psi}_s - \dot{\psi}_i + (\dot{\varphi}_s - \dot{\varphi}_i) \cos \vartheta_i = \omega_o \,.$$

$$\frac{d}{dt} \dot{\vartheta}_s + (\sigma_i + \dot{\varphi}_i) \dot{\psi}_s \sin \vartheta_i = \frac{1}{A_i} (\underline{L}_i \underline{i}'_1 - \underline{\Gamma} i'_1)$$

$$\frac{d}{dt} \dot{\psi}_s \sin \vartheta_i - (\sigma_i + \dot{\varphi}_i) \dot{\vartheta}_s = \frac{1}{A_i} \underline{L}_i \, (\underline{i}_3 \times \underline{i}'_1) \tag{9}$$

$$\dot{\varphi}_s + \dot{\psi}_s \cos \vartheta_i = \omega_{oi} \,.$$

In (8) and (9) are

$$\sigma = \frac{C-A}{A} \, \omega_o \qquad \text{and} \qquad \sigma_i = \frac{C_i - A_i}{A_i} \, \omega_{oi}$$

with the principal moments of inertia C, A, C_i, A_i of the mantle and the inner core.

Comparing (5) and (6) we obtain the relation between the angular velocity components $\dot{\varphi}$, $\dot{\psi}$, $\dot{\vartheta}$, and those contained in the differential equations (8) and (9)

$$
\begin{aligned}
\dot{\vartheta}_s &= \dot{\vartheta}_i + \dot{\varepsilon}\cos(\varphi+\psi_i) + \dot{\psi}\sin\varepsilon \quad \sin(\varphi+\psi_i) \\
\dot{\varphi}_s &= \dot{\varphi}_i + \dot{\varepsilon}\,\frac{\sin(\varphi+\psi_i)}{\sin\vartheta_i} - \dot{\psi}\cos(\varphi+\psi_i)\,\frac{\sin\varepsilon}{\sin\vartheta_i} \\
\dot{\psi}_s &= \dot{\psi}_i + \dot{\varphi} + \dot{\psi}(\cos\varepsilon + \sin\varepsilon\cos(\varphi+\psi_i)\cot\vartheta_i) - \\
&\quad - \dot{\varepsilon}\sin(\varphi+\psi_i)\cot\vartheta_i
\end{aligned}
\tag{10}
$$

and

$$\dot{\varphi} + \dot{\psi}\cos\varepsilon = \omega_o \,.$$

Equations (8), (9) and (10) enable us to determine the relative motion between inner core and mantle, and the motion of the mantle with respect to the inertial co-ordinate system, provided that we can obtain suitable expressions for the torques $\underline{\underline{L}}_m$, $\underline{\underline{L}}_i$ and $\underline{\underline{\Gamma}}$.

2. THE TORQUES INFLUENCING MOTIONS OF THE PRESUMED EARTH MODEL

The components of the presumed Earth model are subject to external and relative torques. Relative torques are caused by electromagnetic forces, which act on the inner core because of its high electrical conductivity and the strong magnetic field in the outer core. The relative motion of the inner core to the outer core will be resisted by viscous drag. Motions of an inner core with an inclined symmetry axis cause an inertial restoring torque.

All these torques are four orders of magnitude smaller than the torque caused by mutual gravitation between core and mantle.

Therefore in Smylie et al. (1984) only the torque due to mutual gravitation is considered. They found that this torque is given by

$$\underline{\underline{\Gamma}} = \Gamma\cos\vartheta_i \sin\vartheta_i \,\frac{\underline{\underline{e}}_3 \times \underline{\underline{i}}_3}{|e_3 \times i_3|} \tag{11}$$

where $\Gamma = 2.52 \cdot 10^{24}$ Nm.

As Smylie et al. (1984) have shown, it is possible to determine the velocity of rotation of the inner core by means

of this torque and assuming the westward drift of the geomagnetic dipole as precession of the inner core.

2.1 The torques due to the attraction of the Sun and the Moon

External torques acting on the inner core and the mantle can be calculated according to

$$\begin{aligned} \underline{L}_i &= \frac{3Gm}{R^5} (C_i - A_i) (\underline{R} \cdot \underline{i}_3) (\underline{R} \times \underline{i}_3) \\ \underline{L}_m &= \frac{3Gm}{R^5} (C - A) (\underline{R} \cdot \underline{e}_3) (\underline{R} \times \underline{e}_3) \end{aligned} \tag{12}$$

where G is the gravitational constant, m is the mass of the Sun or the Moon respectively. R is the geocentric distance of the celestial body which produces the torque.

Formulae (12) are well known from the theory of precession and nutation. In the co-ordinate system connected to the ecliptic the vector $\underline{R}$ is given by

$$\underline{R} = R \cos b \cos l \, \underline{e}'_1 + R \cos b \sin l \, (\underline{E}_3 \times \underline{e}'_1) + R \sin b \, \underline{E}_3 \, . \tag{13}$$

In equation (13) b and l are the latitude and the longitude in the ecliptic of the celestial body considered. R, b and l in equation (13) are temporally variable. The temporal variation of R will be neglected, because we consider only first order effects. Only those can be proved by geomagnetic field variations, because of the insufficient accuracy of the field data.

The vector $\underline{R}$ must be transformed into the inner core and the mantle frame, to evaluate the torques (12). The components of the torques, which must be introduced into the equations (8) can be written in the notation

$$\begin{aligned} \underline{L}_m \, \underline{i}'_1 &= \sum_j K_j \cos(\alpha_j t + \gamma_j) \\ \underline{L}_m \, (\underline{e}_3 \times \underline{i}'_1) &= - \sum_j K_j \sin(\alpha_j t + \gamma_j) \, . \end{aligned} \tag{14}$$

Similar we obtain for the components to be introduced in

equations (9)

$$\underline{\underline{L}}_i \; \underline{\underline{i}}_1' = \sum_j K_{j1} \cos(\alpha_j t + \gamma_j)$$
$$\underline{\underline{L}}_i \; (\underline{\underline{i}}_3 \times \underline{\underline{i}}_1) = -\sum_j K_{j2} \sin(\alpha_j t + \gamma_j). \qquad (15)$$

The circular frequencies α_j in expressions (14) and (15) depend on the temporal variations of the orbital elements of the Earth and the Moon, the rotational velocity of the Earth's mantle and the precessional velocity of the inner core. γ_j depends on the moment to which the calculations of the torques are related. K_j, K_{j1}, and K_{j2} are shown in Tables I and II for different circular frequencies. In the above mentioned tables are

$$(1) = \frac{3}{2} \frac{Gm_{\mathbb{C}}}{R_{\mathbb{C}}^3} \frac{C-A}{A} = 4.2422 \cdot 10^{-16}$$
$$(2) = \frac{3}{2} \frac{Gm_{\odot}}{R_{\odot}^3} \frac{C-A}{A} = 1.9538 \cdot 10^{-16}$$
$$(3) = \frac{3}{2} \frac{Gm_{\mathbb{C}}}{R_{\mathbb{C}}^3} \frac{C_i-A_i}{A_i} = 3.1087 \cdot 10^{-16}$$
$$(4) = \frac{3}{2} \frac{Gm_{\odot}}{R_{\odot}^3} \frac{C_i-A_i}{A_i} = 1.4318 \cdot 10^{-16} \; .$$

In Tables I and II are $\dot{l}_{\odot}$ and $\dot{l}_{\mathbb{C}}$ the velocities of the ecliptical longitudes of the Sun and the Moon. i is the inclination between the Moon's orbit and the ecliptic. – $\dot{\Omega}$ is the velocity of the retrograde motion of the Moon's node.

3. SOLUTIONS OF THE EQUATIONS OF MOTION

3.1 The solution of Smylie, Szeto and Rochester

Comparing the orders of magnitude of the external torques and the torque produced by mutual gravitation between the core and the mantle, we find that the relative torque ($\approx 10^{24}$ Nm) between the core and the mantle is much larger than the torques caused by the attraction of the Sun and the Moon ($\underline{\underline{L}}_i \approx 10^{19}$ Nm

Table I. Amplitudes of periodic terms of the torques exerting the mantle

1. Terms caused by the Sun

α_j	K_j/A
$\dot{\varphi}+\dot{\psi}_i$	$-\frac{(2)}{2}\sin 2\varepsilon$
$\dot{\varphi}+\dot{\psi}_i-2\dot{l}_\odot$	$\frac{(2)}{2}\left(\frac{1}{2}\sin 2\varepsilon+\sin\varepsilon\right)$
$\dot{\varphi}+\dot{\psi}_i+2\dot{l}_\odot$	$\frac{(2)}{2}\left(\frac{1}{2}\sin 2\varepsilon-\sin\varepsilon\right)$

2. Terms caused by the Moon

$\dot{\varphi}+\dot{\psi}_i$	$-\frac{(1)}{2}\sin 2\varepsilon$
$\dot{\varphi}+\dot{\psi}_i-2\dot{l}_\mathbb{C}$	$\frac{(1)}{2}\left(\frac{1}{2}\sin 2\varepsilon+\sin\varepsilon\right)$
$\dot{\varphi}+\dot{\psi}_i+2\dot{l}_\mathbb{C}$	$\frac{(1)}{2}\left(\frac{1}{2}\sin 2\varepsilon-\sin\varepsilon\right)$
$\dot{\varphi}+\dot{\psi}_i-2\dot{l}_\mathbb{C}+\dot{\Omega}$	$-\frac{(1)}{2}(\cos 2\varepsilon+\cos\varepsilon)\tan i$
$\dot{\varphi}+\dot{\psi}_i+2\dot{l}_\mathbb{C}+\dot{\Omega}$	$-\frac{(1)}{2}(\cos 2\varepsilon-\cos\varepsilon)\tan i$
$\dot{\varphi}+\dot{\psi}_i-\dot{\Omega}$	$\frac{(1)}{2}(\cos 2\varepsilon+\cos\varepsilon)\tan i$
$\dot{\varphi}+\dot{\psi}_i+\dot{\Omega}$	$\frac{(1)}{2}(\cos 2\varepsilon-\cos\varepsilon)\tan i$

and $L_m \quad 10^{22}$ Nm).

From this follows that the relative motion between the inner core and the mantle is mainly governed by the torque $\underline{\underline{\Gamma}}$.

The effect of the external torques will be considered as a

Table II. Amplitudes of periodic terms of the torques excerting the inner core

1. Terms caused by the Sun

α_j	K_{j1}/A_i	K_{j2}/A_i
secular	$-\frac{(4)}{2}(1+3\cos 2\varepsilon)\sin 2\vartheta_i$	-
$2\dot{l}_\odot$	$-\frac{3}{8}(4)(1-\cos 2\varepsilon)\sin 2\vartheta_i$	-
$\dot{\varphi}+\dot{\psi}_i$	$-\frac{(4)}{2}\sin 2\varepsilon\ \cos 2\vartheta_i$	K_{j1}/A_i
$\dot{\varphi}+\dot{\psi}_i+2\dot{l}_\odot$	$\frac{(4)}{4}\left(\frac{1}{2}\sin 2\varepsilon-\sin\varepsilon\right)\cos 2\vartheta_i$	$2\frac{K_{j1}}{A_i}\frac{\cos\vartheta_i}{\cos 2\vartheta_i}$
$\dot{\varphi}+\dot{\psi}_i-2\dot{l}_\odot$	$\frac{(4)}{4}\left(\frac{1}{2}\sin 2\varepsilon+\sin\varepsilon\right)\cos 2\vartheta_i$	
$2(\dot{\varphi}+\dot{\psi}_i+\dot{l}_\odot)$	$\frac{(4)}{4}\left(\frac{1}{2}(1+\cos^2\varepsilon)-\cos\varepsilon\right)\sin 2\vartheta_i$	$2\frac{K_{j1}}{A_i}\frac{\sin\vartheta_i}{\sin 2\vartheta_i}$
$2(\dot{\varphi}+\dot{\psi}_i-\dot{l}_\odot)$	$\frac{(4)}{4}\left(\frac{1}{2}(1+\cos^2\varepsilon)+\cos\varepsilon\right)\sin 2\vartheta_i$	
$2(\dot{\varphi}+\dot{\psi}_i)$	$\frac{(4)}{4}\sin^2\varepsilon\ \sin 2\vartheta_i$	

2. Terms caused by the Moon

secular	$-\frac{(3)}{8}(1+3\cos 2\varepsilon)\sin 2\vartheta_i$	-
$\dot{\Omega}$	$-\frac{3}{8}(3)\sin 2\varepsilon\ \tan i\ \sin 2\vartheta_i$	-

Table II (contd)

$2\dot{i}_{\mathbb{C}} - \dot{\Omega}$	$\frac{3}{4}(3)\sin 2\varepsilon \tan i \sin 2\vartheta_i$	-
$2\dot{i}_{\mathbb{C}}$	$-\frac{3}{8}(3)(1-\cos 2\varepsilon)\sin 2\vartheta_i$	-
$\dot{\varphi}+\dot{\psi}_i$	$-\frac{(3)}{2}\sin 2\varepsilon \cos 2\vartheta_i$	$\frac{K_{jl}}{A_i}\frac{\cos\vartheta_i}{\cos 2\vartheta_i}$
$\dot{\varphi}+\dot{\psi}_i+2\dot{i}_{\mathbb{C}}$	$\frac{(3)}{2}\left(\frac{1}{2}\sin 2\varepsilon - \sin\varepsilon\right)\cos 2\vartheta_i$	
$\dot{\varphi}+\dot{\psi}_i-2\dot{i}_{\mathbb{C}}$	$\frac{(3)}{2}\left(\frac{1}{2}\sin 2\varepsilon + \sin\varepsilon\right)\cos 2\vartheta_i$	
$\dot{\varphi}+\dot{\psi}_i+2\dot{i}_{\mathbb{C}}-\dot{\Omega}$	$-\frac{(3)}{2}(\cos 2\varepsilon - \cos\varepsilon)\tan i \cos 2\vartheta_i$	
$\dot{\varphi}+\dot{\psi}_i-2\dot{i}_{\mathbb{C}}+\dot{\Omega}$	$-\frac{(3)}{2}(\cos 2\varepsilon + \cos\varepsilon)\tan i \cos 2\vartheta_i$	
$\dot{\varphi}+\dot{\psi}_i-\dot{\Omega}$	$\frac{(3)}{2}(\cos 2\varepsilon + \cos\varepsilon)\tan i \cos 2\vartheta_i$	
$\dot{\varphi}+\dot{\psi}_i+\dot{\Omega}$	$\frac{(3)}{2}(\cos 2\varepsilon - \cos\varepsilon)\tan i \cos 2\vartheta_i$	
$2(\dot{\varphi}+\dot{\psi}_i)$	$\frac{(3)}{4}\sin^2\varepsilon \sin 2\vartheta_i$	$2\frac{K_{jl}}{A_i}\frac{\sin\vartheta_i}{\sin 2\vartheta_i}$
$2(\dot{\varphi}+\dot{\psi}_i+\dot{i}_{\mathbb{C}})$	$\frac{(3)}{4}\left(\frac{1}{2}(1+\cos^2\varepsilon) - \cos\varepsilon\right)\sin 2\vartheta_i$	
$2(\dot{\varphi}+\dot{\psi}_i-\dot{i}_{\mathbb{C}})$	$\frac{(3)}{4}\left(\frac{1}{2}(1+\cos^2\varepsilon) + \cos\varepsilon\right)\sin 2\vartheta_i$	
$2(\dot{\varphi}+\dot{\psi}_i-\dot{i}_{\mathbb{C}})-\dot{\Omega}$	$\frac{(3)}{4}\left(\frac{1}{2}\sin 2\varepsilon - \sin\varepsilon\right)\tan i \sin 2\vartheta_i$	
$2(\dot{\varphi}+\dot{\psi}_i-\dot{i}_{\mathbb{C}})+\dot{\Omega}$	$\frac{(3)}{4}\left(\frac{1}{2}\sin 2\varepsilon + \sin\varepsilon\right)\tan i \sin 2\vartheta_i$	
$2(\dot{\varphi}+\dot{\psi}_i)-\dot{\Omega}$	$-\frac{(3)}{4}\left(\frac{1}{2}\sin 2\varepsilon + \sin\varepsilon\right)\tan i \sin 2\vartheta_i$	
$2(\dot{\varphi}+\dot{\psi}_i)+\dot{\Omega}$	$-\frac{(3)}{4}\left(\frac{1}{2}\sin 2\varepsilon - \sin\varepsilon\right)\tan i \sin 2\vartheta_i$	

small disturbance of this motion. Smylie et al. (1984) discussed the problem considering the torque of mutual gravitation only. They solved the equations of motion supposing regular precession. This supposition is in accordance with the properties of the westward drift of the dipole field.

Regular precession requires

$$\dot{\psi}_i = \text{const}, \quad \dot{\varphi}_i = \text{const} \quad \text{and} \quad \vartheta_i = \text{const}.$$

Introducing these quantities in (8) and (9) we find also

$$\dot{\psi}_s = \text{const} \quad \text{and} \quad \vartheta_s = \text{const}.$$

With $\vartheta_i = 11^o$ and $\dot{\psi}_i = -0.05^o$ per year. Smylie et al. (1984) obtained the following quantities for the parameters to be determined

$$\begin{aligned}
\dot{\psi}_s &= \dot{\psi}_i + \omega_o + \frac{\Gamma \cos^2 \vartheta_i}{A(\sigma - \dot{\psi}_i)} = -8.4 \cdot 10^{-7} \text{ rad sec}^{-1} \\
\omega_{oi} &= \frac{A_i}{C_i} \dot{\psi}_s \cos \vartheta_i - \Gamma \cos \vartheta_i \left(\frac{1}{C_i \dot{\psi}_s} + \frac{A_i}{C_i A(\sigma - \dot{\psi}_i)} \right) = \\
&= 350^o \text{ d}^{-1} \\
\dot{\varphi}_i &= \Gamma \cos \vartheta_i \left(\frac{C_i - A_i}{A_i} \frac{1}{A(\sigma - \dot{\psi}_i)} - \frac{1}{C_i \dot{\psi}_s} \right) - \\
&\quad - \frac{C_i - A_i}{C_i} \dot{\psi}_s \cos \vartheta_i = -7.2 \cdot 10^{-7} \text{ rad sec}^{-1} \\
\sigma_i &= \frac{C_i - A_i}{A_i} \omega_{oi} = 1.70 \cdot 10^{-7} \text{ rad sec}^{-1}
\end{aligned} \tag{16}$$

3.2 Relative motion of the inner core, caused by external torques

$\Delta\dot{\psi}_{ij}$, $\Delta\dot{\vartheta}_{ij}$, $\Delta\dot{\varphi}_{ij}$, denote disturbances of the parameters of the relative motion caused by a periodic constituent of the external torques. If we mark the parameters valid for regular precession by a zero index, we obtain the following parameters for the complete motion

$$\begin{aligned}\dot{\psi}_i &= \dot{\psi}_{oi} + \sum_j \Delta\dot{\psi}_{ij} \\ \dot{\vartheta}_i &= \sum_j \Delta\dot{\vartheta}_{ij} \\ \dot{\varphi}_i &= \dot{\varphi}_{oi} + \sum_j \Delta\dot{\varphi}_{ij} \cdot\end{aligned} \tag{17}$$

Similar expressions are obtained for the spatial components of the rotational velocity of the inner core $\dot{\psi}_s$, $\dot{\varphi}_s$ and $\dot{\vartheta}_s$.

If we introduce the expressions (17) in the equations (8) and (9), we obtain for each periodic constituent of the external torques following differential equations taking into account the relations (16)

$$\begin{aligned}&\frac{d}{dt}(\Delta\dot{\vartheta}_s - \Delta\dot{\vartheta}_i) - (\sigma - \dot{\psi}_i)(\Delta\dot{\varphi}_s - \Delta\dot{\varphi}_i)\sin\vartheta_i = \frac{K_j}{A}\cos(\alpha_j t + \gamma_j) \\ &\frac{d}{dt}(\Delta\dot{\psi}_s - \Delta\dot{\psi}_i)\sin\vartheta_i + (\sigma - \dot{\psi}_i)(\Delta\dot{\vartheta}_s - \Delta\dot{\vartheta}_i) = \frac{K_j}{A}\sin(\alpha_j t + \gamma_j) \\ &\Delta\dot{\psi}_s - \Delta\dot{\psi}_i = -(\Delta\dot{\varphi}_s - \Delta\dot{\varphi}_i)\cos\vartheta_i\end{aligned} \tag{18}$$

$$\begin{aligned}&\frac{d}{dt}\Delta\dot{\psi}_s + (\sigma_i + \dot{\varphi}_i)\Delta\dot{\psi}_s \sin\vartheta_i = \frac{K_{j1}}{A_i}\cos(\alpha_j t + \gamma_j) \\ &\frac{d}{dt}\Delta\dot{\psi}_s \sin\vartheta_i - (\sigma_i + \dot{\varphi}_i)\Delta\dot{\vartheta}_s = -\frac{K_{j2}}{A_i}\sin(\alpha_j t + \gamma_j) \\ &\Delta\dot{\varphi}_s = -\Delta\dot{\psi}_s \cos\vartheta_i \cdot\end{aligned} \tag{19}$$

In equations (18) and (19) and in all that follows we cancel the index zero of the components of rotation for simplification.

The solution of the equations (18) and (19) yields values for the spatial and relative rotational velocity components of the inner core. After some simple calculations and an additional integration we obtain following periodic constituents of the Euler angles, which stand for the relative motion between inner core and mantle

$$\Delta\vartheta_i = -\frac{1}{\alpha_j}\,\frac{1}{A_i}\,\frac{\alpha_j K_{j1}-(\sigma_i+\dot{\varphi}_i)\,K_{j2}}{\alpha_j^2-(\sigma_i+\dot{\varphi}_i)^2} - \\ -\frac{K_j}{A}\,\frac{1}{\alpha_j+(\sigma-\dot{\psi}_i)}\,\cos(\alpha_j t+\gamma_j) + \Delta\vartheta_{oi}$$

$$\Delta\psi_i = \frac{1}{\alpha_j}\left(\frac{1}{A_i\sin\vartheta_i}\,\frac{\alpha_j K_{j2}-(\sigma_i+\dot{\varphi}_i)\,K_{j1}}{\alpha_j^2-(\sigma_i+\dot{\varphi}_i)^2} - \right. \\ \left. -\frac{K_j}{A}\,\frac{\cot\vartheta_i}{\alpha_j+(\sigma-\dot{\psi})}\right)\sin(\alpha_j t+\gamma_j) + \Delta\psi_{oi} \qquad (20)$$

$$\Delta\varphi_i = \frac{1}{\alpha_j}\left(-\frac{\cot\vartheta_i}{A_i}\,\frac{\alpha_j K_{j2}-(\sigma_i+\dot{\varphi}_i)\,K_{j1}}{\alpha_j^2-(\sigma_i+\dot{\varphi}_i)^2} + \right. \\ \left. +\frac{K_j}{A\sin\vartheta_i}\,\frac{1}{\alpha_j+(\sigma-\dot{\psi}_i)}\right)\sin(\alpha_j t+\gamma_j) + \Delta\varphi_{oi}\,.$$

By means of the amplitudes K_j, K_{j1}, K_{j2} contained in Tables I and II the periodic constituents of the inner core motion can be evaluated according to equations (20). A number of periodic constituents in Tables I and II depend on circular frequencies containing $\dot{\varphi}$. These constituents vary with a nearly diurnal period. Other constituents vary with nearly semi-diurnal periods.

These terms cause variations of $\Delta\vartheta_i$ and $\Delta\psi_i$ which do not exceed 0.01". These small quantities cannot be derived from the coefficients of the spherical-harmonic expansion of geomagnetic field data.

Besides the insufficient accuracy the not determinable origin of these periods prevents the investigation of such small variations. They can be attributed to the internal or external geomagnetic field.

3.21 The influence of long-period variations of the external torques

In Table II we find long-period constituents which depend on temporal variations of the longitudes $l_\odot$ and $l_\mathbb{C}$ and the

velocity of the retrograde motion of the Moon's node. For these terms the coefficients K_{j1} and K_{j2} vanish and we obtain from equations (20) following relations

$$\Delta\vartheta_i = \Delta\vartheta_{oi} + \frac{K_{j1}}{A_i\,(\alpha_j^2-(\sigma_i+\dot{\varphi}_i)^2)}\cos(\alpha_j t+\gamma_j)$$

$$\Delta\psi_i = \Delta\psi_{oi} - \frac{\sigma_i+\dot{\varphi}_i}{\alpha(\alpha_j^2-(\sigma_i+\dot{\varphi}_i)^2)}\frac{K_{j1}}{A_i\sin\vartheta_i}\sin(\alpha_j t+\gamma_j)\,. \tag{21}$$

Using the quantities K_{j1} given in Table II we obtain following amplitudes of the variations $\Delta\vartheta_i$ and $\Delta\psi_i$

period		$\Delta\vartheta_i$	$\Delta\psi_i$
$2\dot{i}_{\mathbb{C}}$	(fourteen days)	0.102"	0.055"
$2\dot{i}_{\mathbb{C}} - \dot{\Omega}$	(nearly fourteen days)	0.042"	0.022"
$2\dot{i}_{\odot}$	(semi-annual)	9.129"	66.084"
$\dot{\Omega}$	(18.6 years)	3.916'	0.291°

Only the 18.6 years period has a sufficient large amplitude to be determined from geomagnetic data. The potential of the geomagnetic dipole is given by

$$V_{Dip} = r_o\,(g_1^0 P_1^0(\cos\vartheta) + (g_1^1\cos\lambda + h_1^1\sin\lambda)\,P_1^1(\cos\vartheta) \tag{22}$$

from which we obtain the parameters of the westward drift

$$\vartheta_i = \arctan\frac{(g_1^{1\,2} + h_1^{1\,2})^{1/2}}{g_1^0}$$

$$\psi_i = \arctan\frac{h_1^1}{g_1^1}\,, \tag{23}$$

which are per definition the parameters of the relative precession and nutation of the inner core. The spherical harmonic expansion of the geomagnetic field is given for different epochs. From these expansions we can construct time series for ϑ_i and ψ_i. These time series allow to determine the periodic constituents contained in the temporal variations of the above

mentioned parameters, according to suitable mathematical procedures. We used a method published by Jochmann (1986). The result is exhibited in Fig. 3. For an 18.6 years period results an amplitude of

$$\Delta\psi_i = 0.103^o$$

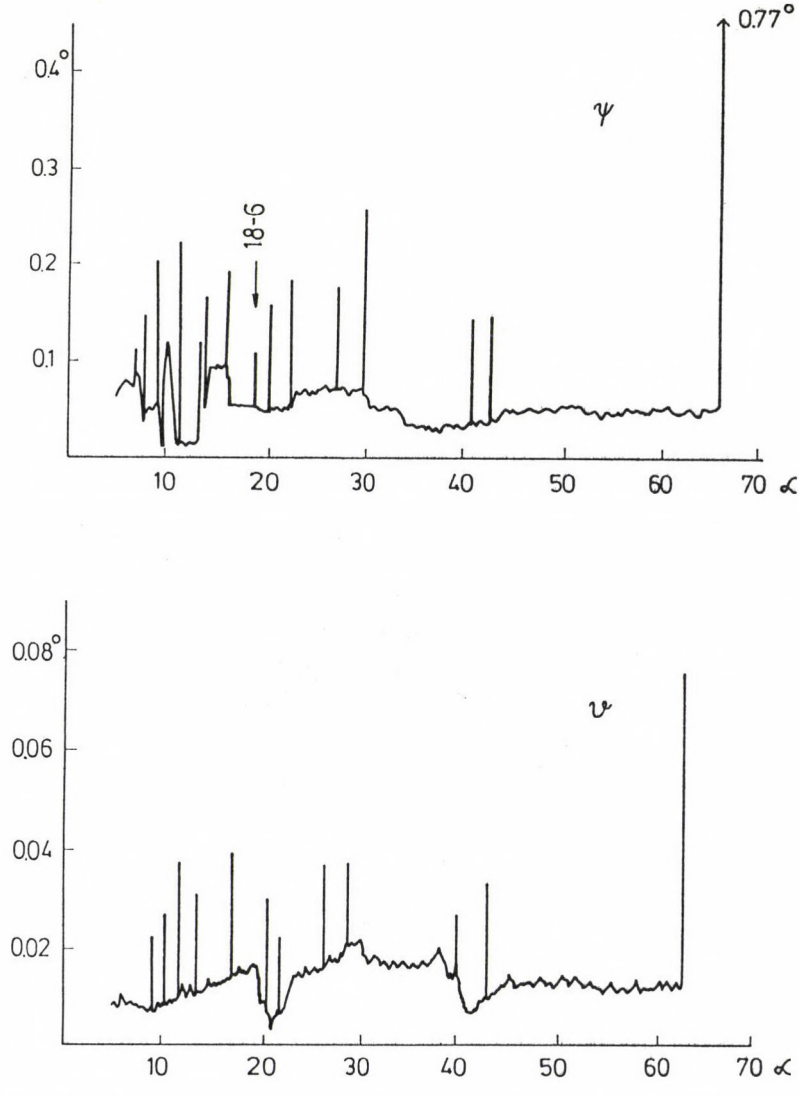


Fig. 3. Amplitude spectra of the direction angles of the geomagnetic dipole axis

which agrees sufficiently with the theoretically calculated value, because the mean square error of the estimated quantity amounts to 0.1^{o}. This result can be supported by an investigation of the 18.2 years periodic constituent of polar motion. This constituent of polar motion is caused by inner core motion and tidal effects. The latter effect is very small and can be neglected. The inner core motion contributes to the excitation function of polar motion. By Jochmann (1989) a method for the construction of the excitation function of inner core motion is given. Applying this and calculating the corresponding periodic constituent of polar motion, we found for the 18.6 years period an elliptical pole path with the semi-axes

$$a = 0.008"$$
$$b = 0 \quad .$$

From the amplitude spectrum of polar motion results

$$a = 0.007"$$
$$b = 0.0004".$$

Both values agree very well. Since by Jochmann (1989) the relation between the variations of the geomagnetic dipole and polar motion was proved for a number of periodic terms, this result confirms the relation between inner core motion and the westdrift of the geomagnetic dipole.

3.22 The effect of secular variations of the external torques

Table II shows that the external torques acting on the inner core contain constant coefficients which are the reason for secular variations of the relative motion between the inner core and the mantle. Solving Eqs (9) and (10) for these terms, we obtain

$$\begin{aligned} \Delta\psi_i &= 6.178 \cdot 10^{-10}\ t + \Delta\psi_o \\ \Delta\vartheta_i &= 2.14\ \ 10^{-4}\ \text{rad} - \frac{K_1}{\sigma_i - \rho_i} \quad . \end{aligned} \tag{24}$$

The secular variation corresponds to a variation of the relative precession with a period of 320 years. This term of the inner core motion can be considered as an additional nutation imposed on the regular precession. The constants $\Delta\psi_0$ and K_1 cannot be determined because of the lack of initial values.

In archaeomagnetic time series, which were investigated by Adam (1983), there is a 360 years period shown, which may correspond to the secular motion (24) (see Fig. 4).

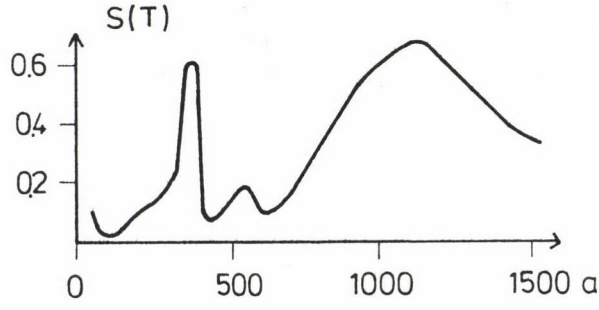


Fig. 4. Spectral density of archaeomagnetic data (according to Adam N V 1983)

4. CONCLUSIONS

In Jochmann (1989) the correlation between variations of geomagnetic field data and polar motion could be proved, using the hypothesis of a rotating, precessing inner core according to Smylie et al. (1984). We extended the theory given in the quoted paper to the influence of external torques caused by the attraction of the Sun and the Moon. It could be proved that periodic terms in the westward drift of the geomagnetic field were caused by inner core nutations due to corresponding terms of the temporally variable external torques. Similar corresponding terms were found in the time series of polar motion.

From these results follows that the applied hypothesis is useful to explain some relations between the geomagnetic westward drift, polar motion and the orbital elements of the Earth

and the Moon. On the other hand the results seem to confirm the hypothesis of a rotating, precessing inner core. But, one must mention that the results are only a first step to prove this hypothesis.

Extended time series of the investigated phenomena, which may be available in future will possibly confirm the obtained results.

REFERENCES

Adam N V 1983: The westward drift of the geomagnetic field. In: Bucha V ed., Magnetic field and processes in the Earth's interior. Academia Prague

Bullard E C 1949: In: Proc. Royal Soc. London, A 197, 443-453.

Jochmann H 1986: Vermessungstechnik, 34, 18-20.

Jochmann H 1989: Astr. Nachr., 310, 435-442.

Schmutzer E 1978: Gerlands Beitr. Geophys. Leipzig, 87, 455-468.

Smylie D E, Szeto A M R, Rochester M G 1984: Rep. Prog. Phys. London, 47, 855-906.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 221-235 (1991)

MAGNETOTELLURIC MODELING OF THREE-DIMENSIONAL BODIES IN A LAYERED EARTH

Chen Jiuping, Chen Leshou, Wang Guange

Department of Geophysics, China University of Geosciences, Beijing 100083, 29 Xue Yuan Road, China

[Manuscript received June 12, 1990]

In this paper, we have developed an algorithm based on the method of integral equations (IE) to simulate the magnetotelluric (MT) response of three-dimensional (3D) bodies in a layered earth. It has two steps. First is to compute the equivalent volume currents in inhomogeneities. The 3D bodies are replaced by an equivalent volume current distribution which is approximated by pulse basis functions. The integral equations of unknown electric fields in 3D bodies are constructed using the tensorial Green functions appropriate to a layered earth. The second step is to find the scattered fields at the earth surface by integrating the electric and magnetic tensorial Green functions over the scattering currents.

In order to reduce computation time and memory, three kinds of processing methods are adopted. They are Hankel transformation and interpolation, group transforms, and the sparse tensorial Green function matrix.

A comparison of our responses over a 3D body in a uniform half-space to these reported in earlier works and a checking of the reciprocity of the Green functions have verified our program. The distortions of 3D MT curves are preliminarily discussed by practical modeling and some new suggestions are presented.

Keywords: Green function; Hankel transform; integral equation; magnetotelluric response; magnetotelluric modeling; numerical modeling; 3D modeling

INTRODUCTION

The straightforward problem of 3D MT responses in an important part of magnetotellurics. The integral equations (IE) method was previously used in electromagnetic simulations to model the inhomogeneities in a half-space by Hohmann (1975). Later Wannamaker et al. (1984) developed an IE method in a layered earth for MT modeling and obtained a lot of useful results. Owing to its advantages less computation time and

small memory, great attention is paid to this method and has became one of the most commonly used simulation tools (Varentsov 1983).

The recent developments in MT field surveys and data interpretation in China have profoundly contributed to our understanding of the relations between the subsurface resistivity structure and observed MT quantities. However, existing 1D and 2D MT theories are not enough to rationally interprete the practical observed data and it is necessary to discuss the characteristics of distortions of 3D MT curves.

Starting with the Maxwell equations and using Schelkunoff symmetric potentials, the dyadic Green functions in a layered earth are derived following Wannamaker et al. (1984). Then the secondary vectorial Fredholm integral equations are constructed for the electric fields in the inhomogeneity. In order to improve the computation efficiency, we have used three methods in digital processing. They are Hankel transformation and interpolation, group transforms, and sparse Green matrix. Finally, we verified our program and studied the distortions of 3D MT based on practical modeling.

BASIC PRINCIPLES OF INTEGRAL EQUATIONS

Consider a 3D body with the conductivity σ_b in a N-layered host, shown in Fig. 1 confined to layer j for simplicity. Impressed with a downward propagating plane wave, the scattering electric and magnetic fields (E_s, H_s) as functions of position R and for an exp ($i\omega t$) time depedence obey the Helmholtz equations (Wannamaker et al. 1984)

$$(\nabla^2+K_l^2)\, E_s = 0 \qquad (l\neq j) \qquad (1a)$$

$$(\nabla^2+K_l^2)\, H_s = 0 \qquad (l\neq j) \qquad (1b)$$

$$(\nabla^2+K_j^2)\, E_s = \left(-\frac{1}{\hat{Y}_j}\nabla\nabla \cdot + \hat{Z}_j \right) J_s + \nabla \times M \qquad (l=j) \qquad (1c)$$

$$(\nabla^2+K_j^2)\, H_s = \left(-\frac{1}{\hat{Z}_j}\nabla\nabla \cdot + \hat{Y}_j \right) M_s - \nabla \times J \qquad (l=j) \qquad (1d)$$

where

$$J_S = (\hat{Y}_b - \hat{Y}_j)\, E \tag{2a}$$

$$M_S = (\hat{Z}_b - \hat{Z}_j)\, H\ . \tag{2b}$$

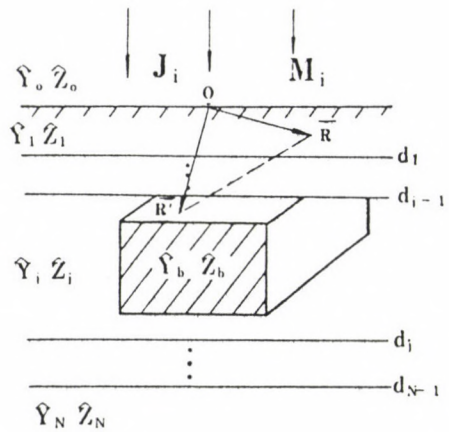


Fig. 1. Geometry of 3D body in layered earth

Here J_S and M_S are called equivalent scattering electric and magnetic current substituting for the body. $\hat{Y}$ is the admittivity and $\hat{Z}$ is the impedance at any point. $K = \sqrt{-\hat{Y}\hat{Z}}$ is the wawe-number in any layer. The relations among $\hat{Y}$, $\hat{Z}$ conductivity σ, dielectric permittivity ε and magnetic permeability μ are

$$\hat{Y} = \sigma + i\omega\varepsilon \tag{3a}$$

$$\hat{Z} = i\omega\mu \tag{3b}$$

where i is the imaginary unit. Subscript b represents a 3D body, j is the layer which contains the 3D body, 1 is any layer.

If $\bar{\bar{G}}_1^E(R|R')$ and $\bar{\bar{G}}_1^H(R|R')$ are the electric and magnetic dyadic tensorial Green functions in half-space with layered earth, the solution of Eqs (1) are given by the method of Green

functions according to the vectorial Green theorem. We only consider the electric difference of layered earth and neglect the displacement currents. The integral equations are as follow

$$E_s(R) = \iiint_V \bar{\bar{G}}_1^E(R|R') \cdot J_s(R')dv \qquad (4a)$$

$$H_s(R) = \iiint_V \bar{\bar{G}}_1^H(R|R') \cdot J_s(R')dv \ . \qquad (4b)$$

The quantitites $\bar{\bar{G}}_1^E(R|R')$ and $\bar{\bar{G}}_1^H(R|R')$ are the electric and magnetic fields which relate a vector field at R in layer 1, to a current element at R in layer j, including the case 1=j.

Using decomposition of the total fields (E,H) and Eq. (2), a vectorial Fredholm integral equation of the second-order can be written about electric fields in the inhomogeneity as

$$E(R) = E_i(R) + (\sigma_b - \sigma_j) \iiint_V \bar{\bar{G}}_1^H(R|R') \cdot J_s(R')dv \qquad (5)$$

where E_i is a plane wave response in a layered earth. A simple matrix equation (6) with order 3N is obtained using the method of collocation with pulse subsection basis functions. N is the number of rectangularly prismatic cells by which the 3D body is approximated.

$$[\bar{\bar{Z}}] \cdot [E] = [E_i] \ . \qquad (6)$$

From these we can calculate all MT responses of the 3D body at the surface of earth by solving for the electric fields within the cells, by computing the scattering fields about the inhomogeneity, and by using the discrete version of Eq. (2).

It should be noted that when the difference of conductivities between host and 3D inhomogeneity is large and the working frequencies are also high, the vortex currents in a 3D body should be considered in the integral equation. Thus integral equation (5) should also be modified corresponding to

eddy currents. This is one of the important parts in which we have modified Wannamaker's program using the IE method. Unfortunately, by now we have not seen any discussion about this aspect in published papers. Our program has demonstrated that eddy currents are significant in 3D MT modeling under some conditions. A detailed discussion is found at Chen Jiuping (1987).

NUMERICAL PROCESSING

1. Hankel transformation and interpolation

The primary tensorial Green functions are analytic expressions, but the secondary ones require much time to calculate the Hankel transformation of the complicated kernel functions of orders 0 and 1. If the inhomogeneity is divided into N cells and M computing points are needed at the surface of the earth, there are 6N+11M*N total transforms. For overcoming this difficulty, we set up six electric and five magnetic Hankel transform tables prior to the matrix equation. Then we compute secondary Green functions by cubic interpolation using Newton's divided difference method (Wannamaker et al. 1984). In Hankel transforms, Anderson's (1975) numerical filters are used with 283 coefficients.

If a 3D body is divided into 128 cells and no interpolation is used, we need 2×10^5 Hankel transforms. Otherwise, only 10^3 transforms are needed. There is 200 times difference between them.

The interpolation reduces not only the computation time, but also satisfies the requirement of a certain precision. Table I is a comparison of analytic to interpolation results using $10 \times 5 \times 5$ basic points.

2. Group transformation

A unitary transformation matrix can be constructed using group representation theory as shown in Fig. 2, if all the

Table I

Calculating Point R = 1500 M,Z = 750 M,Z' = 750 M								
Kernel functions	F_1		F_2		F_3		F_4	
	Re	Im	Re	Im	Re	Im	Re	Im
interpolation	2.4961E-5	3.6623E-9	2.2455E-4	6.7377E-4	2.0132E-8	4.3245E-12	2.0187E-8	4.3444E-12
analytic	2.4546E-5	3.5278E-9	2.3561E-4	6.4676E-4	1.9291E-8	4.1391E-12	1.9291E-8	4.1331E-12
error	1.6 %	3.8 %	4.7 %	4.1 %	4.3 %	4.4 %	4.6 %	5.1 %

horizontal cross sections of a 3D body with vertical edges possess a certain symmetry (Tripp et al. 1984). Through matrix $\bar{\bar{U}}$ the original coefficient matrix $\bar{\bar{Z}}$ of dimension $3N \times 3N$ in Eq. (6), containing redundant information, can be transformed into a block-diagonal matrix, $\bar{\bar{Z}}'$, each of whose four diagonal submatrices have dimensions $3N/4 \times 3N/4$. The transform formula is

$$\bar{\bar{Z}}' = \bar{\bar{U}}\ \bar{\bar{Z}}\ \bar{\bar{U}}^T \tag{7}$$

where $\bar{\bar{U}}^T$ is the transpose of $\bar{\bar{U}}$. Similarly, E and E_i can be transformed into E' and E_i' by Eqs (8).

$$E' = \bar{\bar{U}}\ E \tag{8a}$$

$$E_i' = \bar{\bar{U}}\ E_i\ . \tag{8b}$$

It can be demonstrated that the original Eq. (6) is equivalent to Eq. (9)

$$[\bar{\bar{Z}}] \cdot [E'] = [E_i']\ . \tag{9}$$

Having solved for transformed electric fields by anti-transformation

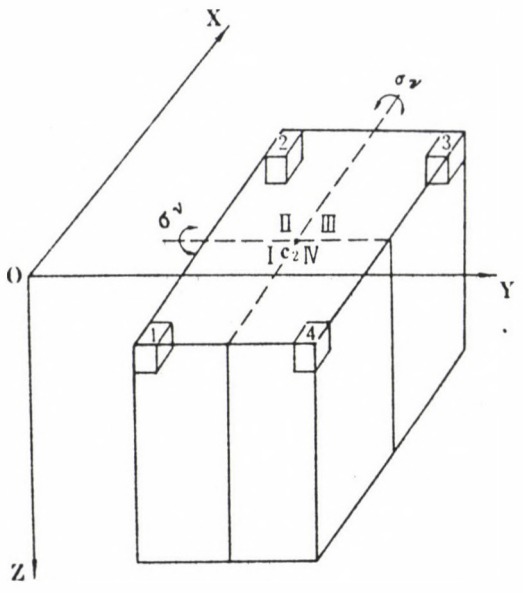


Fig. 2. The horizontal cross sections of a body with vertical sides possess a certain symmetry (point group C_2)

$$E = \bar{\bar{U}}\, E' \tag{10}$$

the electric fields E in Eq. (6) can be easily obtained.

Using this technique we only need 1/16 of the storage capacity and 1/4 at matrix formation time required for the original matrix.

3. Sparse Green matrix

The elements of matrix $\bar{\bar{Z}}$ are related to the relative distance of the observed cell to the source cell according to the physical meaning of the Green functions. Compared to the primary Green functions, the secondary ones are small and can be neglected if the distance between the cells are large enough. Therefore, the secondary Green matrix can become a

sparse matrix according to a certain precision requirement.

3D MODEL STUDY

The program was verified by comparing its 3D MT modeling results with those reported in earlier works and by checking the reciprocity of the Green functions.

Figures 3 and 4 are comparisons of Wannamaker's IE, Lee's hybrid method and our modeling results for TE and TM mode excitations. It is clear that our curves are very close to Wannamaker's ones.

Table II contains the checking of reciprocity of the Green functions. Here the Green function imply their secondary part. The reciprocity relation is given as

$$G_{ij}(R|R') = G_{ji}(R'|R) \qquad i,j=x,y,z \ . \tag{11}$$

Table II

		F_1		F_2	
1		Re	Im	Re	Im
	1	1.7241E-5	5.6768E-8	1.3347E-8	4.9109E-11
	2	1.7241E-5	5.6768E-8	1.3347E-8	4.9109E-11

	F_3		F_4		F_5
2	Re	Im	Re	Im	Re
	2.5321E-6	7.4002E-5	2.6738E-8	1.9005E-10	1.3422E-8
	2.5321E-6	7.4002E-5	2.6738E-8	1.9005E-10	1.3422E-8

	F_5	F_6	
3	Im	Re	Im
	2.9557E-10	2.6736E-8	1.8986E-10
	2.9557E-10	2.6736E-8	1.8986E-10

The model parameters are shown in Fig. 5. The working frequency is 0.01 Hz. Source and observed points are (0, 0, 750), (1, 1000, 1500).

Consider a 3D body of low resistivity in a three-layered earth (Fig. 5). Multifrequency maps of the tensorial apparent

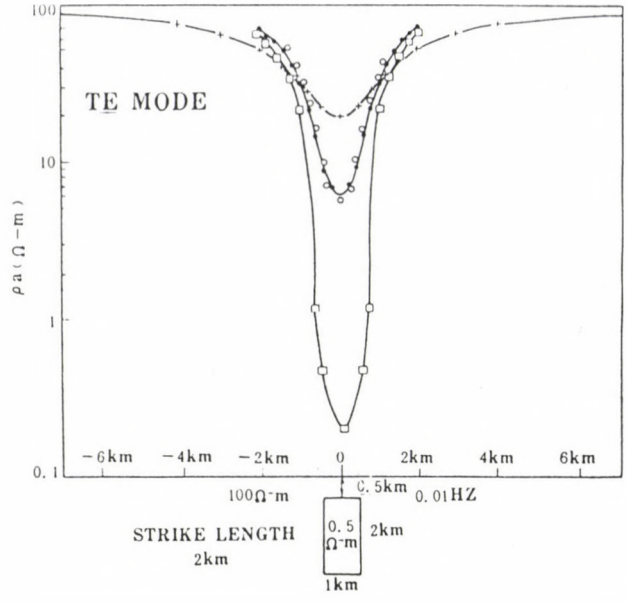


Fig. 3. Comparison of Wan's, Lee's and the present MT results for TE-mode excitation. Legend: + 2D; □ Lee et al.; • Wannamaker et al.; ○ Chen et al.

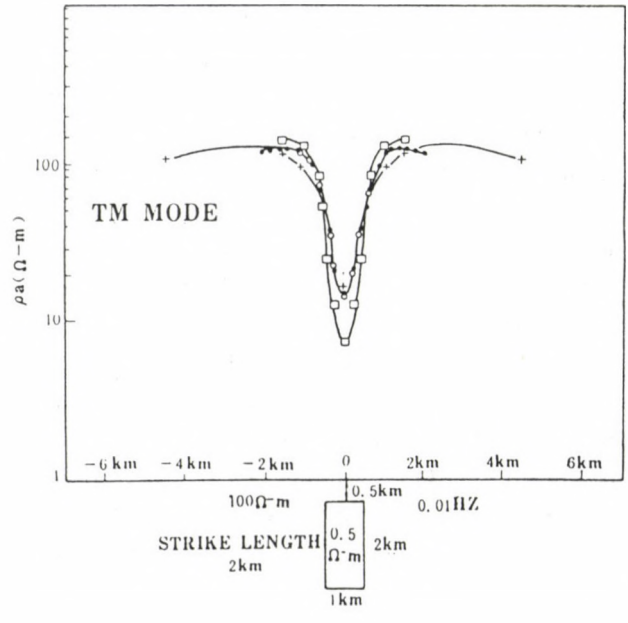


Fig. 4. Comparison of Wan's, Lee's and the present MT results for TM-mode excitation. Legend: + 2D; □ Lee et al.; • Wannamaker et al.; ○ Chen et al.

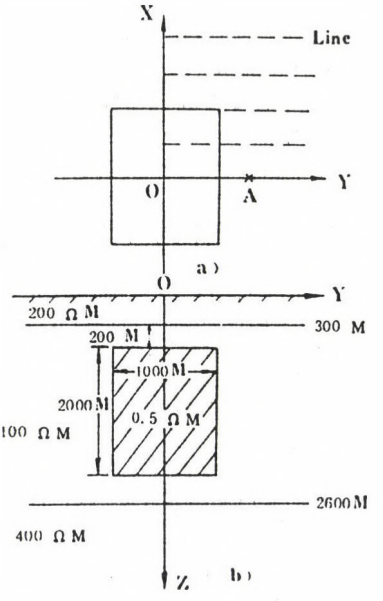


Fig. 5. Theoretical 3D model: a) plan map, b) section map

resistivities ϱ_{xy} and ϱ_{yx} are shown in Fig. 6. At low frequencies (<0.1 HZ), the contours of ϱ_{xy} and ϱ_{yx} are close to each other especially near the edge of 3D body, and form a "resistivity gradient band". If frequencies are higher than 5 Hz, the contours become sparse and anomalies are not clear. This phenomenon is explained as a "static field distortion" (Chen 1987).

Figure 7 shows sounding curves of the apparent resistivity over a 3D body at site A as well as 1D and 2D models which correspond to the 3D model. Note at low frequencies that the 3D ϱ_{xy} and ϱ_{yx} curves are seriously departed. However, the 2D and 3D ϱ_{yx} curves are close to each other at all frequencies.

It should be emphasized that impedance phase curves are different from pseudo-resistivities in distorting characteristics (Fig. 7b). At high and low frequencies, the two phase curves are almost undistorted, but at middle range, they departed clearly.

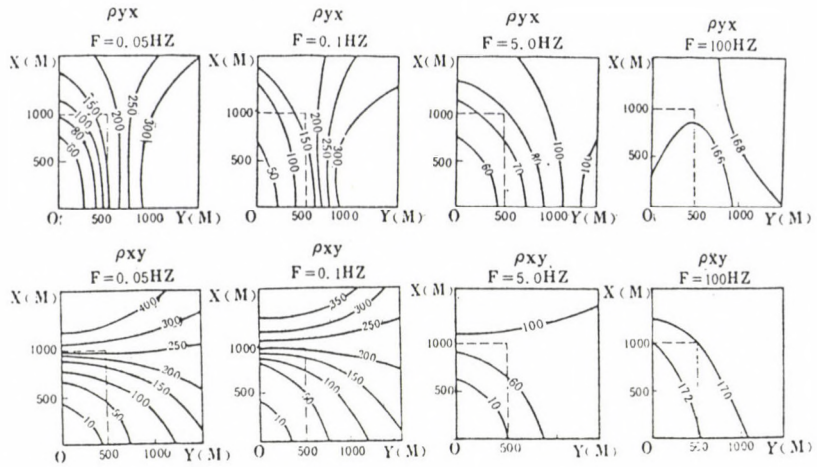


Fig. 6. Multifrequency plan map of apparent resistivities ϱ_{xy} and ϱ_{yx}

A map and a section of the tipper magnitude T are displayed over a 3D body in Fig. 8. The tipper strongly changes vs. frequency and it looks like a band filter. The tipper changes most quickly in the range 1-10 Hz and quickly decreases beyond this range. We define this range as the best range of frequencies for the tipper. It is not difficult to see that this best range is just corresponding to the penetration depth in this example.

In the map, the contours are extended in a direction which is parallel to the strike of the elongation in 3D and gathered near the edge of the inhomogeneity.

Generally speaking, the skew S is an important parameter which is an indicator of the dimensions of subsurface targets. In the map of Fig. 9a, the contours appear as a bundle of inclined ellipses. As the observed sites are on the symmetry axis of the prismatic 3D body, the skew approaches zero. This illustrates that S cannot present the characteristics of 3D bodies in some special positions.

As shown in the pseudo-sections of the skew, Fig. 9b, the contours look like an inverted V, the axis is located at the right side and is parallel to the frequency axis. Moreover,

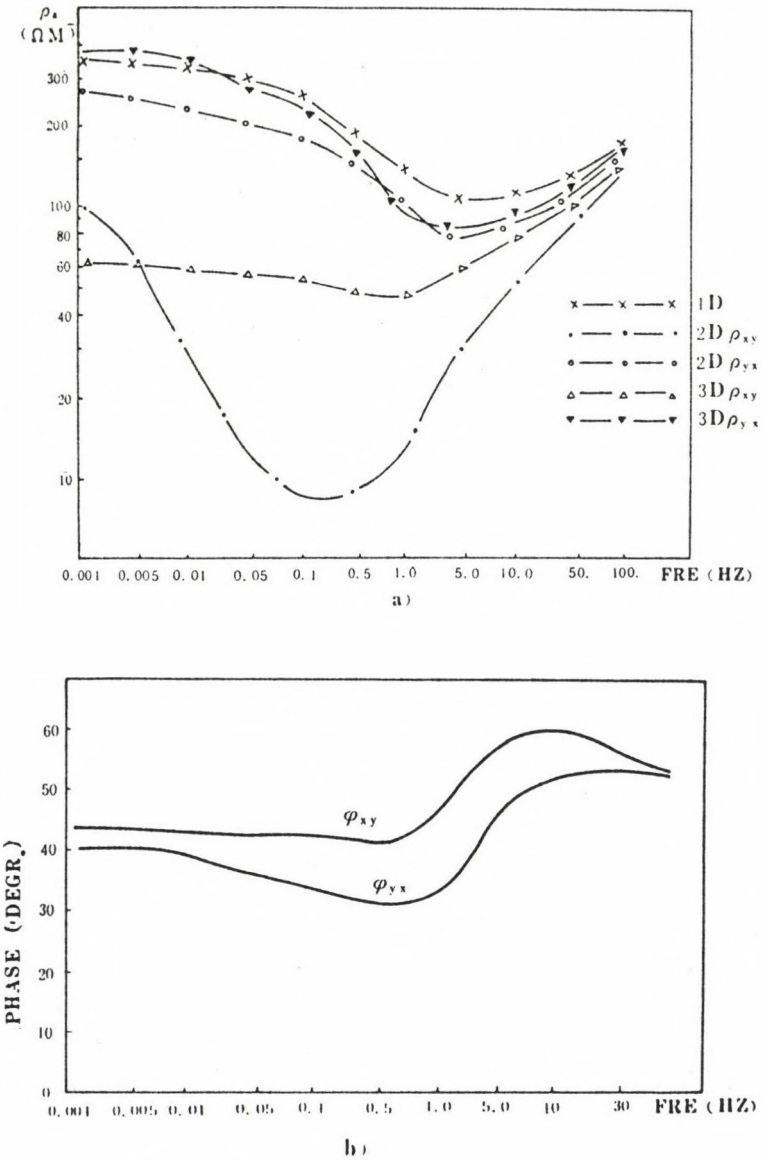


Fig. 7. MT sounding curves at site A

the magnitude decreases if the frequency increases. Therefore, the effective frequency of the skew is at low frequencies. The skew is affected by the distortion due to inhomogeneities.

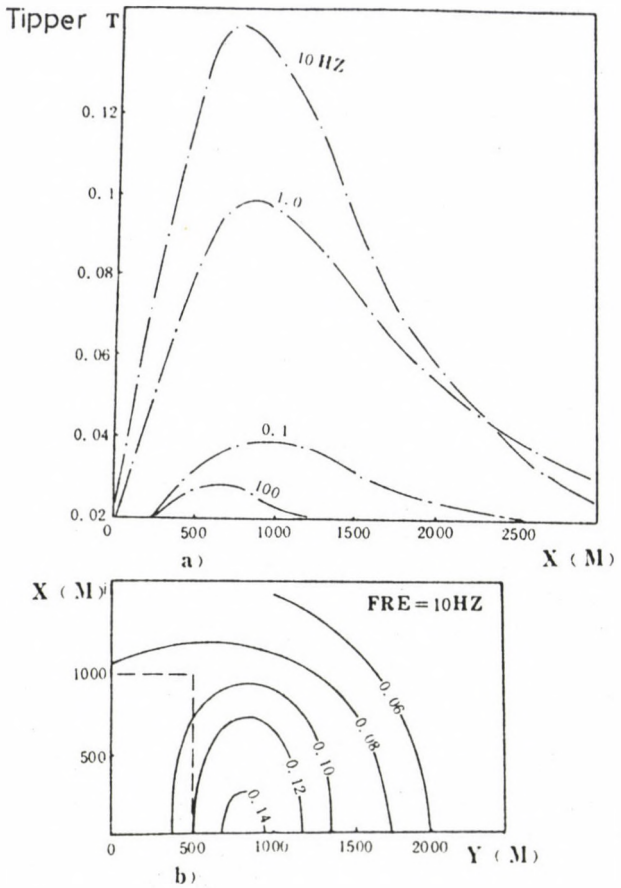


Fig. 8. Tipper T: a) profiles at x=Ø m, b) plan map

DISCUSSIONS AND CONCLUSIONS

Based on the modeling results over a 3D body in a layered earth, we conclude that

1. The distortions of 3D MT responses can be still divided into two types: galvanic and induced ones. But galvanic ones are more significant than induced ones. Essentially, galvanic distortions may be due to the static effect which is the direct reason of the departing of ρ_{xy} and ρ_{yx} at all frequencies.

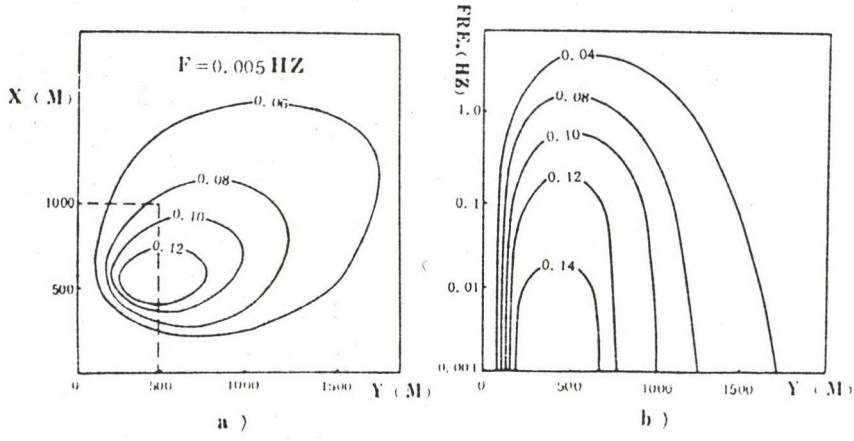


Fig. 9. Skew S: a) plan map, b) pseudosections at x=500 m

2. The tipper has the characteristic to indicate the working frequencies. At the best frequency range, the tipper reflects the existence of a subsurface target. The strike and the border of an elongated inhomogeneity can be determined according to the map of tipper magnitude, so the tipper is a very useful parameter.

3. The effective frequency range of the skew corresponds to low frequencies. It is not enough to determine the dimensions of a 3D body only from the values of the skew at a single site.

REFERENCES

Anderson W L 1979: Geophysics, 44, 1287-1305.

Hohmann G W 1975: Geophysics, 40, 309-324.

Jiuping Chen 1987: Three-dimensional MT modeling of integral equations. Dissertation for Master degree in China University of Geosciences

Lee K H, Pridmore D F, Morrison H F 1981: Geophysics, 46, 796-805.

Tripp A L, Hohmann G W 1984: Block diagonalization of the EM impedance matrix of a symmetric buried body using group theory. IEEE Trans. on geoscience and sensing, GE-22, 62-69.

Varentsov I M 1983: Geophys. Survey, 6, 55-78.

Wannamaker P E, Hohmann G W, Sanfilipo W A 1984: Geophysics, 49, 60-74.

Wannamaker P E, Hohmann G W, Ward S H 1984: Geophysics, 49, 1517-1537.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 237-251 (1991)

A JOINT VIEW OF GEOMAGNETIC, IONOSPHERIC AND THERMOSPHERIC DISTURBANCES*

P Bencze

Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H-9401 Sopron, POB 5, Hungary

Common features of the storm time and disturbance daily variations of the geomagnetic field, the ionosphere and the thermosphere are discussed. An attempt is made to use the results achieved in the interpretation of one of these phenomena, e.g. in the geomagnetic disturbances for the explanation of the others. It has been found that as energy sources producing these variations not only auroral processes, but also particle precipitation from the ring current belt should be taken into consideration.

Keywords: energetic particles; geomagnetic disturbances; ionospheric storms; ring current; thermospheric disturbances

INTRODUCTION

If the variations of the geomagnetic field, the ionosphere and the thermosphere related to geomagnetic disturbances are considered, several common features can be detected. Both the horizontal component of the geomagnetic field and the electron density in the F region of the ionosphere - the former at low and mid-latitudes, the latter at middle and high latitudes - show a storm time variation consisting of a lasting decrease of the field and the electron density, respectively (storm time is reckoned from the commencement of the storm) (Figs 1, 2) (Chapman and Bartels 1951, Matsushita 1959, Bencze 1965, Obayashi and Matuura 1971). In this paper the F region electron density is represented by the critical frequency of the F2

*Paper presented at the 10th KAPG Winter School on „Physical Processes in the Plasma of the Magnetosphere" and dedicated to the memory of the late professor Dr Ch-U Wagner

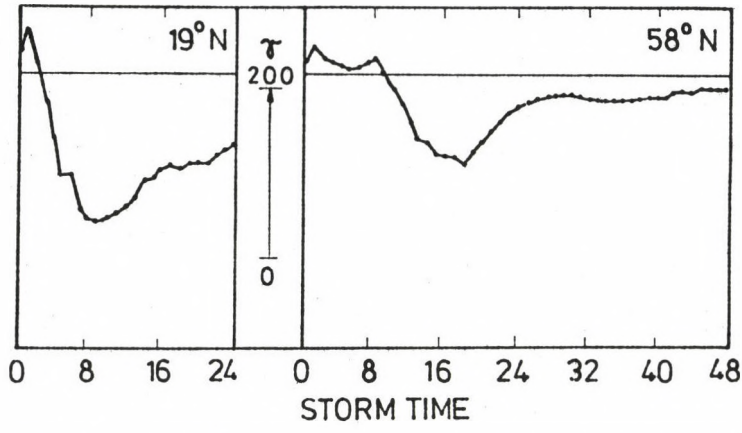


Fig. 1. Storm time variation in the horizontal component of the geomagnetic field at low and middle latitudes (after Chapman and Bartels 1951)

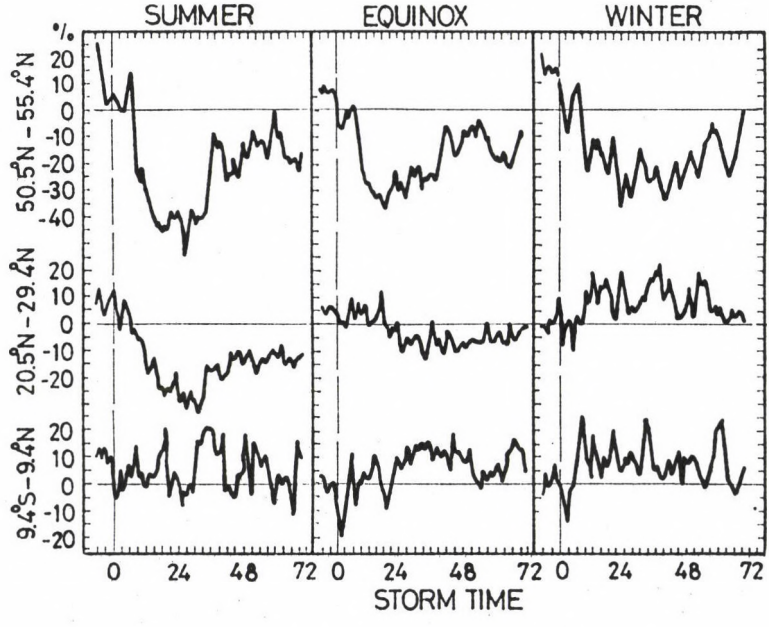


Fig. 2. Storm time variation of foF2 (proportional to the maximum electron density of the F2 layer) in different seasons at low, middle and high latitudes (after Matsushita 1959)

layer foF2 indicating the maximum electron density of the ionosphere. In the same latitude zone (30^{o}-60^{o}), where the

decrease of the electron density during geomagnetic disturbances (a negative ionospheric storm) can be observed, at the same time the composition of the neutral atmosphere is changed in the thermosphere (composition disturbance). The concentration of the molecular constituents (N_2, Ar) is increased as compared to that of the atomic constituents (O,He) (Fig.3) (Prölss 1980). At low latitudes the increase of the electron density in the F region during geomagnetic disturbances (positive ionospheric storm) (Fig. 2) coincides with the zone of unchanged composition in the thermosphere (Fig. 3). The relation between the F region electron density variation and the composition disturbance in the thermosphere during geomagnetic disturbances is confirmed by the corresponding change of the latitudinal extension of the decreased electron density and that of the composition disturbance with season, too. In summer the zone of decreased electron density is extended equatorward as compared to winter

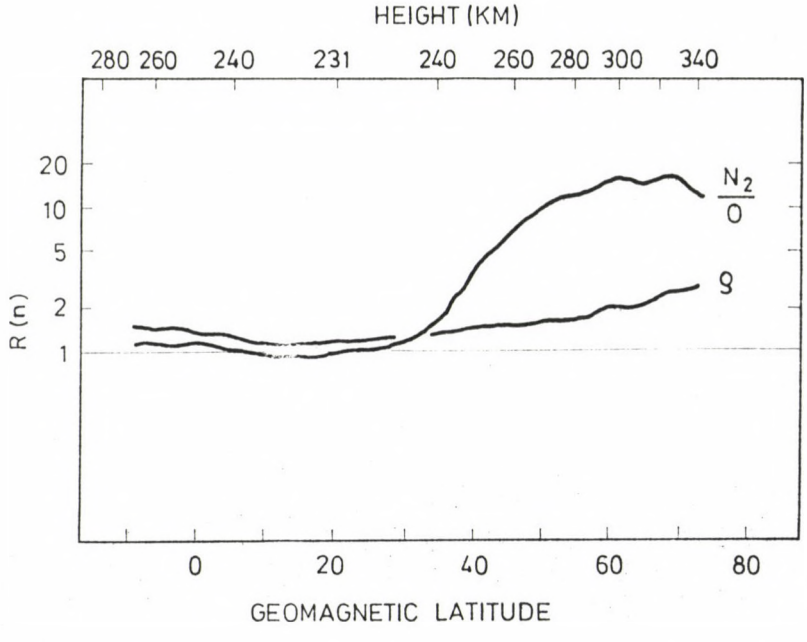


Fig. 3. Latitudinal variation of the composition disturbance in the thermosphere (after Prölss 1980)

in accordance with which the zone of the composition disturbance is also enlarged equatorward (Figs 2, 4) (Matsushita 1959, Prölss 1980).

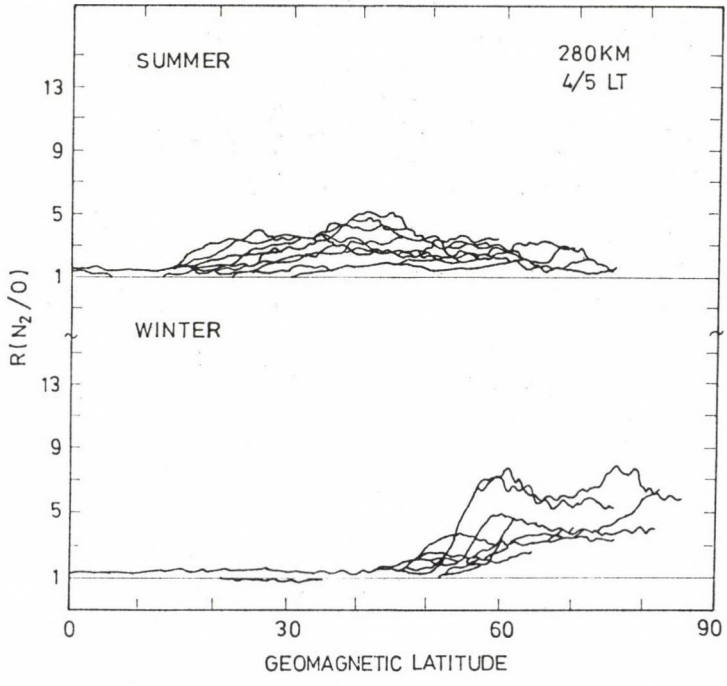


Fig. 4. Seasonal variation of the latitudinal extension of the composition disturbance in the thermosphere (after Prölss 1980)

Further similarities can be found in the disturbance daily variations (SD) of the above quantities. Both the disturbance daily variation of the geomagnetic horizontal component and that of the F region electron density changes similarly the maximum being shifted with increasing latitude from the forenoon hours at low latitudes to the afternoon hours at mid-latitudes (Figs 5, 6, 7) (Chapman and Bartels 1951, Matsushita 1959, Bencze 1965). In the density of the thermosphere a disturbance daily variation could also be traced at low latitudes the mean variation determined for the latitude zone $\pm 30^{o}$ having a maximum in the day-time hours (Fig. 8) (Illés-Almár et al. 1989). This picture is in agreement with the well known fact that at low latitudes the F region electron density

increases with geomagnetic activity. Here a morning maximum in the disturbance daily variation of the electron density corresponds to the morning maximum of the disturbance daily variation of the geomagnetic horizontal component (indicating the geomagnetic activity). As for the disturbance daily variation of the F region electron density at middle latitudes, the circumstance that at the time of maximum geomagnetic

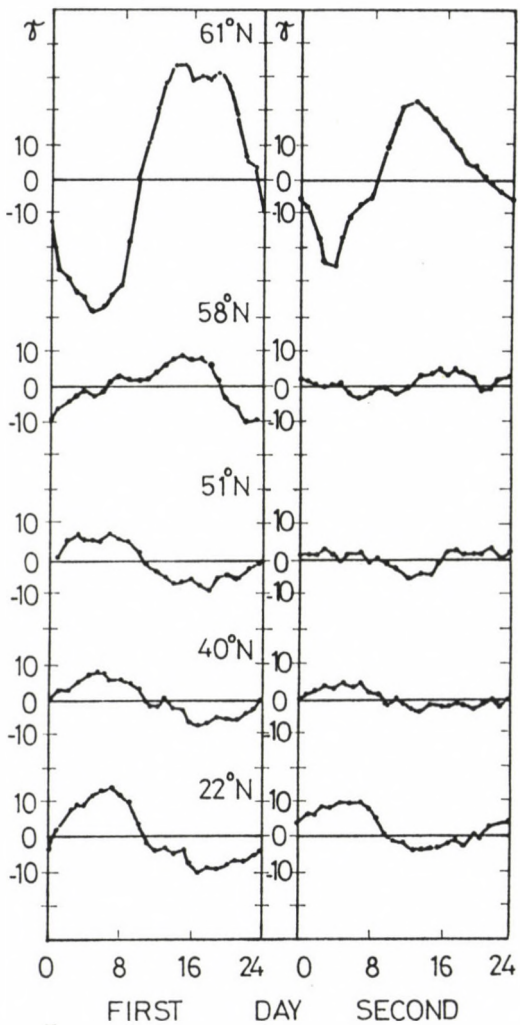


Fig. 5. Disturbance daily variation in the horizontal component of the geomagnetic field determined for the first and second days of the disturbance (after Chapman and Bartels 1951)

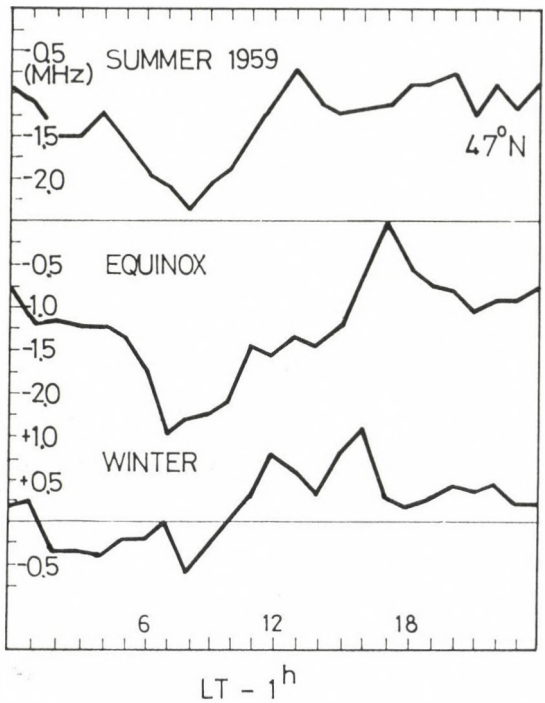


Fig. 6. Disturbance daily variation of foF2 (proportional to the maximum electron density of the F2 layer) in different seasons at middle latitude (Bencze 1965)

disturbance in the afternoon instead a minimum a maximum is observed, hints at a process differing from that producing the storm time variations. Similarly, since rising geomagnetic activity is accompanied by increased thermospheric density, enhanced day-time density corresponds to the morning maximum of the disturbance daily variation of the geomagnetic horizontal component (reminding that the disturbance daily variation of the thermospheric density considered here represents a spatial mean and the delay in the reaction of the thermosphere must also be taken into account).

DISCUSSION AND INTERPRETATION

As the background of the geomagnetic disturbances is fairly well known, it can support the interpretation of

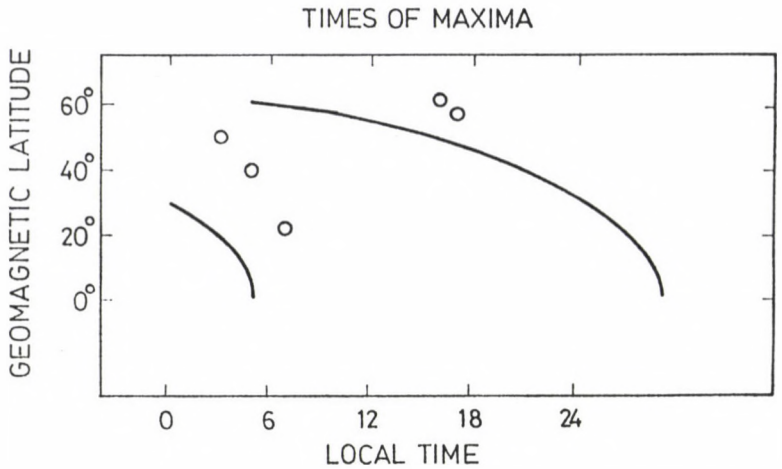


Fig. 7. The latitudinal variation of the phase (local time of the maxima) of the disturbance daily variation in foF2 (full line) (after Matsushita 1959) and in the geomagnetic horizontal component (circles) (after Chapman and Bartels 1951)

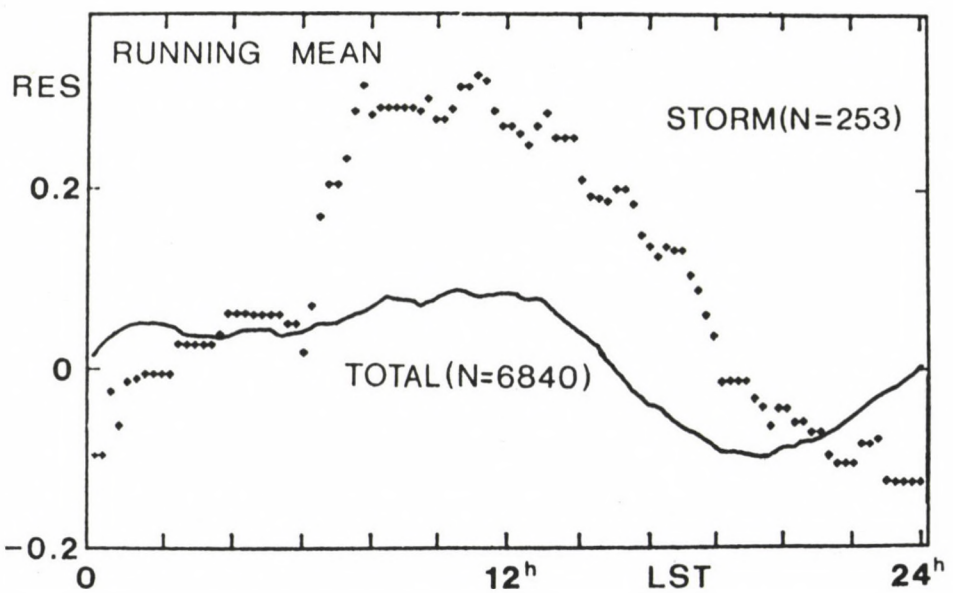


Fig. 8. Disturbance daily variation of the running mean residuals of the neutral density measured at about 400 km (deviations from the model values corrected for storm time variation determined fro the total material (full line) and for disturbed days (dots) (Illés-Almár et al. 1989)

ionospheric and thermospheric phenomena occurring during geomagnetic disturbances.

The lasting depression of the horizontal component of the geomagnetic field at low and mid-latitudes - called the main and recovery phases of a geomagnetic storm - are due to the ring current. It has been found that as a consequence of the interaction between the cross tail (convection) electric field generated by the solar wind and the geomagnetic field a plasma sheet is formed along the central plane of the tail. After earthward transport particles get from the plasma sheet both to the auroral ovals and the ring current. The particle injection to both of these regions during geomagnetic disturbances is due to the increased convection. The particles enter into the ring current belt as a result of gradient and curvature drifts in the geomagnetic field. The main phase depression of the geomagnetic field at low and mid-latitudes is the effect of the intensification of the ring current due to this particle injection from the plasma sheet. The decrease of the electron density in the F-region and the composition disturbance in the thermosphere at middle and high latitudes can be related to the effects of the particle injection from the plasma sheet to the auroral ovals occurring simultaneously with the injection to the ring current.

The modelling of thermospheric storm effects has shown that heating caused by precipitating electrons in the auroral zones establishes an upwelling of air rich in molecular constituents thus, increasing the concentration of molecular constituents as compared to that of the atomic components (Rishbeth et al. 1985). The effect of the upwelling, however, spreads over a latitude zone of only 5^{0} equatorward from the auroral ovals, the vertical diffusion being faster than the equatorward transport by the horizontal wind (Rishbeth et al., 1985). Thus, at lower latitudes - at middle latitudes - thermal expansion due to additional heating can assure the increased concentration of the molecular constituents as compared to that of the atomic constituents.

Hence at middle and high latitudes negative ionospheric

storms are connected with composition changes mentioned above. The increase of the concentration of the molecular components (N_2,O_2) as compared to that of the atomic oxygen can enhance the rate of dissociative recombination and a reduction of the electron density is observed in the F region.

At low latitudes thermal expansion caused by additional heating can also be necessary to compensate the effect of sinking of air rich in atomic components corresponding to the upwelling of air at high latitudes and establishing unchanged composition.

The development of the positive ionospheric storm is less clear. Considering the state of the thermosphere during geomagnetic disturbances, it can not be due to composition changes. Thus, it has been assumed that the increase of the F region electron density is related to the vertical transport of ionization caused by equatorward winds and/or zonal (eastward) electric fields. The circumstance that at low latitudes thermal expansion due to additional heating is necessary for the maintenance of unchanged composition, hints at the role of increased temperature. An increase of the temperature would also decrease the recombination rate and result in enhanced electron density.

The disturbance daily variation of the geomagnetic field is attributed partly to the asymmetry of the ring current, partly to auroral sources. Thus, an interpretation of the disturbance daily variation of the F region electron density and that of the thermospheric density might be attempted on the basis of this assumption. The ring current is asymmetric first of all from the point of view of its intensity the latter being greater at the night side than on the day side. The asymmetry of the intensity is due partly to the limited range of the longitudinal drift of the particles, partly to the partial ring current established by field aligned currents (region II currents) (Iijima et al. 1988). The ring current is also asymmetric considering its composition. It has been found that O^+ ions are the dominant ions on the dayside during the main and recovery phases of geomagnetic storms (Lundin et al. 1983).

The decay of the ring current is attributed partly to charge exchange between ring current ions (H^+,O^+,He^+) and neutral atoms in the geocorona (Dessler et al. 1961), partly to wave-particle interaction (Cornwall et al. 1970, Williams and Lyons 1974a,b, Williams et al. 1976, Joselyn and Lyons 1976, Lyons and Speiser 1982). As a result of the former process energetic neutral atoms are produced precipitating into the upper atmosphere at low latitudes, while due to the latter process energetic electrons leave the radiation belt (Fig. 9) (Torr 1983). The particles lose their energy in collisions with the ambient atmosphere. The wave-particle interaction can be considered as a third factor contributing to the asymmetry of the ring current. The efficiency of the wave-particle interaction as a loss process for the ring current is influenced namely by the interaction between the hot plasma of the ring current and the cold plasma of the plasmasphere (Cornwall et al. 1970, Kozyra 1988). The plasmasphere being larger on the dayside than on the nightside and having a bulge in the afternoon sector (Decreau et al. 1982, Gringauz 1983), the particle loss due to this process is not uniform along the belt. Thus, the particle loss can be larger on the dayside, especially in the afternoon than on the night side.

Making use of these findings, in case of the disturbance daily variation of the geomagnetic horizontal component the maximum in the morning hours at low latitudes hints at the relation to the asymmetry in the intensity of the ring current. As to the disturbance daily variation of the F-region electron density and the thermospheric density the compositional asymmetry (presumably the flux of energetic oxygen atoms produced by charge exchange being larger and resulting also in a greater energy deposition on the dayside in the thermosphere than protons) can also contribute to the development of the maximum in the morning-midday hours. The shift of the maximum from the forenoon to the afternoon hours with increasing latitude can be connected with the increasing importance of the wave-particle interaction in the afternoon as compared to charge-exchange in the production of energetic particles. An

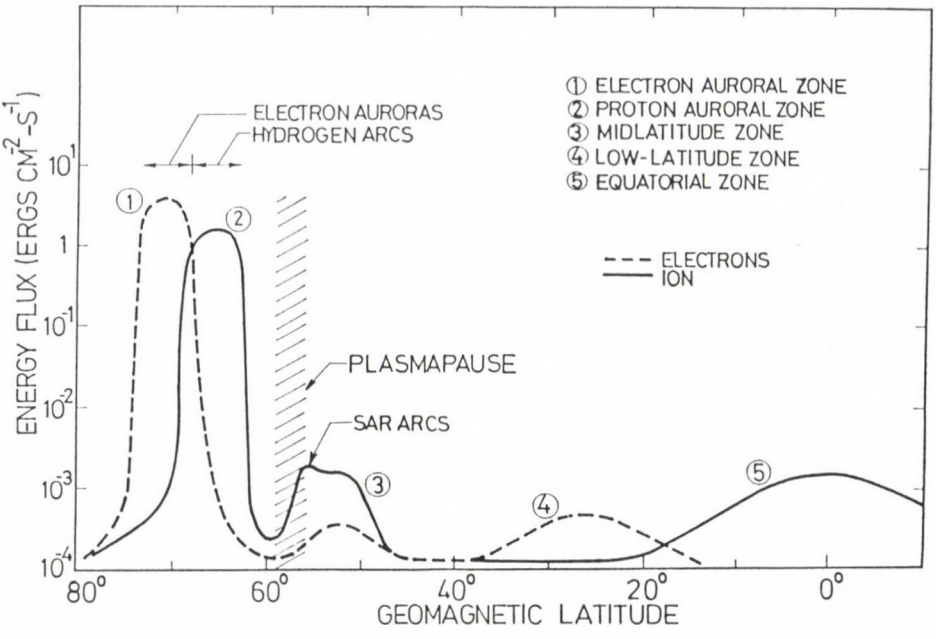


Fig. 9. The nighttime zones of precipitating electrons (dashed line) and ions (solid line) as a function of geomagnetic latitude, energy flux referring to 200 km, moderately disturbed conditions (Torr 1983)

argument in favour of this explanation can be the circumstance that at the time of the maximum geomagnetic disturbance a maximum of the disturbance daily variation of the F region electron density occurs also at mid-latitudes as at low latitudes, indicating a process analogous to the conditions at low latitudes, probable heating but not composition change.

CONCLUSIONS

Up to recent years the auroral oval has been considered as the place of energy input into the upper atmosphere during magnetospheric storms. The uncommon phenomena (aurora, geomagnetic, ionospheric disturbances) taking place in these cases at high latitudes directed the attention to the auroral zone. Since such phenomena of similar intensity were observed

nowhere else, the investigation of the terrestrial effects of the solar activity has been focussed on the polar regions. The observation of atmospheric emissions at low latitudes was probable the first observation, indicating that the storm energy has access to the upper atmosphere at these latitudes, too (Barbier 1947, Seaton 1956, Chapman 1957 a,b, Chamberlain 1961, Levasseur and Blamont 1973, Meier and Weller 1975, Paresce 1979, Tinsley 1979, Burnside et al. 1980, Meriwether and Walker 1980, Thomas 1980, Cazes and Emerich 1980, Tinsley and Burnside 1981, Tinsley et al. 1982, Rohrbaugh et al. 1983, Tinsley et al. 1984, Ishimoto et al. 1986, Tinsley et al. 1986, Abreu et al. 1986). At first it was not quite clear how these emissions are produced. A systematic study of atmospheric emissions has enabled the determination of the energy source of these emissions, particles transferring the energy stored in the magnetosphere (Meier and Weller 1975, Tinsley 1979, Tinsley and Burnside 1981, Tinsley et al. 1982, Rohrbaugh et al. 1983, Ishimoto et al. 1986, Tinsley et al. 1986, Abreu et al. 1986). Ionospheric observations have also shown that at low latitudes there is an excess ionization (intermediate layer formation) during geomagnetic disturbances hinting at an additional energy source (Rowe 1974, Goldberg 1974, Kelley et al. 1977). Increased electron temperature has also been observed in the recovery phase at middle latitudes (Fig. 10) (Wagner et al. 1986). The increase of the neutral temperature at low latitudes deduced from satellite drag data could partly be explained by particle (energetic hydrogen atoms) precipitation from the storm ring current (Prölss et al. 1973). However, not only these effects of particle precipitation, but also the particles themselves, e.g. quasi-trapped particles (2-20 keV electrons or protons) have been observed by rocket measurements above 200 km at low latitudes (Goldberg 1974, Kelley et al. 1977, Guzik et al. 1989) supporting the idea of multiple charge-exchange between radiation belt ions and neutrals (Scholer et al. 1975).

Thus, there are many observations indicating the access of energy into the upper atmosphere during geomagnetic disturbances also at low latitudes. Consequently, in addition

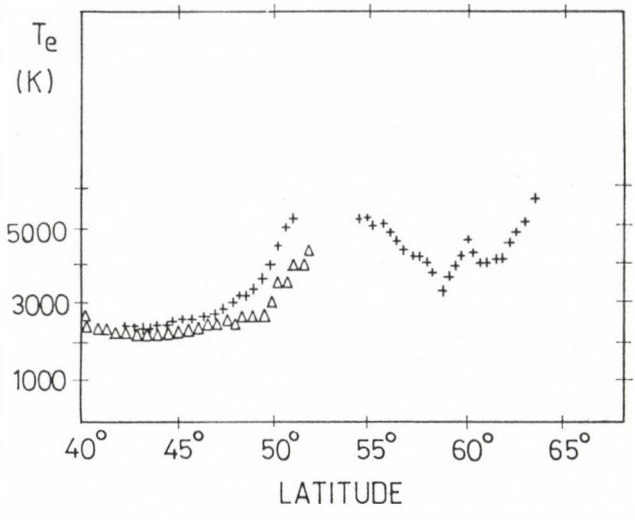


Fig. 10. Latitudinal variation of the electron temperature measured by the satellite IK-10 on 29.1.1974 (+) and 31.1.1974 (Δ) at 5 MLT in about 1000 km altitude (Wagner et al. 1986)

to the auroral processes the particle precipitation at low and mid-latitudes should also be taken into account. As it has been shown, in disturbed periods the different typs of precipitating particles can contribute to the variations of the geomagnetic field, the ionosphere and the thermosphere.

REFERENCES

Abreu V I, Eastes R W, Yee J H, Solomon S C, Chakrabarti 1986: J. Geophys. Res., 91, 11365-11368.

Barbier D 1947: Ann. Geophys., 3, 227-228.

Bencze P 1965: Thesis, Sopron

Burnside R G, Meriwether J W Jr, Walker J C G 1980: J. Geophys. Res., 85, 767-773.

Cazes S, Emerich C 1980: J. Geophys. Res., 85, 6049-6054.

Chamberlain J W 1961: Physics of the Aurora and Airglow. Academic Press, New York

Chapman S 1957a: Nature, 179, 7-11.

Chapman S 1957b: Ann. IGY. 4(II), 25-40.

Chapman S and Bartels J 1951: Geomagnetism. University Press, Oxford

Cornwall J M, Coroniti F V, Thorne R M 1970: J. Geophys. Res., 75, 4699-4709.

Decreau P M E, Beghin C, Parrot M 1982: J. Geophys. Res., 87, 695-712.

Dessler A J, Hanson W B, Parker E N 1961: J. Geophys. Res., 66, 3631-3637.

Goldberg R A 1974: J. Geophys. Res., 79, 5299-5303.

Gringauz K 1983: Space Sci. Rev., 34, 245-257.

Guzik T G, Miah M H, Mitchell J W, Wefel J P 1989: J. Geophys. Res., 94, 145-150.

Iijima T, Potemra T A, Zanetti L J 1988: Paper presented at COSPAR Meeting, Helsinki

Illés-Almár E, Almár I, Bencze P, Horváth A 1989 (in press)

Ishimoto M, Torr M R, Richards P G, Torr D G 1986: J. Geophys. Res., 91, 5793-5802.

Joselyn J A, Lyons L R 1976: J. Geophys. Res., 81, 2275-2282.

Kelley M C, Schwartz W E, Tayan Y, Torbert R 1977: J. Atmos. Terr. Phys., 39, 1263-1268.

Kozyra J U 1988: Paper presented at COSPAR Meeting, Helsinki

Levasseur A C, Blamont J E 1973: J. Geophys. Res., 78, 3881-3893.

Lundin R, Hultquist B, Pissarenko N, Zacharov A 1983: In: Energetic Ion Composition in the Earth's Magnetosphere, Terrapub, Tokyo, 307-351

Lyons L R, Speiser T 1982: J. Geophys. Res., 87, 2276-2286.

Matsushita S 1959: J. Geophys. Res., 64, 305-321.

Meier R R, Weller C S 1975: J. Geophys. Res., 80, 2813-2818.

Meriwether J W Jr, Walker J C G 1980: J. Geophys. Res., 85, 1279-1284.

Obayashi T, Matuura N 1971: In: Solar Terrestrial Physics/1970, Reidel, Dordrecht, Holland, 199-211.

Paresce F 1979: J. Geophys. Res., 84, 4409-4412.

Prölss G W 1980: Rev. Geophys. Space Phys., 18, 183-203.

Prölss G W, Najita K, Yuen P C 1973: J. Atmos. Terr. Phys., 35, 1889-1901.

Rishbeth H, Gordon R, Rees D, Fuller-Rowell T J 1985: Planet. Space Sci., 33, 1283-1301.

Rohrbaugh R P, Tinsley B A, Rassoul H, Sahai Y, Teixeira N R, Tull R G, Doss D R, Cochran A L, Cochran W D, Barker E S 1983: J. Geophys. Res., 88, 6317-6330.

Rowe J F 1974: Radio Si, 9, 175-182.

Scholer M, Hovestadt D, Morfill G 1975: J. Geophys. Res., 80, 80-85.

Seaton M J 1956: In: The Airglow and the Aurorae. Pergamon Press, London, 225-243.

Thomas G E 1980: J. Geophys. Res., 85, 6055-6062.

Tinsley B A 1979: J. Geophys. Res., 84, 1855-

Tinsley B A, Burnside R G 1981: Geophys. Res. Letters, 8, 87-90.

Tinsley B A, Rohrbaugh R P, Sahai Y, Teixeira N R 1982: Geophys. Res. Letters, 9, 543-546.

Tinsley B A, Rohrbaugh R P, Rassoul H, Barker E S, Cochran A L, Cochran W D, Wills B J, Wills D W, Slater D 1984: Geophys. Res. Letters, 11, 572-575.

Tinsley B A, Rohrbaugh R P, Rassoul H, Sahai Y, Teixeira N R, Slater D 1986: J. Geophys. Res., 91, 11257-11269.

Torr D G 1983: Rev. Geophys. Space Phys., 21, 245-262.

Wagner C-U, Förster M, Johanning G 1986: Paper presented at COSPAR Meeting, Toulouse

Williams D J, Lyons L R 1974a: J. Geophys. Res., 79, 4195-4207.

Williams D J, Lyons L R 1974b: J. Geophys. Res., 79, 4791-4798.

Williams D J, Hernandez G, Lyons L R 1976: J. Geophys. Res., 81, 608-616.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 253-262 (1991)

GEOMAGNETIC PULSATIONS AT LOW- AND MID-LATITUDES

J Verő[1], L Holló[1], B P Singh[2]

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

[2]Indian Institute of Geomagnetism, Colaba, Bombay, 400 005 India

[Manuscript received August 30, 1990]

One day of geomagnetic pulsation records have been compared at the Indian magnetotelluric station Dharampur with those of the mid-latitude stations Nagycenk and Niemegk. It was found that at the low latitude station, the pulsation activity is different in its spectrum from the mid-latitude ones, and the differences are most substantial if there is no latitude dependence of the period at mid-latitudes. In such a case even impulsive events may be absent at one of the locations and/or the frequency contents may also be different. In the case of latitude dependent events, there is some similarity between low- and mid-latitude records.

Keywords: dynamic spectrum; geomagnetic pulsations; L-value

1. INTRODUCTION

Low-latitude pulsations have been studied almost from the beginning of pulsation research. Kunetz and Richard (1952) used records from Madagascar and Venezuela when he discovered the simultaneous world-wide appearance of pulsations. Hutton (1962) was the first to show that at an equatorial station pulsations have daily variations with significant maxima in night-time hours. Saito (1961) also included low-latitude stations in his study on the daily variation of pulsation amplitudes.

Models based on the field line resonance predicted very short periods at low latitudes (Jacobs and Watanabe 1962). Later there were calculations which agreed better with experimental data by using non-homogeneous plasma distributions (Jacobs and Watanabe 1963).

The equatorial enhancement of at least a part of the pulsations has been studied by Sastry et al. (1983).

A significant change occurred in the interpretation of low-latitude pulsations when the possibility of the co-existence of two (or more) different types of pulsations has been established (see e.g. Yumoto 1986). Thus waves of solar wind origin may penetrate the magnetosphere to rather low latitudes on the one hand, conserving their original properties (frequency), on the other hand field line resonances may be excited in the magnetosphere triggering signals of frequencies changing with L-value (or geomagnetic latitude). A comparison has shown in the sixties that instead of the periods 20-25 s (at $L \sim 2$), at $L \sim 1.22$ (Tamanrasset) periods of 8 to 12 s may occur. In other cases (Cz Miletits et al. 1990) the period did not change at latitudes below $L \sim 2.0$ to 2.2.

Japanese studies (Yumoto et al. 1985, Yumoto 1985) included stations at L-values of 1.15; 1.3 and 1.8 and data of the year 1981. No mention is made on any systematic variation of the frequency with the L-value, and the functions $T=f(B)$ do not differ considerably even from those at $L \sim 5.6$, moreover the scatters are similar. Yumoto (1986) presented a model for low-latitude pulsations based on stations at $L \sim 1.06$-1.57, but without a study of the periods.

Figure 1 gives a scheme of the distribution of periods vs. L-value (latitude) for the two types of pulsations at two values of the interplanetary magnetic field.

2. RECORDS USED IN THIS STUDY

The coverage of the present study was determined by the fact that digital records were available from the station Dharampur, DHA ($L \sim 1.04$, near Bombay) for a day April 10, 1988 12 00 UT till April 11, 12 00 UT for the 5 EM components but in this study only records of the component H_x were used. These data were compared to records of the component E_y in the observatory Nagycenk, NCK ($L \sim 2.14$) and H_x in Niemegk, NGK ($L \sim 2.63$). The longitude at these observatories differs by about 55° from that of DAA. The sampling rate was generally 3 s.

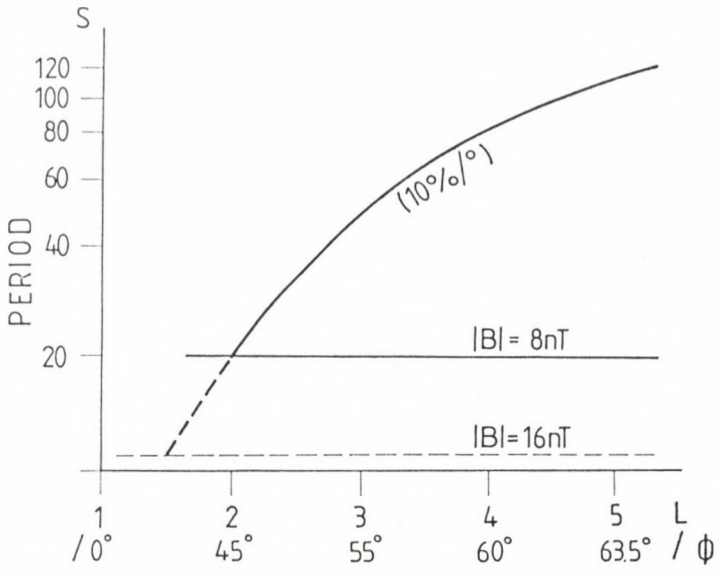


Fig. 1. A scheme of the function period vs. L-value for the two pulsation types; the period of the magnetospheric type changes here at a rate of 10 percent for a degree of latitude. Full line represents the situation for an interplanetary magnetic field of 8nT, dotted one for that of 16 nT

The most interesting part of the records is April 11, 0353 - 0843 UT, when all stations were in the sunlit part of the Earth (during night-time hours, pulsation activity was low at DHA, too). This five hours-section was studied in more detail.

High-resolution dynamic spectra were computed from these five hours long sections of the records, and these dynamic spectra were used for the comparison of the characteristic periods at the three sites, further impulsive events were selected from them and then compared. Average amplitudes over these sections were also computed and used as amplitude spectra.

3. RESULTS

3.1 Pulsation activity in the investigated interval

During the five hours of the investigated interval (April 11, 0353 - 0843 UT) three sections could be separated of different pulsation activities.

1) 0353 - 0638: strong pulsation activity which ends with a switch-off. The period is the same at the observatories NCK and NGK, it means that this event is of a latitude (or L-value) independent period.
2) 0638 - 0801: very low pulsation activity at all the three stations
3) 0801 - 0843: switch-on, followed by strong pulsation activity. The spectrum at NCK is similar to that of the first part, in NGK, the period is longer (peak at NCK, 23 s, at NGK 36 s, i.e. the period changes by 9 percent for a degree of latitude, being a normal value).

Figure 2 shows the dynamic spectrum of a half hour section of the first pulsation event (1) from DHA and NCK; Fig. 3 the same from the second pulsation event (3); here the NGK dynamic spectrum is also presented for comparison.

A long-scale comparison of the pulsation activities at DHA and NCK (and NGK) shows that the time evolution of the pulsation activity is similar at all stations, mainly due to the mentioned switch-off and switch-on which are present at each of the three stations. The amplitude spectra are, however, different at the three stations (Fig.4): at NGK, there is a narrow peak at about 30 s with $f_0/\Delta f_{0.7}= 2$. Here f_0 is the frequency corresponding to the amplitude peak, A_0, and $f_{0.7}$ that corresponding to 0.7 A_0. At NCK, there are two peaks, one at 20 s, with similarly $f_0/\Delta f_{0.7}=2$, and a second one at 100 s. The 30 s NGK and the 20 s NCK peaks correspond to the local field line resonant periods. At DHA, a flat maximum covers the full Pc 3-4 range (20- to 120 s) with $f_0/\Delta f_{0.7}=0.85$. The difference of the spectra is also reflected in the fact that at NCK, 50 percent of all short-lived events (as e.g. short

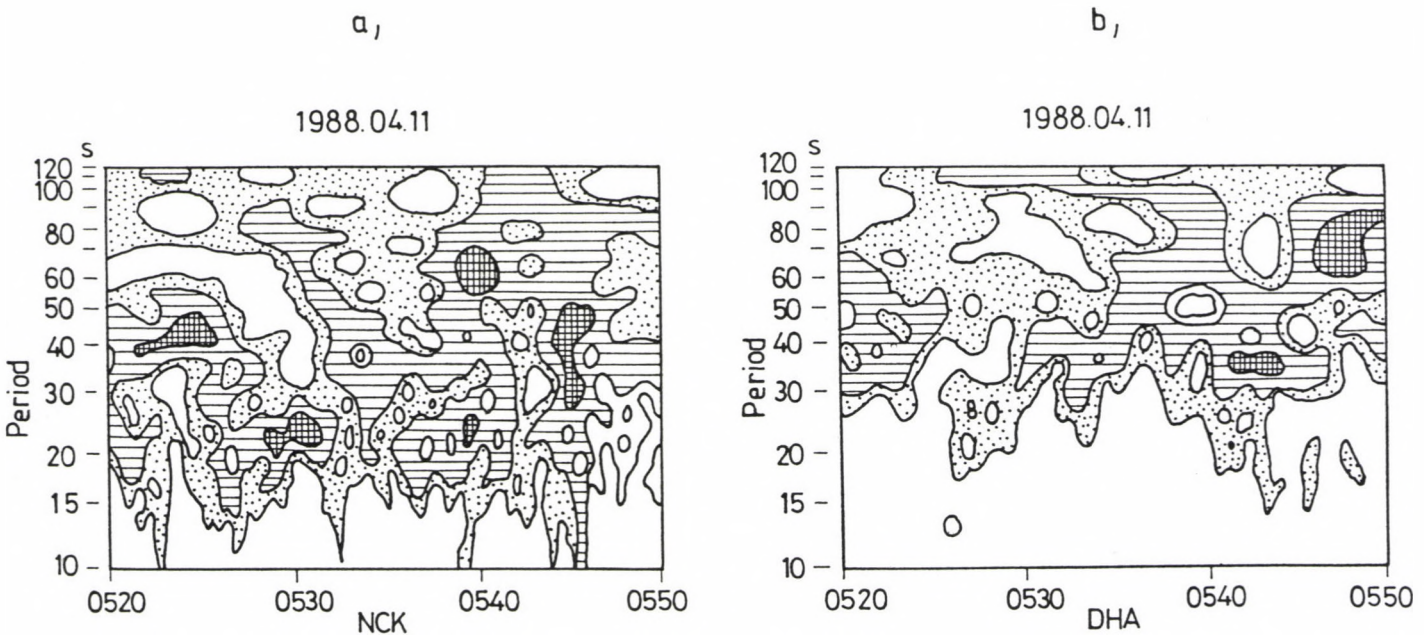


Fig. 2. Dynamic spectra of NCK and DHA for the first event on April 11, 1988. On the vertical axis, periods are given (no frequencies). Isolines are plotted for amplitudes differing by a factor of 2; the three highest amplitude values are indicated

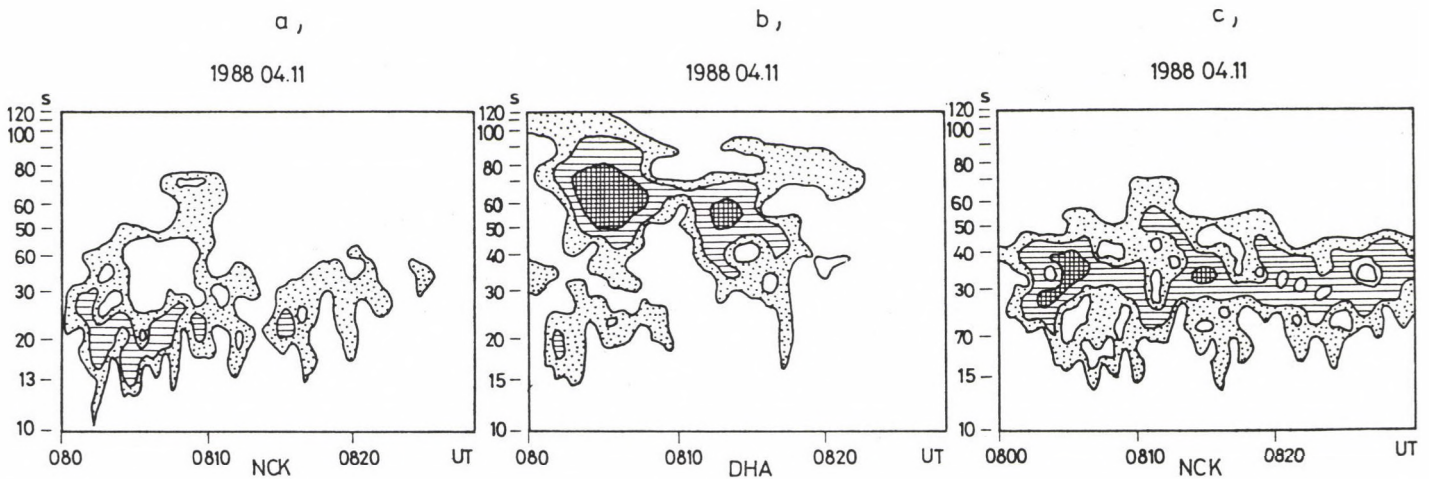


Fig. 3. Dynamic spectra of NCK, DHA and NGK for the second event on April 11, 1988 (with a period changing with latitude). Isolines are similar to those in Fig. 2

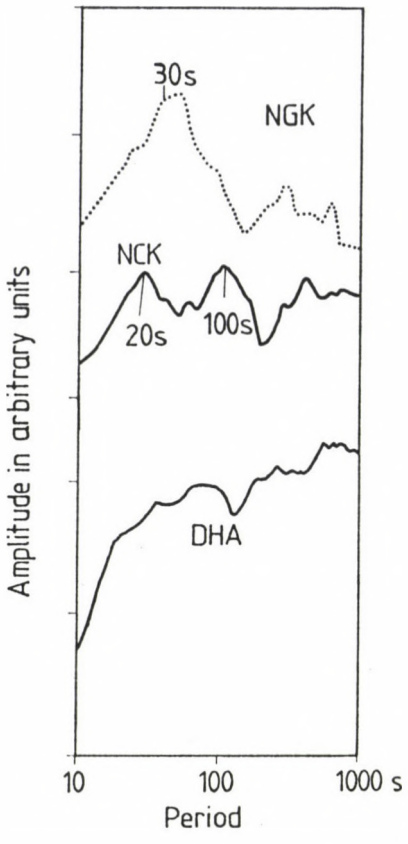


Fig. 4. Amplitude spectra of the full 5h section investigated at NCK and DHA, and spectrum of the event shown in Fig. 3c (0801-0843 v.) at NGK. NCK spectrum is electric (Ey), the other two magnetic (Hx)

impulses, switches of short duration, amplitude peaks etc.) occur in the 20-25 s interval, at DHA, the same amount occurs in the 15 to 35 s range.

As the pulsation activity is determined by the switch-off at 0638 and by the switch-on at 0801, at all the three stations, the correlation between peak amplitudes is high. This high correlation (similarity) does not mean a similarity of the corresponding records in detail, as the following comparison will show.

During the five hours interval, a total of about 50- to 60 short-lived events were counted in the DHA and NCK records.

From them, about 30 percent coincide in time at the two stations what is not significantly more than expected in a random distribution. It is, however, worth mentioning that in the first part, only 15 to 20 percent of these events coincided, whereas in the third part, about 50 percent. There are cases when a most characteristic short-lived event at a mid-latitude station is absent at low latitudes, in other cases the spectra are strongly different as in the event presented in Fig. 5, when there is no increase at DHA in the range of the NCK-impulse (15-40 s).

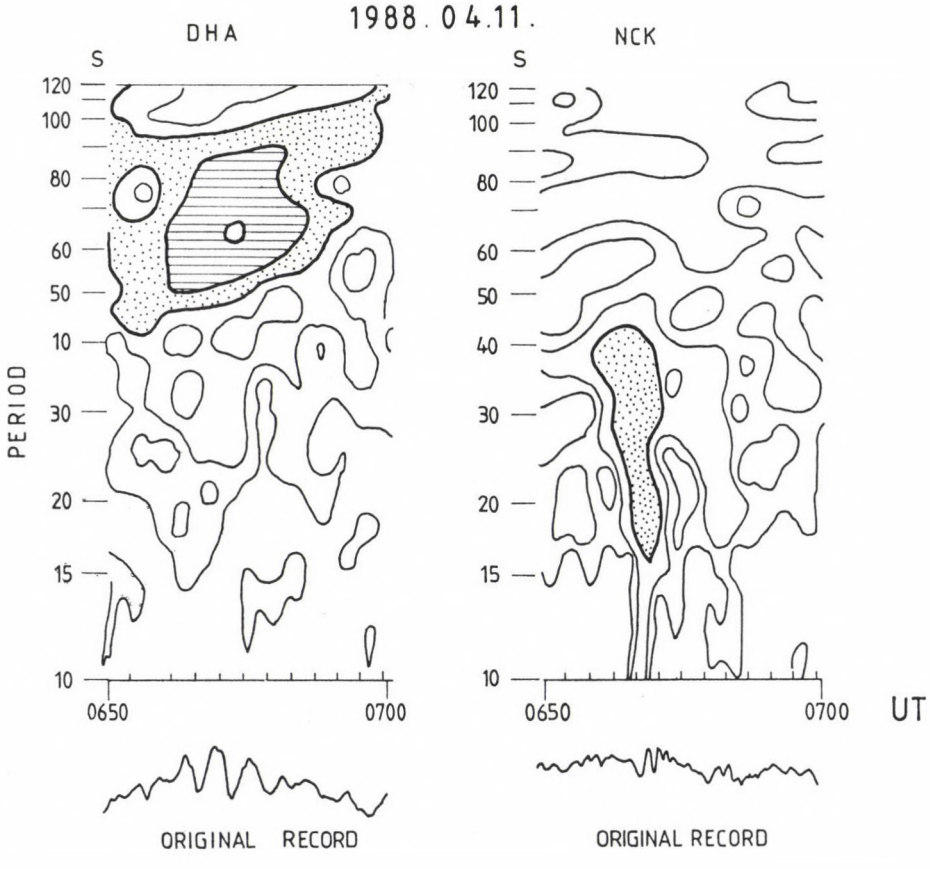


Fig. 5. Dynamic spectra of a section when an impulse at NCK has no counterpart at DHA in the same period range

In the pulsation event with L-dependent periods (3), the activities are in a rather complicated connection. Correlations between NCK and DHA activities are generally higher than between NCK and NGK (a low correlation between short section NCK and NGK pulsation amplitudes is normal for a pulsation event with L-dependent period). The correlation between the amplitudes at NCK and DHA is high for the shortest period ranges (10 to 15 s), further for the NCK 20-25 s range (the local field line resonance) with a rather wide range of periods at DHA. Both for the short period pulsations and for the NCK 20-25 s range, the correlation coefficient reaches a value of 0.7, being significant at the 1 percent level. Pc 4 activities are not correlated.

4. CONCLUSION

In the events investigated in the present study, pulsations with L-independent periods (compressional waves of solar wind origin) did not reach the equatorial region. This event compared with a number of others (e.g. in Cz Miletits et al. 1990) shows that the equatorial edge of the penetration region of the compressional waves may shift quite significantly.

In the case of field line resonances, however, equatorial stations „see" at least parts of the pulsations at different L-values (of different periods), thus in such events low- and mid-latitude pulsation activities are better correlated, even if the spectra are different. It seems that in the equatorial zone due to the mixture of signals from different L-values, spectra are much flatter than at somewhat higher latitudes, where field line resonances create narrow spectral peaks. In the equatorial zone, longer period activity is stronger than at somewhat higher latitudes. Nevertheless, the dynamic spectra of three stations show that in spite of the similarity of the peak amplitude periods at NGK and DHA ($L \sim 2.35$ and 1.04), the pulsation activities are less correlated than between DHA and NCK ($L \sim 2.14$) in spite of the fact that the periods are

different at the latter stations.

A comparison of these results with Yumoto's (1986) model shows a possibility for their interpretation. Yumoto supposed that low-latitude pulsations are caused by an ionospheric Pedersen eddy current induced by the inductive electric field of fast magnetosonic source waves plus (in order to explain the observed elliptic polarization instead of the theoretically expected circular one) a Hall eddy current by a standing field-line oscillation excited near $L \sim 2.0$. Now if the latter, the Hall eddy current may reach the equatorial region from higher L-shells, too, the appearance of the primary source wave (the Pedersen eddy current) is no more prerequisite of the low-latitude pulsation activity, the observation in the present paper can be well explained.

REFERENCES

Cz Miletits J, Verő J, Szendrői J, Ivanova P, Best A, Kivinen M 1990: Planet. Space Sci., 38, 85-95.

Hutton V R S 1962: Nature, 195, 269-270.

Jacobs J A, Watanabe T 1962: J. atmos. terr. Phys., 24, 413-434.

Jacobs J A, Watanabe T 1963: Planet. Space Sci., 11, 869.

Kunetz G, Richard H 1952: Atti del settimo convegno nazionale del metano e del petrolio. I. Taormina, 511.

Saito T 1961: Sci. Rept. Tohoku Univ., Ser 5, 13, 53-61.

Sastry T S, Sarma Y S, Sarma S V S, Sanker-Narayan P V 1983: J. atmos. terr. Phys., 45, 733-741.

Yumoto K 1985: Planet. Space Sci., 33, 239-249.

Yumoto K 1986: Planet. Space Sci., 34, 1329-1334.

Yumoto K, Saito T, Akasofu S, Tsurutani B T, Smith E J 1985: J. geophys. Res., 90, No. A7, 6439-6450.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 263-272 (1991)

GROUND-BASED MONITORING OF IMF MAGNITUDE

T A Plyasova-Bakounina[1] and J W Münch[2]

[1]Institut Fiziki Zemli, Moscow, USSR
[2]Siegen, University, FRG

[Manuscript received September 20, 1990]

A method of ground-based monitoring of B, using geomagnetic pulsations Pc2-4 is presented in this paper. The method we used consists of computing the coherence function of the signals of two stations. This method is simple and requires the following steps:

1. Two stations (or more) equipped with identical instruments (digital fluxgate magnetometers) located at the same latitude and spread in longitude at a distance $> 20^{o}$ are chosen for the measurement of geomagnetic pulsations in the frequency range of 10-200 mHz. Recording of the two horizontal components H and D of the magnetic field should be organized from dawn to afternoon.

2. The modulus of the coherence function is computed for the two stations for the H- and D-components separately for selected time-intervals.

3. The frequency interval which corresponds to the width of the maximum of the coherence-function above a level of 0.8, is determined for H and D: ΔF_H, ΔF_D respectively.

4. The common part ΔF_c of two the frequency bands ΔF_H and ΔF_D is evaluated.

5. B is calculated, substituting ΔF_c into Eq. (1).

This method was tested with data of the stations Lindau and Sverdlovsk ($\Delta\lambda=38^{o}$). B_{calc} matched with the averaged value of the measured B_{real}.

The error of B_{calc} is composed of the error of F and the error of C (Eq. 3). The error of C is ~ 5 %, as numerous statistical investigations of the coupling coefficient in the F-B relationship have demonstrated.

The accuracy of the determination of B can be improved by increasing the accuracy of the determination of F. This can be achieved in three ways:

1. by improving the resolution of the frequency in the calculation of the power spectrum and the coherence function, choosing the optimal parameters of the computer program,

2. by undertaking a massive statistical study for improving of the methodics, partly as the basis for choosing the level of significant coherence function,

3. by mastering the present method by placing stations along the meridian.

Keywords: field line resonance; geomagnetic pulsations; ground-based methods; interplanetary magnetic field; spectrum

INTRODUCTION

The magnitude of the Interplanetary Magnetic Field B is an important parameter which determines the input of energy from the solar wind into the magnetosphere (Akasofu 1981, Pudovkin and Semenov 1986).

There exists a diagnostic tool to measure B by means of micropulsations, using the one station data - „B-index" (Gul'elmi and Bolshakova 1973). It is based on the idea of the external origin of Pc2-4 and on the observation that the following relationship exists:

$$T = 160/B \text{ or } F = CxB \qquad (1)$$

where C = 6.25 mH/nT (see Troitskaya et al. 1971) and T and F are the period and frequency of the pulsations, respectively.

The disadvantage of this method is that the strong correlation of equation (1) is not fully fullfilled with the observations taken into account: the percentage of success is low (only 33 %), if the method cited by Gul'elmi and Bolshakova (1973) is used. The attempt to change the method has not lead to a considerable success (55 % of success of „Advanced B-index", see in detail Russel and Fleming 1976).

The failure of the B-index occurred because pulsations in the frequency range of Pc2-4 are waves, which are generated by different sources, not only external ones as it was suggested in Troitskaya et al. (1971) and other paper (see review Odera 1986), but mainly by internal ones: K-H instability at the magnetopause, plasmapause, drift and bounce instabilities and others (Southwood 1974, Chen and Hasegawa 1974, Siebert 1964).

The cloud of points out of the main dependence T = 160/B, which one can see at Fig. 1 in Troitskaya et al. (1971) and Fig. 2 in Gul'elmi (1974) show that the correlation between B and T differs with the source of the pulsations considered.

Hence, the problem of B-diagnostic results in the problem of the separation of the waves of external origin from the other ones. The report of Plyasova-Bakounina et al. in Budapest

EGU-meeting (published Plyasova-Bakounina et al. 1983) demonstrated that the same frequencies of Pc2-4, which are observed at two stations (Borok and Hartland) located over more than 20^{o}-distance apart (what we believe is the scale of the correlation of the local wave fields of external origin), are highly correlated with B. Hence, they made a conclusion, that the solar wind-controlled Pc have a global wave-field. This result was confirmed by studies of Green et al. (1983) (also Borok and Hartland data) and Odera (1986), who also showed, that the correlation between F and B is best when the latitudinal distance between stations, which are observing the same frequencies F, is more than 20^{o} for Pc3 and more than 60^{o} for Pc4 pulsation.

The following simple arguments support a global wave-field of solar wind controlled Pc. The external source - the upstream waves - can be found on the whole day-side before the bow shock (Fairfield 1969). Its projection on the ground into the region of the day-polar cusp creates in the upper ionosphere a „secondary" source of several ten degrees size (Plyasova-Bakounina et al. 1986, Lanzerotti et al. 1986, Engebretson et al. 1986) i.e. a current, which probably closing through the middle latitudes, forms a global wave-field there.

The size of the region of the internal sources is rather indefinite, but it seems that it is not larger than the resonance region and at any rate not larger than the external source region.

Certainly, both external pulsations and internal ones can amplify in the corresponding resonance tubes, but the intensity of the external pulsations can be high out of the resonance region, while the intensity of the internal ones seems to be limited by the resonance region.

So, we believe that the size of the resonance region can be that critical size which divides wave-fields of the pulsations into two groups: global and local ones.

The sizes of such a resonance region on the ground according to the satellite data (see Hughes et al. 1978, Hughes 1980) are $5\text{-}10^{o}$ in latitude and $< 20^{o}$ in longitude for Pc3 and

a little more for Pc4.

The aim of the present work is to work out methodically a ground-based method of monitoring of B which would be simple, easy to reproduce, provided a satisfactory accuracy of the definition of B and an uninterrupted set of the predicted values of B are available.

DATA AND ANALYSIS

The pulsation data of a pair of stations were processed Lindau ($\lambda = 10^{o}$, $\Phi = 53^{o}$) and Sverdlovsk ($\lambda = 48^{o}$, $\Phi = 52^{o}$). The data were recorded during the Soviet-German experiment campaign of 1974 (August-October) over a network of middle latitude stations. The longitudinal distance between the stations is 38^{o}. The stations were equipped with typical measuring instruments (fluxgate magnetometers) and recording instruments (digital tape recorders). The horizontal H- and D-components of the magnetic field were recorded in the frequency range 0-1000 mHz (Münch et al. 1975). The magnetograms were filtered in the frequency range of Pc2-4 pulsations, 10-200 mHz.

The intensity of the wave-field in this frequency range was analysed simultaneously at both stations from dawn to afternoon 06.00-14.00 UT. No special selection of the analysed intervals has been done. 20 half-hour intervals from 05.09 till 08.09 1974 were analysed. The list of events is shown in Table I. The data were processed in the following way (index "1" denoting Lindau, index "2" denoting Sverdlovsk):

1. Power spectra were computed for the H- and D-components in Lindau and in Sverdlovsk: $S_1^H(F)$, $S_1^D(F)$, $S_2^H(F)$, $S_2^D(F)$.
2. Modulus of the coherence function γ_H between two sets of data of the H-components was calculated.

$$\gamma_H(F) = \sqrt{\frac{S_{12}^H(F)^2}{S_1^H(F) \times S_2^H(F)}} \qquad (2)$$

where S_{12}^H is the cross-spectrum between the H-components of

Table I

No.	Data Day month year	Time UT	ΔF_H (mHz)	ΔF_D (mHz)	ΔF_C (mHz)	B_{calc}	B_{real}
1.	5.09.74	06.00-06.30	40-37	37	37	5.9	5.9±4
2.		06.30-07.00	38-33	36	36	5.8	5.9±4
3.		07.00-07.30	34	100	-	-	5.6±4
4.		07.30-08.00	71	55	-	-	5.6±4
5.	6.09.74	06.00-06.30	30	43	-	-	4.4±3
6.		06.30-07.00	27.8		27.8	4.4	4.4±3
7.		07.00-07.30	28-25	37-28	28	4.5	4.5±3
8.		07.30-08.00	25-33	31-24	25	4.0	4.5±3
9.	7.09.74	06.00-06.30	25	20-14	-	-	5.5±3
10.		06.30-07.00	72	59	-	-	5.5±3
11.		07.00-07.30	20	25;14	-	-	3.6±3
12.		07.30-08.00	20	20;14	20	3.2	3.6±3
13.		08.00-08.30	40-31	33	33	5.3	5.3±4
14.		08.30-09.00	25-20	29	-	-	5.3±4
15.	8.09.74	06.00-06.30	34-27	27	27	4.3	4.3±3
16.		06.30-07.00	32-27	27	27	4.3	4.3±3
17.		07.00-07.30	29	46-29	29	4.7	4.5±3
18.		07.30-08.00	-	27	-	-	4.5±3
19.		08.00-08.30	27	14	-	-	5.3±1
20.		08.30-09.00	22-18	25	-	-	5.3±1

stations 1 and 2.

3. Moduli of the coherence function γ_D of two sets of the data of the D-components were also calculated.

Fig. 1 shows γ_H and γ_D as a function of frequncy F. γ changes from 0 to 1.

Taking into account that the observed coefficient of correlation between F and B is > 0.8 (see for instance Gul'elmi et al. 1973), i.e. rather high, we have adopted this value for the estimation the optimal significant level of γ.

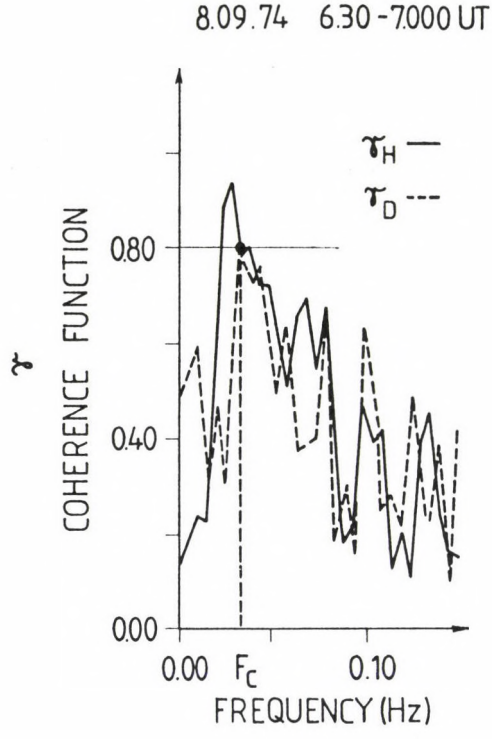


Fig. 1.

We suppose that if γ reaches the level 0.8, the corresponding frequencies can be considered conditionally as coherent ones.

Calculating the frequency width of the maxima of the coherence funcion at the level 0.8, we define ΔF_H and ΔF_D. Then we define the common part for these two frequency ranges - ΔF_C.

Putting ΔF_C in Eq. (1), the B-value is calculated.

ΔF_H, ΔF_D, ΔF_C and B_{calc} are listed in Table I.

DISCUSSION

The results of the analyses are shown in Table I. Besides B_{calc}, the values B_{real} measured at satellites, are cited in

the form: averaged B $\pm$ deviation (see catalogue by King 1977).

The comparison of B_{calc} with B_{real} demonstrates that B_{calc} lies within the limits of the deviation, i.e. in the limit of 5 % from the averaged value B.

The Table I shows that for 10 events from 20, no common frequencies were observed in the spectra. For the remaining 10 events B_{calc} was calculated. 9 of them revealed a very good coincidence with B_{real}. The error of the definition of B is within the limit of 5 %.

For our limited statistics we can estimate the percent of success of the method being – 90 %, to be compared with the B-index-method which had 33 % of success, and the „Advanced B-index" – method which had 55 %.

Let us estimate the accuracy of Eq. (1). The error of calculating B results from sources:

$$\frac{\Delta B}{B} = \frac{\Delta F}{F} + \frac{\Delta C}{C} . \qquad (3)$$

Which of the two parameters C or F give the larger contribution to the error of B-definition?

First of all we consider the error of C.

The coefficient C was investigated in many studies:

1. ground-based pulsation measurements at one station and at networks of stations (Troitskaya et al. 1971, Vinogradov and Parkhomcv 1974, Gul'elmi and Bolshakova 1973, Gul'elmi 1974, Barchatov et al. 1974, Russel and Fleming 1976, Verő and Holló 1978, Plyasova-Bakounina et al. 1978, Verő 1980, Odera 1982, 1984, Green et al. 1983, Verő et al. 1985, Engebretson et al. 1986, Best and Best 1987);
2. pulsations at geosynchronous orbit (Yumoto et al. 1984);
3. waves in the solar wind in the vicinity of the Earth's bowshock (Russel and Hoppe 1981, Odera 1982);
4. waves in the vicinity of other planets (Hoppe and Russel 1982, Russel and Hoppe 1983).

The last studies (Hoppe and Russel 1982, Russel and Hoppe 1983) represented strikingly that coefficient in the F-B relationship given there practically (within the limit of

deviation) coincides with the same one received from ground-based data by Troitskaya et al. (1971) what is important for the clear understanding of the problem of the mechanism of acceleration of protons at shocks in space.

The average value deduced from the above mentioned papers is:

$$C = 6.2 \pm 0.3 \text{ mHz/nT}$$

(compare Gul'elmi 1988).

Since the contribution of C = 6.25 to the error of B is small (5 %), the accuracy of the calculated B-value depends mainly on the accuracy of the determination of F. How may the accuracy of the determination of F be improved?

First of all - the methodology can be improved, e.g. by a massive statistical study in order to find the significant optimum level of modulus of coherence function.

One of the weaknesses of the present method is the possibility of the appearance of a false maximum in the coherence function due to a coincidence of local resonances in Lindau and Sverdlovsk, because they occupy the same L-shell. This seems to be happened in event 8 of Table I, the frequency of which corresponds probably to the third harmonic of the fundamental model (Pilipenko et al. 1988). By arranging stations along a meridian can avoid the appearance of false maxima.

It is possible to master the method itself by the installation of the stations along a meridian. Baransky et al. (1985) analysing Pc2-4 data on the meridional network of stations, discovered that the D-component of pulsations has a constant spectrum along the meridian. On the contrary the spectrum of the H-component of pulsations changes along meridian according to the prediction of the resonance theory. It means that the spectrum of the H-component is defined mainly by resonanting pulsations in comparison with the spectrum of the D-component which is defined mainly by solar wind-controlled pulsations, the period of which is supposed to be

constant in a large area, according to the theory of extramagnetospheric origin.

Hence it is possible to select solar wind controlled pulsations by analysing only the D-component of Pc2-4 for stations located along the meridian. The distance between stations should be more than 10^{o}, if one takes into account the estimated radial size of the resonance region in the magnetosphere (Hughes et al. 1978, Hughes 1980).

REFERENCES

Akasofu S I 1981: Space Sci. Rev., 28, No. 2, 121-190.

Baransky L N, Borovkov J E, Gochberg M B, Krylov S M, Troitskaya V A 1985: Planet Space Sci., 33, 1369-1374.

Barchatov N A, Levitin A E, Rudneva N M, Feldstein J I 1974: Interplanetary plasma and pulsations Pc3-4. In: Solnechni veter i magnetosfera, Moscow, IZMIRAN (in Russian)

Best I, Best A 1987: Day-variation of the maximum of the power spectrum in the Pc3-4 range in connection with Borok-index. HHI-Rept., 21, 209.

Chen L, Hasegawa A 1974: J. Geoph. Res., 79, 1024-1032.

Engebretson M I, Meng Ch I, Arnoldy R L, Cahill L J Jr 1986: J. Geophys. Res., 91, 8909.

Fairfield D H 1969: J. Geoph. Res., 74, 3541.

Green C A, Odera T J, Stuart W F 1983: Planet. Space Sci., 31, 559.

Gul'elmi A V 1974: Space Sci. Rev., 16, 331.

Gul'elmi A V 1988: Geomagn. i Aeron., 28, 3, 465.

Gul'elmi A V, Bolshakova O V 1973: Geomagn. i Aeron., 13, 559.

Gul'elmi A V, Plyasova-Bakounina T A, Chepetnov R V 1973: Geomagn. i Aeron., 13, 382.

Hoppe M M, Russel C T 1982: Nature, 295, 41-42.

Hughes W J 1980: J. Geomagn. Geoelectr., 32, Supplement, 41-45.

Hughes W J, McPherron R L, Barfield J N 1978: J. Geoph. Res., 83, 1109-1116.

King J H 1977: National Space Data Center, NASA

Lanzerotti L J, Maclennan C G, Medford L V, Carpenter D L 1986: J. Geophys. Res., 91, 375.

Münch J W, Wilhelm K, Troitskaya V A, Kasak B N, Baranskiy L N 1975: In: Kleinheubacher Berichte, Fernmeldetechnisches Zentralamt, Darmstadt, Band Nr. 18, 123-132.

Odera T J 1982: The control and generation of magnetic pulsations on the ground and interplanetary space by parameters of the solar wind. Ph. F. Thesis, University of Edinburgh

Odera T J 1984: Acta Geod. Geoph. Mont. Hung., 19, 305-338.

Odera T J 1986: Rev. Geophys., 24, 55-74.

Pilipenko V A, Povzner T A, Savin I V, Nikomarov J S 1988: Physics of the Earth, No. 10, 54-61 (in Russian), Izvestija AN SSSR, Fiz. Zemli, No. 10, 54-61.

Plyasova-Bakounina T A, Golikov Yu V, Troitskaya V A, Hedgecock P C 1978: Planet Space Sci., 26, 547.

Plyasova-Bakounina T A, Stuart W F, Troitskaya V A, Kharchenko I P 1983: In: Issledovania po geomagnetismu i aeronomii i fizike Solntza, Irkutsk (in Russian)

Plyasova-Bakounina T A, Troitskaya V A, Münch J W, Gauler H F 1986: Acta Geod. Geoph. Mont. Hung., 21, 143-153.

Pukovkin M I, Semenov V S 1986: Geomagn. and aeronomie, 26, 1026-1028.

Russel C T, Fleming B K 1976: J. Geophys. Res., 81, 5882.

Russel C T, Hoppe M M 1981: Geophys. Res. Lett., 8, 615.

Russel C T, Hoppe M M 1983: Space Sci. Rev., 34, 155-172.

Southwood D J 1974: Planet. Space Sci., 22, 483.

Siebert M 1964: Planet. Space Sci., 12, 137-147.

Troitskaya V A, Plyasova-Bakounina T A, Gul'elmi A V 1971: Dokl. Akad. Nauk SSSR, 197, 1312.

Verő J 1980: J. Atmos. Terr. Phys., 42, 371-380.

Verő J, Holló L 1978: J. Atmos. Terr. Phys., 40, 857.

Verő J, Holló L, Potapov A S, Polyushkina T N 1985: J. Atmos. Terr. Phys., 47, 557.

Vinogradov P A, Parkhomov V A 1974: Geomagn. Aeronom. Engl. Transl., 15, 109.

Yumoto K, Saito T, Tsurutani B T, Smith N J, Akasofu S I 1984: J. Geophys. Res., 89, 9731.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 273-285 (1991)

DETECTABILITY OF HIGH-CONDUCTIVITY PLATES BY THE CSAMT METHOD ON BASIS OF ANALOGUE MODELLING RESULTS*

An interesting analogue modelling experience

L Szarka

Geodetic and Geophysical Institute of the Hungarian Academy of Sciences,
H-9401 Sopron, POB 5, Hungary

[Manuscript received October 19, 1990]

In the CSAMT method it is very rare that in the anomaly maps the pure geometry of the model would be directly reflected. In an analogue model experiment 3-D thin-sheet like graphite plates were put on the salt water/high resistivity bottom boundary and amplitude and phase of the electric component parallel to a distant HED transmitter were measured. The position of the phase anomaly proved to be stable in the whole period range (i.e. in the wave-, the transition-, and in the quasi-near zones) serving a correct basis for the interpretation. The amplitude maps were more variable and unusual behaviour was discovered in a narrow period range (that is at the very beginning of the evolution of the long-period anomaly): the small changes over the 3-D thin-sheet like plate were found in an unbelievable close relation with the real shape of the model, leading to much better detectability, than at any other periods. In spite of that this effect in the field in most cases would be masked by measuring errors, this experiment is a clear proof that we must pay more attention to such transition phenomena.

Keywords: analogue modelling; controlled source methods; electromagnetic methods; horizontal electric dipole; three-dimensional problem

INTRODUCTION

In 1989 at the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences CSAMT analogue model experiments were carried out for the Geophysical Exploration Company of the Hungarian Oil and Gas Trust. In the model tank several high-conductivity thin-sheet like plates having the same area but having different geometrical shape were put and the object of the experiment was to find relations between the anomalies and the geometrical shape of the models.

*Paper presented at the 10th Workshop on Electromagnetic Induction in the Earth IAGA WG I-3, Ensenada, August 1990

(A description of the field problem is given by Nagy et al. 1989.)

As a transmitter a distant HED was used and the amplitude and phase of five electromagnetic field components were measured at eight different frequencies. (Later the measurements were limited to the electric components.) The results are summarized in our Research Report (1989). (Other 3-D examples are collected by Nagy and Szarka 1989.) Here only an interesting modelling experience, as a "by-product" of the measurements, is presented.

DESCRIPTION OF THE MODELS

The description of the modelling equipment is discussed in details by Märcz et al. (1986) and Ádám et al. (1981). A plan-view in Fig. 1 shows

- the HED transmitter,
- three different models: a, b and c,
- orientation of the measuring dipole MN,
- the measuring grid.

In Fig. 2 the cross section is shown. Figures 1 and 2 are proportional in all directions.

ABOUT THE DATA USED FOR THE INTERPRETATION

Relative amplitude (amplitude with graphite/amplitude without graphite) and relative phase (phase with graphite - phase without graphite) of the horizontal electric field measured by MN (see Fig. 1) were determined.

Figure 3 shows typical sounding curves at three different measuring sites. (Along the horizontal axis lambda is the wavelength in salt water and h1 is the depth of the models.) In agreement with 1-D theoretical calculations it can be seen that where the phase reaches its maximum, the amplitude anomaly is practically zero.

We were interested first of all in the "big" amplitude anomalies: in the short-period ABS(E_x)-increase (at

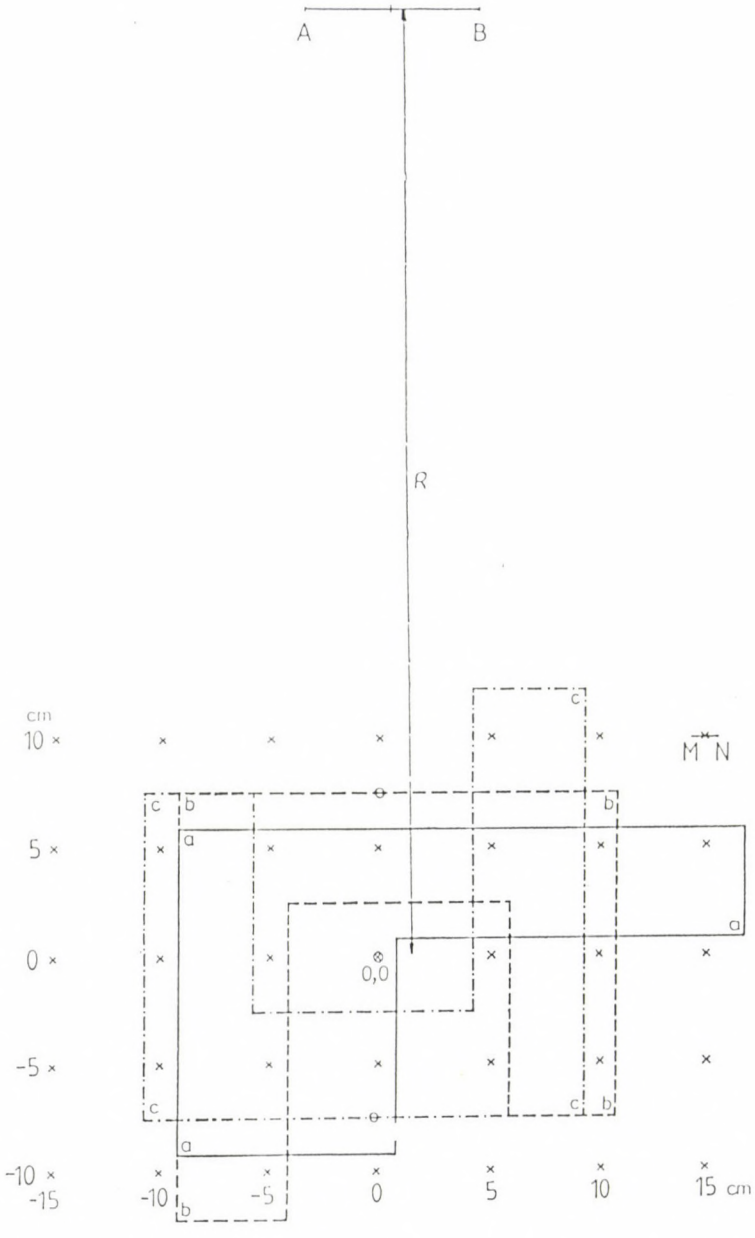


Fig. 1. Plan-view of the CSAMT model experiment. AB: current electrodes, MN: measuring dipole for E_x
model a: pistol
model b: reversed letter U
model c: letter U

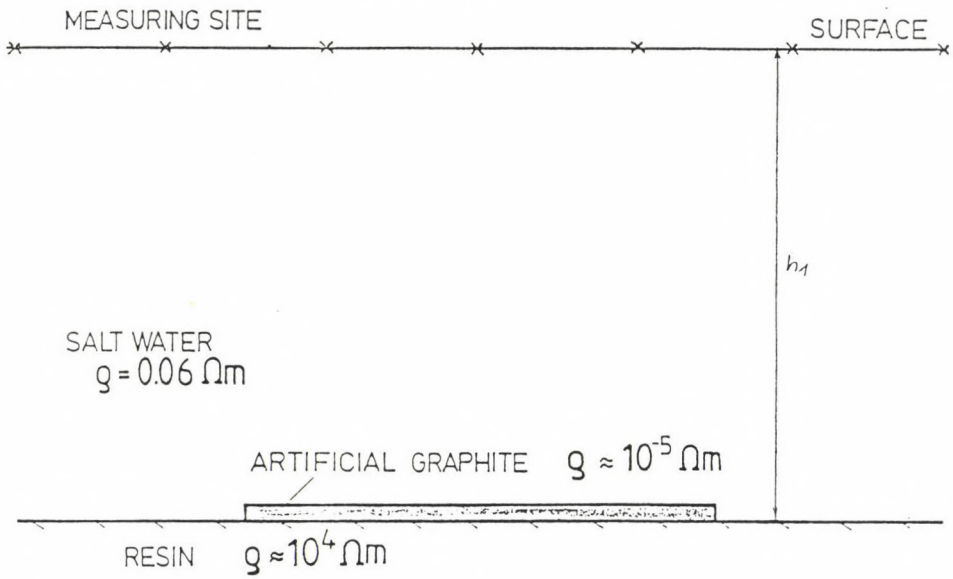


Fig. 2. Cross section

lambda/h1 = 4-5), in the long-period ABS(E_x)-decrease, and also in the phase anomaly.

In Fig. 4, 5, 6 and 7 amplitude and phase anomaly maps are shown at a characteristic "short" period and a characteristic "long" period for all the three models. In Fig. 8 the phase anomalies are shown at lambda/h1 = 5.49.

In figures it can be seen that the amplitude anomalies are more variable, while the phase anomalies are more stable. The resolution power in neither case can be stated to be high. Those parts of the model which are perpendicular to the direction of the transmitter, remain hidden.

A SURPRISING RESULT

In a narrow transition zone, where the ABS(E_x) anomaly is just changing from its short-period maximum to a more and more important long-period minimum, the ABS(E_x) values should be close to 1. Of course, at different sites these values 1 are

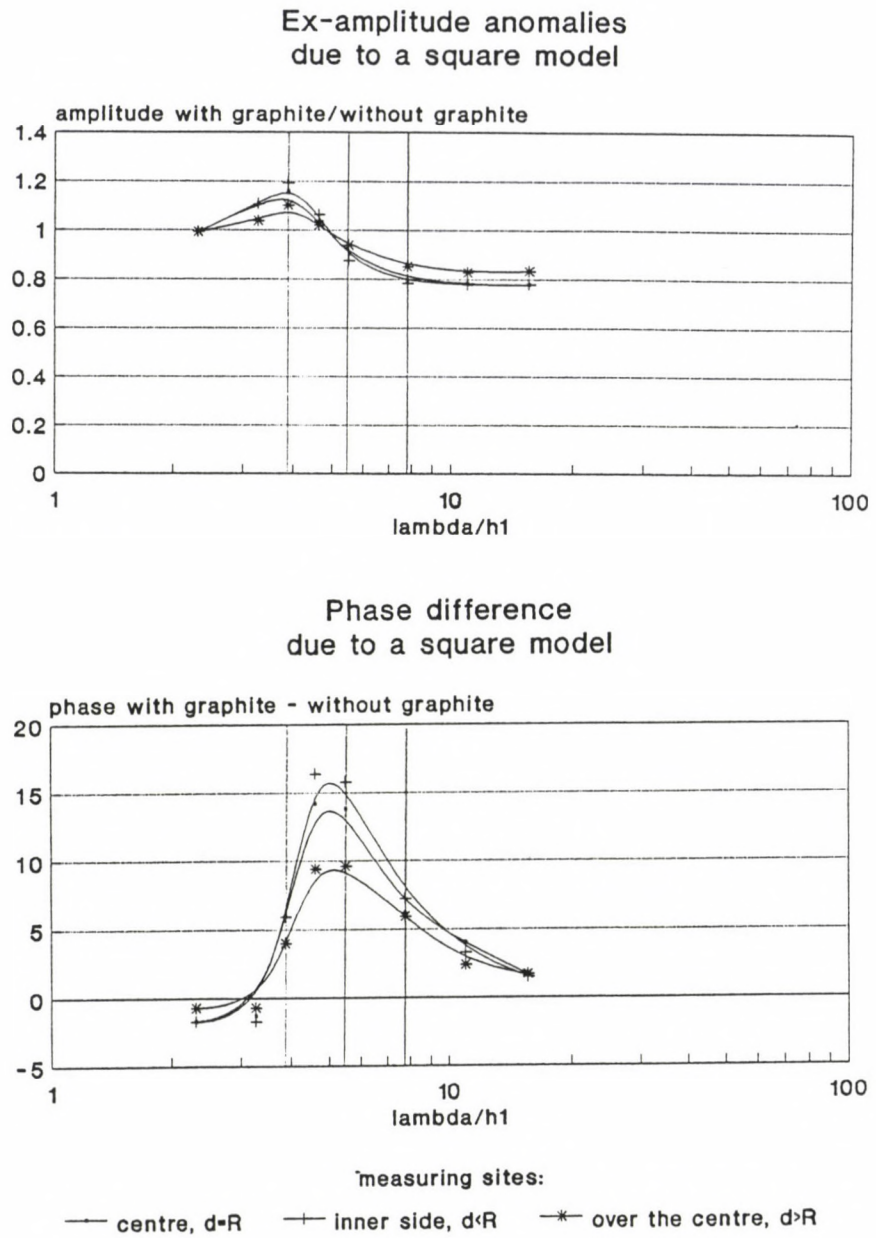


Fig. 3. Relative E_x amplitude and phase sounding curves. The anomaly maps were measured at relative wavelengths indicated by the three vertical lines: lambda/h1 = 3.82 ("short" period), lambda/h1 = 5.49 (in the "critical" period range), lambda/h1 = 7.77 ("long" period)

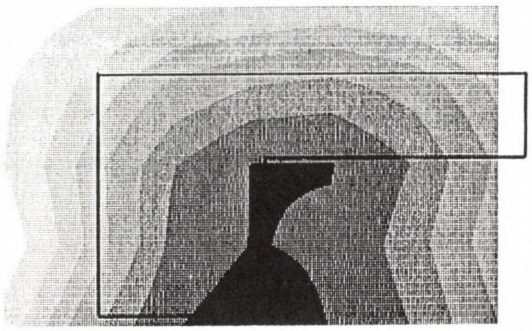
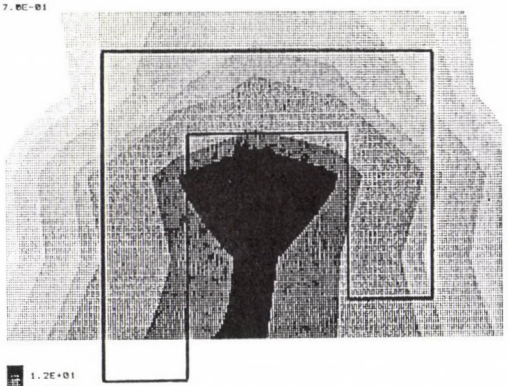


Fig. 4a. E_x phase over model a at short period



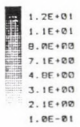
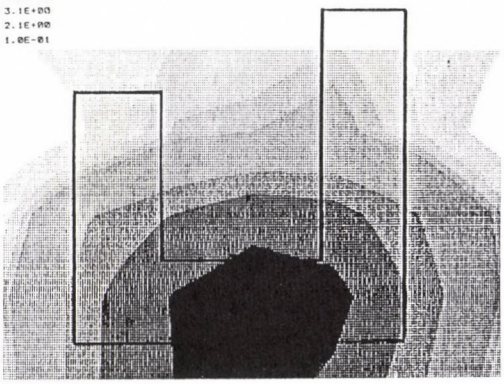


Fig. 4b. E_x phase over model b at short period



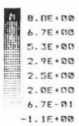
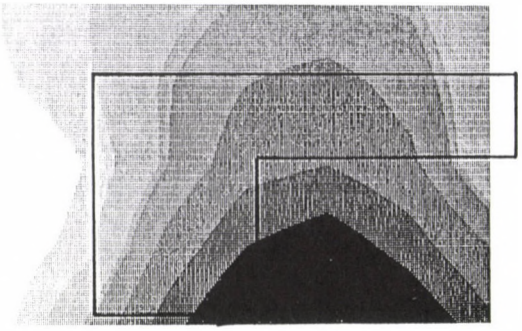


Fig. 4c. E_x phase over model c at short period



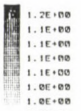
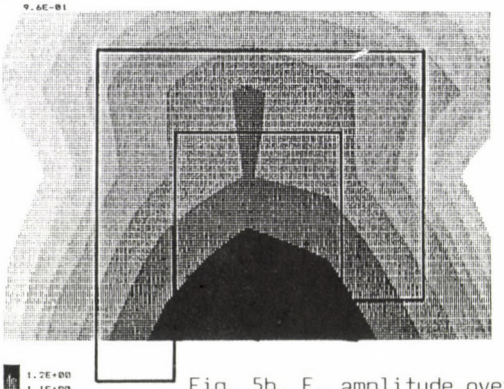


Fig. 5a. E_x amplitude over model a at short period



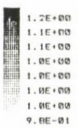
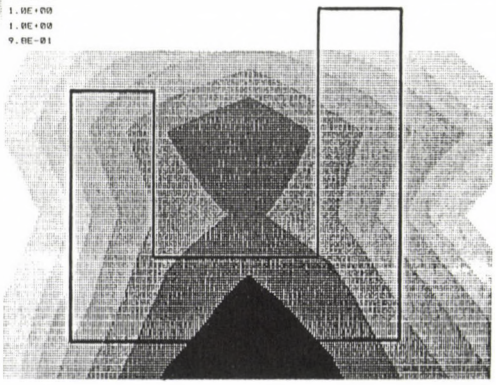


Fig. 5b. E_x amplitude over model b at short period



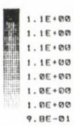


Fig. 5c. E_x amplitude over model c at short period

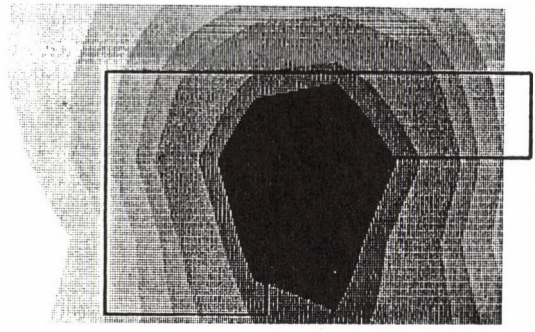
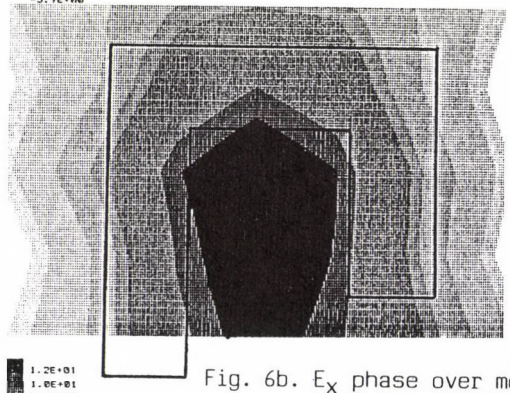


Fig. 6a. E_x phase over model a at long period



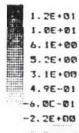
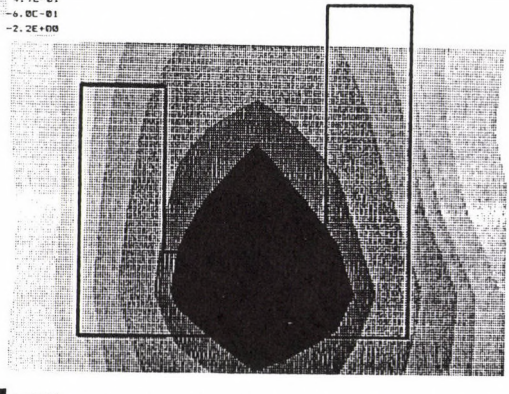


Fig. 6b. E_x phase over model b at long period



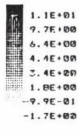
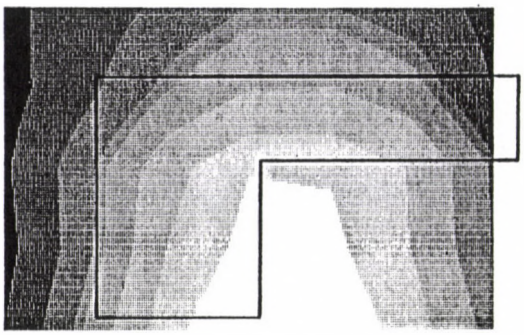


Fig. 6c. E_x phase over model c at long period



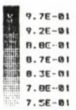
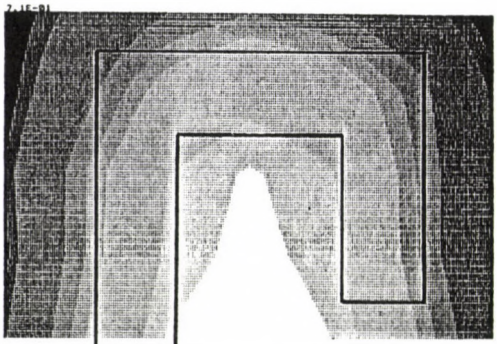


Fig. 7a. E_x amplitude over model a at long period



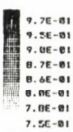
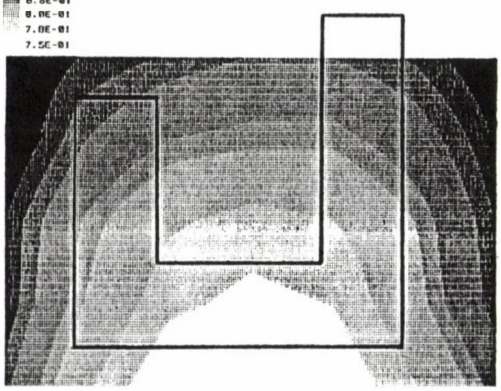


Fig. 7b. E_x amplitude over model b at long period



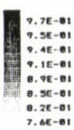
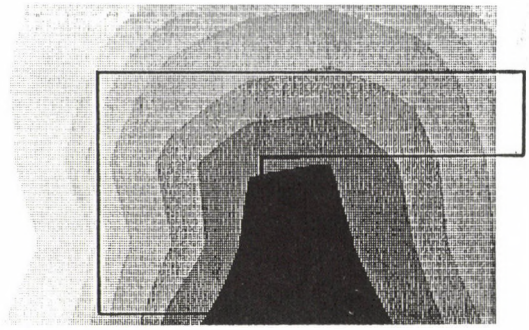


Fig. 7c. E_x amplitude over model c at long period



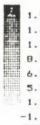
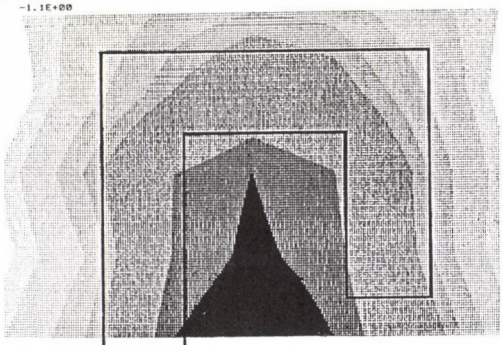


Fig. 8a. E_x phase over model a at the critical period



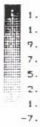


Fig. 8b. E_x phase over model b at the critical period

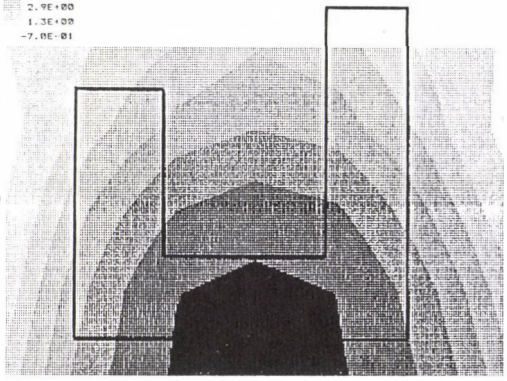


Fig. 8c. E_x phase over model c at the critical period

taken up at slightly different periods, depending on the transmitter-receiver distance and the actual conductivity structure. Consequently if we have a map measured at a certain period, slightly different values will belong to different measuring sites. Since at this period the variation over the model is expected to be somewhere at the measuring errors, nobody has paid attention to the details of these small anomalies. Therefore it was a surprise for us to have a really good correlation between the isolines and the shape of the three structures as it is shown in Fig. 9. (E.g. a rotation of the model b by 180 degrees to get the model c is followed by the isolines only at this critical period.)

REMARKS

According to Fig. 9 the evolution of 3-D anomalies takes place in a very exciting way as far as the isolines in the transition zone seem to be closely connected to the geometric shape of the models. It seems that the bigger the anomaly, the worse is the lateral resolution.

Unfortunatelly this delicate phenomenon in the field is probably masked by measuring errors, hindering any direct application. Further studies (model measurements and comparisons with 3-D numerical results) are needed yet to give satisfactory explanation of the phenomenon. The aim of this paper is just to call the attention to this problem.

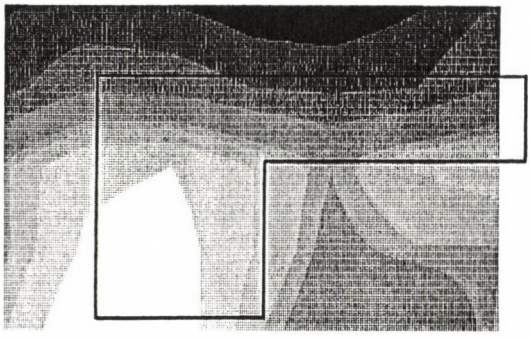
REFERENCES

Ádám A, Pongrácz J, Szarka L, Kardeván P, Szabadváry L, Nagy Z, Zimányi I, Kormos I, Régeni P 1981: Acta Geod. Geoph. Mont. Hung., 16, 359-378.

Märcz Gy, Pongrácz J, Szarka L 1986: Scientific Instrumentation, 1, 119-133.

Nagy Z, Szarka L 1989: 3-D CSAMT analogue modelling results. Paper presented at the EAEG 51st Annual Meeting, Berlin, 1986

Nagy Z, Beke B et al. 1989: Electromagnetic investigations in the EOR program by the CSAMT method. Paper presented at



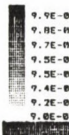
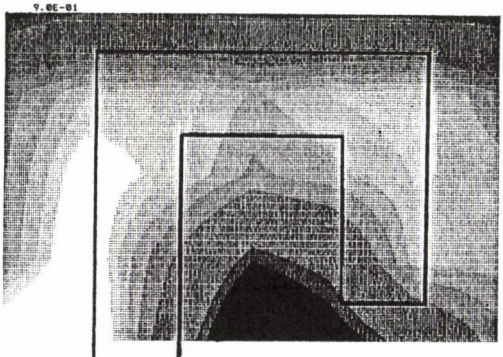


Fig. 9a. E_x amplitude over model a at the critical period



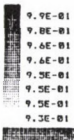


Fig. 9b. E_x amplitude over model b at the critical period

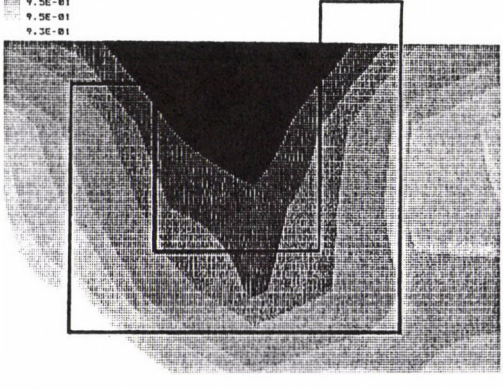


Fig. 9c. E_x amplitude over model c at the critical period

the International Geophysical Symposium, Budapest

Research report for the Geophysical Exploration Company on the EOR analogue modeling results 1989: Geod. Geoph. Research Institute, Sopron

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 287-300 (1991)

ON THE PROTECTION AND OPERATION OF NUCLEAR POWER PLANTS AND GREAT STRUCTURES IN CENTRAL EUROPE WITH RESPECT TO THE TECTONICS OF EUROPEAN ALPIDS. A CONCEPTION STUDY

J Nedoma

Inst. of Computer Sciences, Czech. Acad. Sci., 182 07 Prague 8-Libeň, Pod vodárenskou věží 2, Czechoslovakia

[Manuscript received October 26, 1990]

This conception study wants to show that the financial resources, needed for exploitation of the energetical resources as well as for the security of great energetical structures (i.e. nuclear power plants (NPP), deep mines, great waterdams, etc.), can be employed with high effectivity only when using appropriate basic and applied geological and geophysical investigations. One can only by this way save money expended during the last years to great structures built in critical regions. The breakdown of the Chernobyl NPP and the catastrophic earthquake in Armenia as well as the construction of a NPP there, situated on the Ararat deep fault, in a relatively small distance from the earthquake region, warn us and constitute the main reason for the realization of the research programme presented in this paper as well as for the recommendation of this research programme to be solved within the framework of a Pentagonal group.

Keywords: Alpids; earthquake risk; nuclear power plants; tectonics; waterdams

INTRODUCTION

A substantial part of the increase of energy needs in the countries of Central Europe is to be covered by nuclear power plants (NPPs) for which there is a good raw material supply e.g. in the deposit of the Bohemian Massif. With respect to the fact that the whole Central Europe is densely populated, the detailed observation of physico-tectonic processes in the area of construction of a NPP and other great critical structures themselves as well as in their farther vicinity is considered as an immediate necessity. The observation has to take into account the regional viewpoint of geotectonic processes in the central and southern parts of Europe and is to be based on a detailed fundamental geodynamic research. Here, the tectonics

is represented by the collision of the African plate (more exactly, the Adriatic microplate) with the Euroasian one which has resulted in folding of the Alps and of the Carpathian Arc.

The breakdown of a NPP in the world has disastrous alarming consequences for the population of large regions of our planet. To build NPP-s in Central Europe, in contrast to other countries as US, etc., means to construct in densely populated areas geologically influenced by Tertiary and older tectonics. A great deal of deep fault zones are active there even now.

GEOLOGICAL STRUCTURE OF CENTRAL EUROPE

The geological structure of Central Europe is strongly influenced by the Alpine orogene mainly in the Tertiary era. From a geographical point of view the region of Central Europe covers the Bohemian Massif as fundamental part of this domain, the platform part of Europe to the North, the West Carpathians to the East and the Eastern Alps to the South (the latter are immediate results of the collision of the Adriatic microplate with the Euroasian plate) and the Pannonian Basin. According to the plate tectonics, Central Europe is characterized by a mosaic of lithospheric blocks. In Czechoslovakia a NPP or other critical structures are being built in the Bohemian Massif and the West Carpathians.

The greater part of the Bohemian Massif is in the territory of Bohemia and Moravia, marginal parts extend to Austria, to Germany and to Poland. The massif is bordered to the S and SE by the Alpine-Carpathian young fold-belt, its Neogene foredeep is situated on the submerged parts of the massif. It is bounded to W by a number of faults separating it from the Mesozoic Bavarian plate. Its margins are covered to N and NW either by Upper Permian and Lower Triassic sediments or by thick Quaternary deposits. It is bounded by the great Odra fault to NE.

In the Bohemian Massif geological blocks can be defined of a relatively regular mosaic structure. The predominant

directions of tectonic lines in the Bohemian Massif are the same as in whole Europe. The directions are approximately NE (Krušné Hory fault (6), Litoměřice fault (1), Klatovy fault (2), Benešov fault (3), Kladno (4) and Stříbro (5) faults), approximately NW (i.e. Poděbrady fault (7), Jáchymov fault (8), Tachov fault (9), Mariánské Lázně fault (10), Horní Slavkov fault (11), Tábor fault (13) Říčany fault (15), Cínovec fault (16)) and approximately NNE-SSW (i.e. Blanice graben (17) and Pilsen fault (18)) i.e. the so-called Rhenish and Jizera direction as well as faults perpendicular to the Rhenish direction which are of a less importance (see Fig. 1 after Röhlich and Štovíčková 1968). These deep faults divide in the Bohemian Massif blocks of different geological structure and development. The following blocks have been defined: Krušné Hory Mts block, Teplá-Barrandian block, Central Bohemian transitional segment, Český Les segment, Moldanubian block divided into Bohemian and Moravian segments, Elbe zone segment, West Sudeten block and Moravian-Silesian block. A recent diagonal pattern of deep faults limit these blocks.

The development of the deep structure of the Bohemian Massif can be summarized in four stages of tectonic development, i.e. the Pre-Cadomian, the Cadomian, the Variscan and the platform stages. Although the contact between the platform and the Alpids is at present relatively quiet, there is a number of conclusive geological evidences about motions in this region during Pliocene and Late Quaternary (Pleistocene), but as geophysical measurements (e.g. recent crustal movements) show present lithospheric motions also exist in this region.

The European Alpids (the Alps, the Carpathians, etc.), represent a system of zonal mountains. Their origin is due to Alpine folding. The European Alpids are in a sharp tectonic contact with the European platform. Their basic and characteristic feature is a large number and variety of tectonic units formed by the Alpine folding, mostly from sediments of the Alpine cycle. The basement units built up of pre-Alpine complexes were rebuilt and incorporated into the Alpine structural plan. From the viewpoint of Tertiary

Fig. 1.

tectonics, the Alpine-Carpathian mountain system has two different zones.

The West Carpathians are divided into several zones, each of which is characterized by its historical evolution and structure: 1. Inner (Central) Carpathians, 2. the Klippen zone, 3. Outer (Flysch) Carpathians, 4. Carpathian Foredeeps. There are several deep-seated fault zones in the Bohemian Massif, also in the West Carpathians which divide the zonal structure into large blocks. These faults only became more active in the Late Tertiary, but they are active also at present. During the Neogene the Inner Carpathians split into a series of blocks, their motions are associated with Miocene volcanism forming the Inner Carpathian volcanic arc.

The most expressive seismically active geological boundary is the Klippen Belt between the Inner and Outer Carpathians (the so-called Záhorie-Humenné deep fault or the Peri-Pienniny zone, respectively). Other boundaries are: a) the Vepor deep fault or the Raab's line, respectively which crosses the inner West Carpathians from Komárno to Poprad and continues to the NE into the Outer Carpathians, b) the Štiavnica-Přerov deep fault in NW-SE direction, c) the Slanský deep fault in East Slovakia running N to S along the Slanské pohorie mountain range (see Fig. 2, after Fusán et al. 1979). The mentioned deep faults define the fundamental neotectonic blocks of the West Carpathians, i.e. Danubian, South-Slovakian, East-Slovakian, Fatra-Tatra, Slovak-Silesian and Slovak-Moravian blocks. All these blocks are divided into several sub-blocks.

In the southern part of the Danubian block there are two seismically active faults. The northern one extends from Pezinok to Komárno. In the Malé Karpaty mountains it is reflected by an axial depression between Bratislava and Modrá. The southern one extends from Bratislava towards SE along the river Danube. In the northern part of the Danubian block, the Dobrá Voda zone is an important seismically active fault zone. It extends from the depression between Malé Karpaty and the Nedzev Range towards SE, i.e. to Trnava and Nové Zámky, passes through the central part of the Danubian block. It divides the

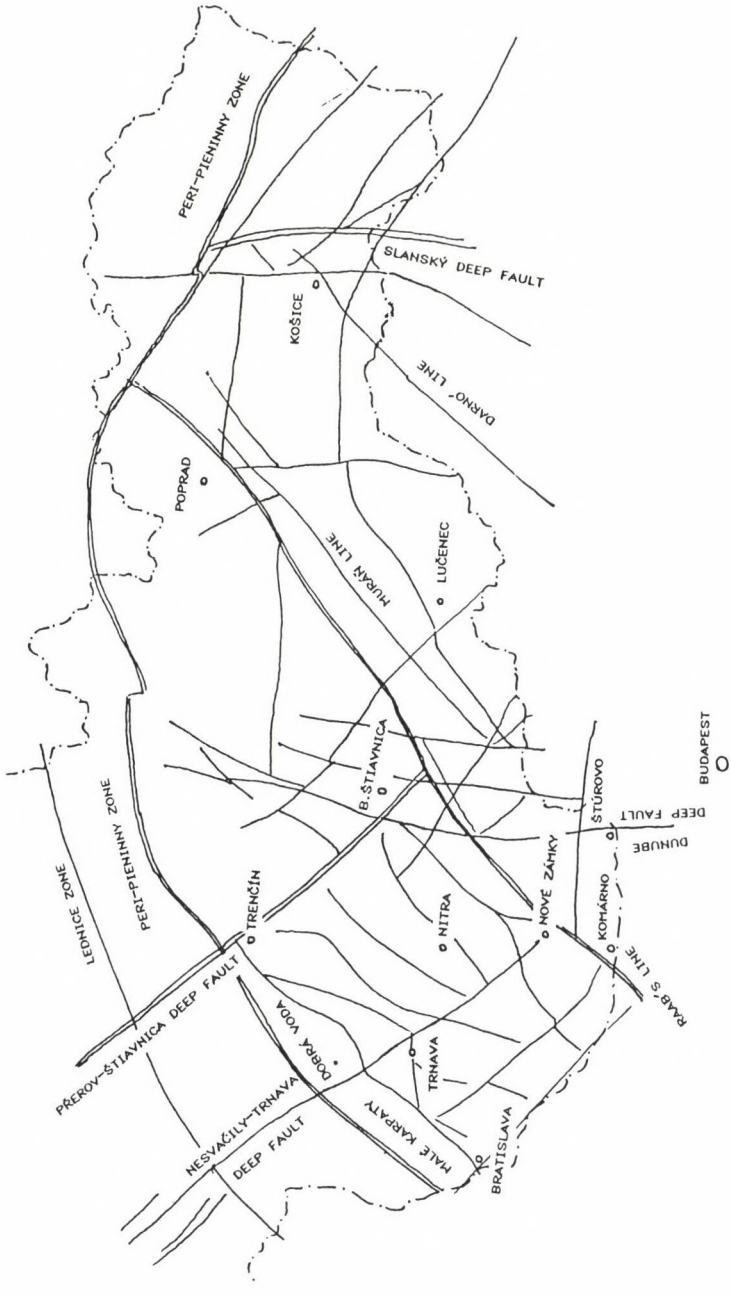

PERI-PIENINNY ZONE
SLANSKÝ DEEP FAULT
KOŠICE
DARNO- LINE
POPRAD
HUSÁR LINE
LUČENEC
B.ŠTIAVNICA
BUDAPEST
ŠTÚROVO
DEEP FAULT
DUNAJ
NOVÉ ZÁMKY
KOMÁRNO
RAAB'S LINE
TRENČÍN
NITRA
LEDNICE ZONE
PERI-PIENINNY ZONE
PŘEROV-ŠTIAVNICA DEEP FAULT
DOBRÁ VODA
TRNAVA
NESVAČILY-TRNAVA DEEP FAULT
MALÉ KARPATY
BRATISLAVA

Fig. 2.

Danubian block into a northern and a southern part. In other fundamental blocks, deep faults also exist. An interesting fact is that the fundamental blocks are separated by seismically active zones. The partial blocks are bounded as a rule by active zones, many of which are seismically active and simultaneously form boundaries in the intensity of recent vertical movements and of the geothermal field between individual blocks (see e.g. Fusán et al. 1979, Zátopek 1979, Dudek 1981, Beránek and Zátopek 1981, Mahel' et al. 1974, Ádám 1976).

DISCUSSION

A substantial part of the increase of the energy consumption in Central Europe is to be covered by NPPs. Unfortunately, their sites are not always optimally chosen. With regard to geologically complicated conditions in Central Europe, they are often located in seismically active areas. A series of earthquakes in the area of Dobrá Voda (northern epicentral zone of Malé Karpaty) at a rather short distance from the NPP Jaslovské Bohunice in 1987 is a first serious warning. It is well known that only part of the motion of lithospheric blocks results in seismic effects. Geodetic measurements have proved the recent motion in the Malé Karpaty region. The sites of NPPs situated in seismically active areas with the effects of recent seismic activity (NPP Jaslovské Bohunice, a depression area with active faults in the vicinity of the NPP Temelín, situated in the Bohemian Massif, horizontal movements in the area of Blahutovice situated on the Lednice deep fault, etc.) raise the necessity of a detailed observation of geodynamical processes in the area of construction of NPPs and of other great critical structures as well as in their broader vicinity (see e.g. Nedoma 1986). The breakdown of a NPP or of a great waterdam, respectively, with disastrous consequences for the population of large regions of our planet is alarming. Moreover, building NPPs in conditions of Central Europe, in contrast to other countries with spacious

uninhabited territories (China, USA, USSR, etc.), means to construct in densely populated areas. The more so we should do our best to ensure the safety of NPPs in these areas and, with it, also our safety. The construction of NPPs should be decided - after a detailed and, in particular, trustworthy investigation and paying attention to the public opinion - in the first place to scientists and not to politicians who, even with the best intentions, cannot professionally judge advantages or drawbacks of the possible sites. Having the recommendation of scientists, the subject should be again discussed. We can only avoid a possible failure by means of consistent control and criticism of the project before its realization stage with immense consequences or, in a better case, with a multi-billion economic loss.

One of the principal factors that determine the safety of NPPs is the maximum possible earthquake which still does not endanger the operation of a NPP. Several methods are used to estimate it. They are based on calculations from simple observed data or on an evaluation of geological-geophysical data from the area concerned. The methods involve estimates of the maximum magnitude of earthquakes M_{max} in the region of the NPP by seismology, statistical data analysis with the help of the Gumbel distribution, and analyses of seismic, gravimetric, and geodetic data from the investigated area. All these methods evaluate observations on the Earth's surface. None of them takes into account the accumulation of stress in geological blocks defined by tectonic lines in the areas considered and in their broader vicinity and, therefore, none of them studies the causes of activity in these areas. The project which is the subject of my proposal and whose realization is possible from a technical point of view, is based on stress data.

Since it is not possible to increase the number of structural boreholes in the area of the construction of a NPP beyond an acceptable measure for reasons of both economy and resources, the evaluation of the actual stress and structural state has to use an analysis of geodynamical processes in the area considered and in its vicinity, i.e. tectonics,

sedimentology and, in particular, stress state in the neighbourhood of a few boreholes in the area. We have to be aware that while the interest of builders of a NPP (as well as of other great critical structures) is concentrated on a detailed evaluation of a few square miles of the area concerned, the laws influencing tectonics and stress state in this area must be studied from a regional point of view, i.e. in a large area of which the area of NPP is only a part. Without this information, our knowledge would be incomplete and could therefore result in an image false from the viewpoint of the tectonics of the area. This is the reason why all the available boreholes in the area should serve this purpose. They ought to be equipped with probes (which are certain small laboratories) capable of measuring time changes of the stress (from the initial stress values found in the area of the NPP, i.e. by the overcoring method) caused by tectonics and, in addition, also by civilization activity in the rock massif during the construction of NPP as well as thermal and further geophysical fields. Since we can only find the stress state in the area at discrete points, we use mathematical modelling and simulation of geodynamic processes in the area concerned to extrapolate the stress information to greater depths and longer distances from the place of stress measurement in situ.

At present, the problems mentioned are only very sparsly studied by complex research projects at geophysical or geological institutes. No complex research does exist in the branches of interest governed by a unified methodology in the whole field. This is an enormous chance for fundamental research in Central Europe to get funds for a unified and complex geological and geophysical study in the central region of the contact zone between the platform and the Alpids. The fact that Czechoslovakia as a whole lies in the central part of the largest transversal elevation crossing the contact of the platform with the Alpids in Europe gives an opportunity to study geologically and geophysically the contact, its foreground as well as hinterland in a number of structural depths that can be reached by boreholes both in Bohemia and in

the Carpathians, too.

For the benefit of fundamental geophysical and geological research (with straightforward applications to the construction, operation and safety of NPPs, deep mines, great water dams, etc.), the proposed investigation should include the study of further geophysical fields and processes. In such a case the fundamental research of the regional area has to include:

a) long-term tectonophysical research of the present motion of the lithosphere by the means of the stress-strain field observed in accessible places of the lithosphere (in deep boreholes at various depths, in gallery etc.). In addition, observation of the recent motion of the Earth's crust (i.e. of both components, vertical as well as horizontal) and of the seismicity in the area (detectors, i.e. wide-band geophones, are placed in a probe to measure the stress-strain field in the borehole during long time) and detailed determination of the geological structure of the area by methods of seismic and electromagnetic tomography in boreholes, seismic microzoning, etc.

b) Theoretical analysis of the stress-strain field in the area completed with the map of the stress-strain field and with a database for various depths. Comparison of theoretical results based on mathematical modelling and simulation of tectonophysical processes in the regional area with the results of measurements in situ. Extrapolation of the results of measurements to longer distances and greater depths from the place of measurement in situ. Choice of a most probable geodynamical model of the area based on expert systems and further improvement of this model facilitated by continuing research.

c) Measurement of the seismoacoustic activity in the long-term observation points (detectors-geophones, are placed in the probes for the long-term measurement of the stress-strain field in the boreholes).

d) Long-term observation of the stress-strain field caused by tidal forces in the area of interest and the following

comparison with analogous measurements in tidal research stations and stations in other areas.

e) Observation of the distribution and orientation of macro- and microcracks by pattern recognition methods and determination of the stress in the rock massif by means of the study of the deformicity of rock crystallic structure and of the acquisition of rock cores from places where the overcoring method is applied to study the present stress-strain state in the rock massif (this is part of the investigation in item a).

f) Long-term measurement of temperature and heat flow in observation of the stress-strain field (detectors-thermometers, are placed in the probes at long-term observatiqn points in accessible places of lithosphere).

g) All the currently performed measurements of physical quantities (as density, Young modulus of elasticity, Poisson's constant, porosity, etc.) of the rock-cores obtained in the overcoring method.

h) Paleomagnetic measurements of the rock core samples from the places of application of the overcoring method and of long-term observation points. They enable the paleomagnetic reconstruction of the orientation of the magnetic field and the determination of motion and historical rotation of the individual blocks that form the area of interest.

i) Magnetic measurements at long-term observation points.

j) Magnetotelluric measurements carried out at long-term observation points and, moreover, on exactly defined profiles that characterize the tectonics of the regional area.

k) Magnetometric measurements for determining fractures and their motion.

l) Deep seismic sounding on the profiles in item j) for determination of the geological structure of the area of interest.

m) Aerial and satellite photography of the considered area followed by evaluation.

n) Gradual creation of databases for the area of interest containing the results of the proposed complex research of the whole area.

Further improvement of geodynamical models of the regional

area based on the actual level of knowledge. This will allow the elimination of the influence of civilization on the area considered and facilitate finding criteria for the prediction of mine shocks and earthquakes in the whole area.

We admit that the task is extremely difficult. It is connected with a great deal of problems not yet solved that are outside the scope of geology and geophysics. The present state of the Austrian, Hungarian, Italian, Yugoslavian and Czechoslovak geology and geophysics as well as the state of these branches of science in Germany, Poland, Switzerland, USSR and Rumania, however, guarantee that it can be carried out and successfully solved. The geological, but even the economic community does not realize sufficiently the consequences of the project proposed in other branches of economy, e.g. mining, safety of construction and operation of deep mines, great waterdams, transportation, nuclear power plants, etc. (e.g. the problem of a low portion of obtainable reserves of solid fuel is not only the consequence of economic, technological, ecological and social circumstances, but, to a considerable extent, also of natural resources and, unfortunately, even that of the insufficient knowledge of geological structural phenomena of coal seams). Many legal problems should be immediately solved in this point, too. An example that hampers the development of complex geological and geophysical research is the problem of the residual value of the borehole after its completion for geological prospecting. The price of a borehole depends on its depth and ranges from several hundred thousands (for shallow boreholes) to several ten or hundred millions (for deep boreholes) in Austrian schilling or in Czechoslovak crown. After completing a borehole, the geological survey is not interested in these structural boreholes, they are closed up and destroyed. But such a borehole could be a valuable laboratory for complex geological and geophysical research, in particular for the study of temperature, stress-strain field in the rock massif, and, in addition, for further geophysical fields, microcracking of rock, safety of mines, waterdams, etc., including nuclear power plants. Moreover, it could

facilitate the prediction of mine shocks in areas of mines and great waterdams and earthquakes in seismically active zones in the vicinity of NPPs. The profit of such a study would then affect both fundamental and applied research and praxis, represented in the first place by mines, waterdams and NPPs. But the methodology of some research projects is applicable also to other branches of economy (e.g. the methodology of the stress-strain measurement in the rock massif is applicable to building industry and transportation - instability of slopes, tunnels, tubes, supporting walls, etc.). Moreover, the exploitation of boreholes for the above mentioned research valories them and the cost of their building related to their utilization decreases. Every geological borehole should thus be used for the above proposed research with a unified programme for the whole area (that may cover even several countries) and equipped with devices of unified type to ensure comparable results.

The inexorable logic of my note presenting the importance of organizing the complex geological and geophysical research lies in the fact that this subject improves the safety of the operation of NPPs and the activity in other branches of economy - as in mining. This research makes the life of coal mines and quarries in the next century many tens of years longer and protects thus future generations from great troubles with the energy balance. It helps to avoid catastrophic situations connected with breakdowns of NPPs, mines, and great waterdams. Moreover, the results of this research can be applied to other branches of economy. The contribution of fundamental research will apparently be much greater than the investment put in it. But we cannot do without changes of the present legal regulations, financial resources, development of new instruments and measurement devices (for the author's and his coworkers' results see Nedoma 1990a, b), and also a new, more flexible approach to urgent economic problems represented by the energy balance of the present, but above all the future generations and a new way of economic as well as political thinking in both local and international scale.

REFERENCES

Ádám A ed. 1976: Geoelectric and Geothermal Studies. KAPG Geophys. Monograph, Budapest, 752 pp.

Beránek B, Zátopek A 1981: Earth's crust structure in Czechoslovakia and in Central Europe by methods of explosion seismology. In: Zátopek A ed. 1981: Geophysical synthesis in Czechoslovakia. VEDA, Bratislava

Dudek A 1981: The principal features of the geological structure of Czechoslovakia with regard to adjacent region. In: Zátopek A ed. 1981: Geophysical synthesis in Czechoslovakia. VEDA, Bratislava

Fusán O, Ibrmajer J, Plančár J 1979: Neotectonic blocks of the West Carpathians. Geodynamic investigations in Czechoslovakia. VEDA, Bratislava

Mahel' M ed. 1974: Tectonics of the Carpathian-Balkan Regions. Geological Inst. of D. Štúr, Bratislava

Mantič V, Nedoma J 1990: In: Proc. of the 4th Int. Symp. on the Analysis of Seismicity and Seismic Risk. Bechyně Castle, Czechoslovakia, 1989, Geophys. Inst. Czech. Acad. Sci.

Nedoma J 1986: Tensometric probe for stress determination in a rock massif - its principle and use for stress monitoring in areas of construction and operation of nuclear power plants. Manuscript (in Czech)

Nedoma J 1990a: In: Report No. V-477, Inst. of Computer Sciences Czech. Acad. Sci., Prague (in Czech)

Nedoma J 1990b: In: Report No. V-476, Inst. of Computer Sciences Czech. Acad. Sci., Prague

Nedoma J, Hiršl P, Müller R, Šmuk K 1990: Automatic measuring and processing systems for monitoring stress-strain fields in situ. Manuscript (in Czech)

Röhlich P, Šťovičková N 1968: Geologie, 17, 670-694.

Zátopek A ed. 1979: Geodynamic investigations in Czechoslovakia. VEDA, Bratislava

Zátopek A ed. 1989: Geophysical synthesis in Czechoslovakia. VEDA, Bratislava

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 301-322 (1991)

DERIVING VELOCITY ESTIMATES FROM VSP DATA: A NOVEL APPROACH USING EDGE DETECTION

Z Wéber and I Bondár

Department of Geophysics, Eötvös University, H-1083 Budapest,
Kun Béla tér 2, Hungary

[Manuscript received January 8, 1991]

One of the prime purposes of vertical seismic profiling (VSP) is to find wave velocity as a function of depth. In the literature several proposed inversion methods can be found. The interval velocity function calculation from the first break time is the simplest method, but the results may have significant errors due to errors in the estimates of the first break times. The general linear inversion (GLI) method, that takes into consideration the whole VSP wavefield, gives more reliable results but it needs a good initial model, true amplitudes and its execution is very time consuming.

The inversion method described in this paper is free of the above mentioned disadvantages. After careful wavelet filtering and wavefield separation the gradient of the travel time curve of the first arrivals as a function of depth is estimated, on the basis of which the velocity function can also be estimated. The gradient of the travel time curve is determined from the instantaneous phase section calculated from the downgoing wavefield by means of edge detection well known from image processing.

The method is illustrated by synthetic and real examples of zero-offset VSP data.

Keywords: edge detection; inversion; seismic attributes; VSP

INTRODUCTION

Exploration geophysics is involved with finding a number of parameters that describe the types and states of rocks in a given region. Many tools and techniques are commonly used in this pursuit, among those being the vertical seismic profile (VSP). One of the prime purposes of this in situ survey is to find the wave velocity as a function of depth. This velocity is then used in the lithologic delineation and mapping process.

Most frequently the interval velocity function is calculated from the travel times of direct arrivals.

Unfortunately the velocity values estimated by this method may have significant errors because the travel times cannot be determined with proper accuracy neither by interactive nor by automatic methods (Balch and Lee 1984, Stewart 1984, Vermes 1984).

Methods using the whole wavefield in the VSP inversion yield more reliable results (Grivelet 1985, Mace and Lailly 1986, Ursin 1986). The proposed methods use basically an iterative modeling approach in which a processor defines an initial estimation of the acoustic impedance function, computes the synthetic VSP wavefield and then compares this calculated response with the actual VSP data. If the synthetic and the measured data do not agree within an accepted tolerance, then the impedance function is adjusted, and the process is repeated until the difference between the measured and calculated VSPs becomes - in a least squares sense - minimum. This method is called general linear inversion (GLI).

As the problem is nonlinear, the initial estimation of the acoustic impedance function has to be close enough to the real one in order to get the global - and not the local - minimum of the objective function to be minimized. Because the GLI method uses the dynamic features of the VSP data, instead of the velocity function, it is the acoustic impedance function which can be estimated and, on the other hand, the amplitudes of the VSP data have to be correctly reconstructed during data processing. Another disadvantageous feature of the GLI technique is that the calculation of realistic synthetic VSPs and the solution of the large sets of linear equations require extremely fast computers.

The inversion method described in this paper does not possess the above mentioned disadvantages. Amplitudes do not have to be reconstructed, the first break times do not have to be determined, calculation of synthetic VSPs, solution of large sets of linear equations and definition of an initial model are not needed. Careful wavelet filtering and wavefield separation are sufficient.

THE METHOD

The series of the first breaks on a VSP defines the travel time curve of the first arrivals. As it is well known, the gradient of this travel time curve is in a very close connection with the wave velocity: in the case of zero-offset VSP and horizontal layering the gradient and the velocity are equal.

The gradient of the travel time curve vs. depth function can be estimated by method of edge detection well known from image processing (see Appendix). On the VSP section which is considered as an image, the gray levels change rapidly in the vicinity of the first arrivals because of the large amplitude and the high signal to noise ratio. These rapid changes can be detected by edge detection operators. After edge detection the gradient of the travel time curve of the first arrival can be estimated by a statistical method using the gradient values in the vicinity of the first breaks. The statistical processing of the large number of gradient data involves that in addition of the estimated velocity values the error of the estimation can also be given. In this paper the most frequent value procedure is used because of its robustness and resistance (Ferenczy 1988, Ferenczy et al. 1988, Steiner 1988).

Experiences obtained in the tests of the above described method suggest that edge detection should be carried out on the instantaneous phase section (Taner et al. 1979), calculated from the VSP rather than on the original VSP data. The reason for this is explained in Fig. 1. Figure 1a shows a Klauder wavelet with peak frequency of 40 Hz repeated five times, and Fig. 1b is the instantaneous phase of this section. It is evident that in the same time interval the instantaneous phase section contains sharper "jumps" than the original section of Klauder wavelets. As a consequence the gradient values obtained from the instantaneous phase data are more accurate than that calculated from the original VSP data.

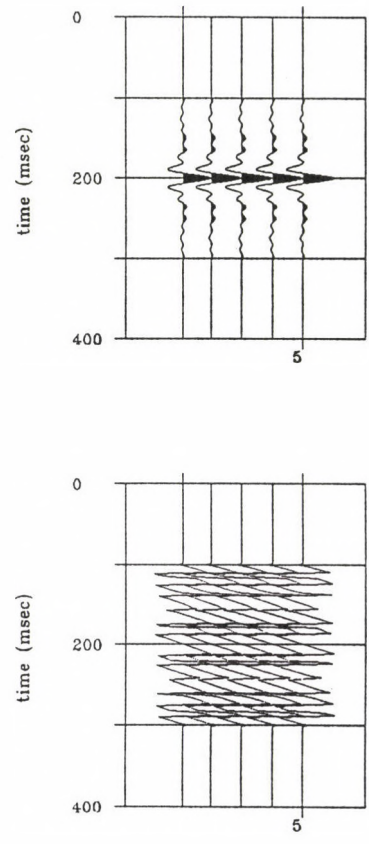


Fig. 1. Top: Klauder wavelets with the peak frequency of 40 Hz. Bottom: The corresponding instantaneous phase

SYNTHETIC EXAMPLES

The features of the inversion method described in the previous paragraph were studied on synthetic data. By means of the following examples we examine how the method is influenced by random noise and model errors.

The velocity function used in synthetic zero-offset VSP calculation is shown in Fig. 2. The model consists of 1000 layers. Each layer is 2 metres thick.

The calculated synthetic VSP can be seen in Fig. 3. The section consists of 149 traces, the distance between the

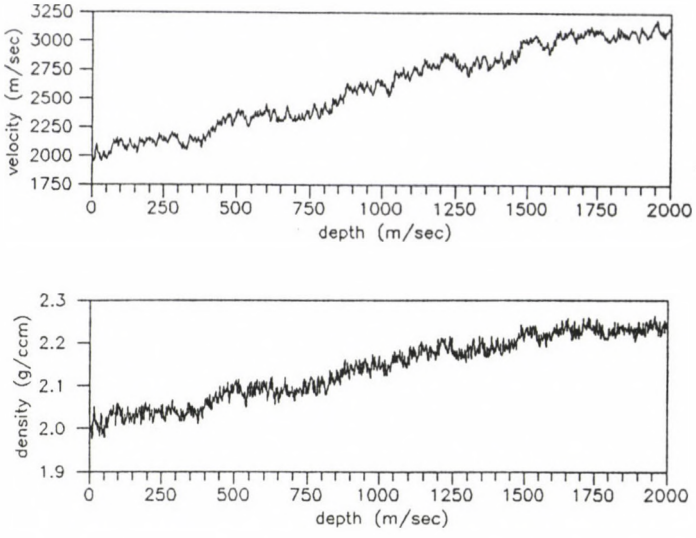


Fig. 2. Velocity and density functions for calculating synthetic VSP

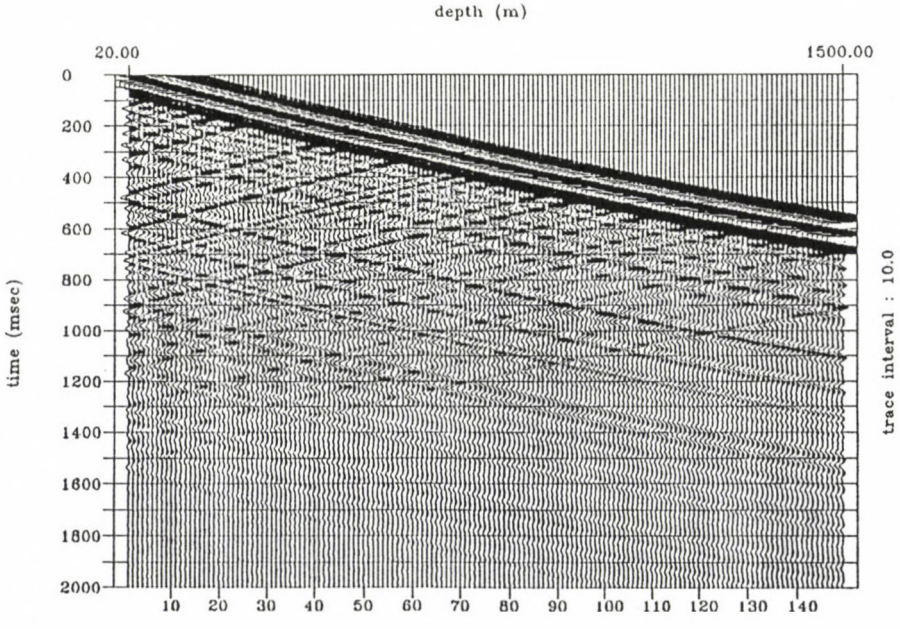


Fig. 3. Synthetic zero-offset VSP calculated from the velocity model shown in Fig. 2 (wavelet: Klauder with the peak frequency of 30 Hz; source depth is 10 m)

equally spaced geophones (traces) is 10 metres, time sampling is 2 msec, the depth of the source is 10 metres. The calculated impulse response is convolved by a Klauder wavelet with the peak frequency of 30 Hz.

The downgoing wavefield used in velocity estimation and the corresponding instantaneous phase section are shown in Figs 4 and 5, respectively. According to our expectation, the instantaneous phase section contains much sharper "jumps" than the original VSP, so the gradient data, demanded by the velocity estimation procedure, were calculated from the instantaneous phase section. For the velocity estimation those gradient data were used which were found in 200 msec wide time intervals around the first breaks. The final results are shown in Fig. 6. (The estimation uncertainty shown in the lower part of Fig. 6 means that the true velocity values are between the lower and the upper limits of error with a probability of about 66 percent.)

In the case of noise-free VSP the estimated velocity

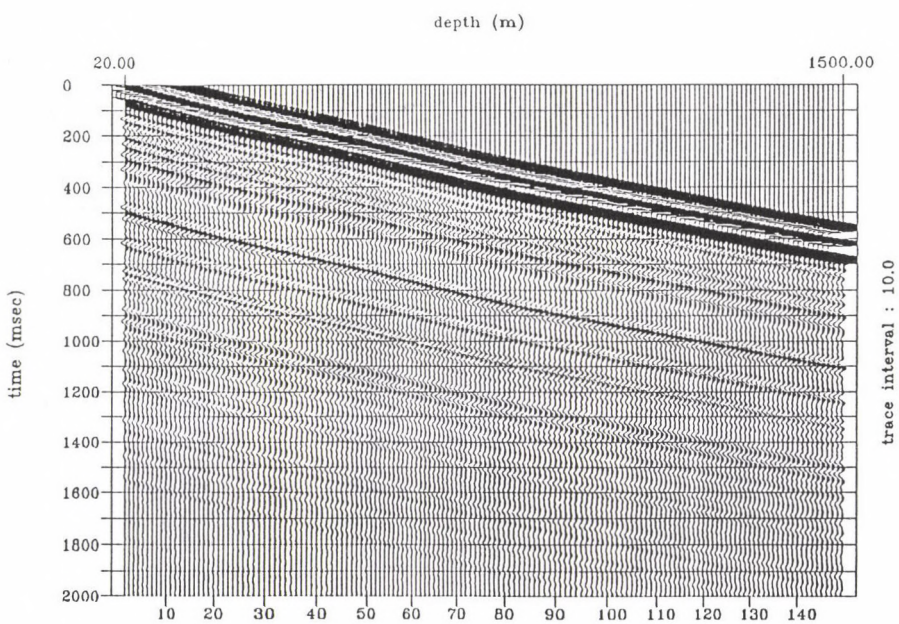


Fig. 4. Downgoing wavefield separated from the VSP shown in Fig. 3

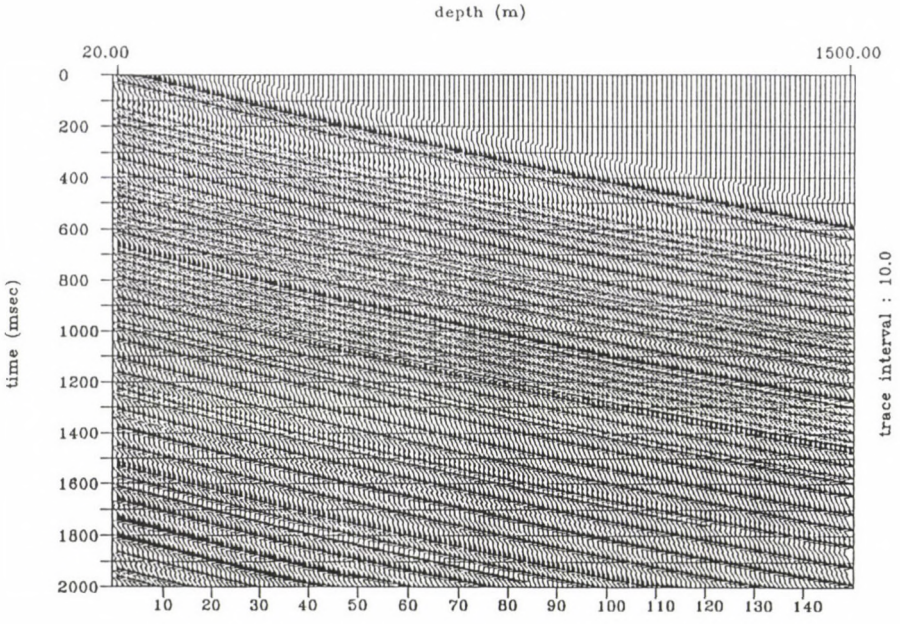


Fig. 5. Instantaneous phase section calculated from the downgoing wavefield

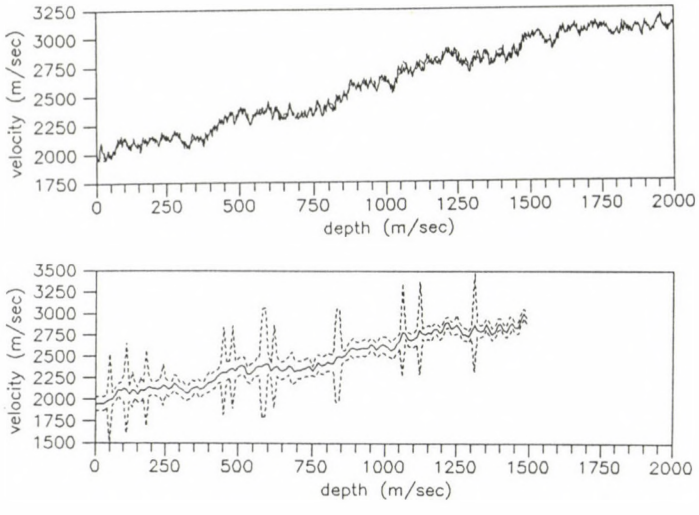


Fig. 6. Top: The true (solid line) and the estimated (dashed line) velocity functions. Bottom: The estimation uncertainty (solid line: estimated velocity function; dashed lines: the upper and the lower limits of error)

function is a very good approximation of the true velocity data. The mean difference between the two velocity functions is only 1.2 percent and the estimation uncertainty originating from the deviation of the gradient data is less than ±5-6 percent of the estimated values.

As was mentioned in the previous section, during the statistical processing of the gradient data the most frequent value procedure is used. Since this method is much more robust and resistant than the usual least squares method, the final results - mainly in the case of noisy VSP data - are also expected to be more reliable. In order to ascertain this expectation, the previous velocity estimation was carried out by application of the least squares statistics. The results (Fig. 7) are really more inaccurate than that obtained by the application of the most frequent value procedure (Fig. 6): the mean error of the estimation is 2.4 percent and the estimation uncertainty originating from the deviation of the gradient data has also increased.

In order to investigate how the proposed inversion method

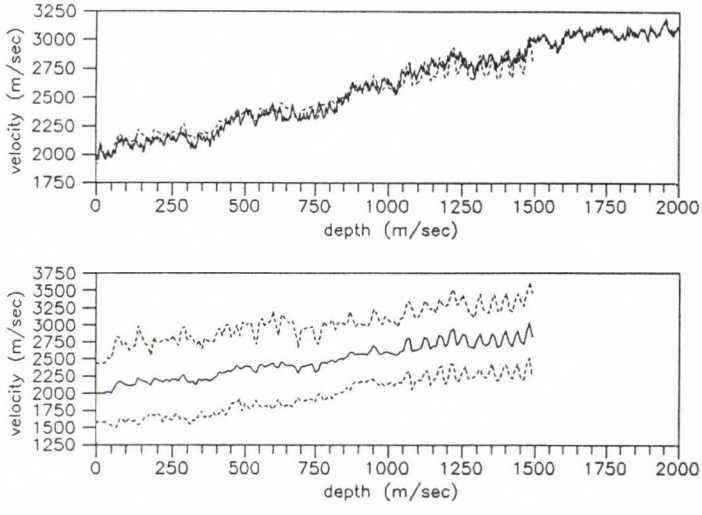


Fig. 7. Top: The true velocities (solid line) and the velocity function estimated by using the least squares statistics (dashed line). Bottom: The estimation uncertainty

is affected by additive noise, 5 and 10 percent, resp. coloured noise was added to the downgoing wavefield shown in Fig. 4; the mean amplitude of the noise was calculated from the maximum amplitude of the noise-free VSP, random numbers were added to each sample of the synthetic impulse response and the sum was convolved by a Klauder wavelet with the peak frequency of 30 Hz. After convolution white noise was not added to the section because during data processing this type of noise becomes coloured in any case. The downgoing wavefield corrupted by 10 percent coloured noise is shown in Fig. 8.

The inversion of VSPs corrupted by 5 and 10 percent noise gives the results shown in Figs 9 and 10, respectively. For the velocity calculation gradient data were used which were found in 200 msec wide time intervals around the first breaks.

In the case of 5 percent additive noise the estimated velocity data - except a few layers - approximate the true values fairly well (the mean difference is 1.7 percent) and the

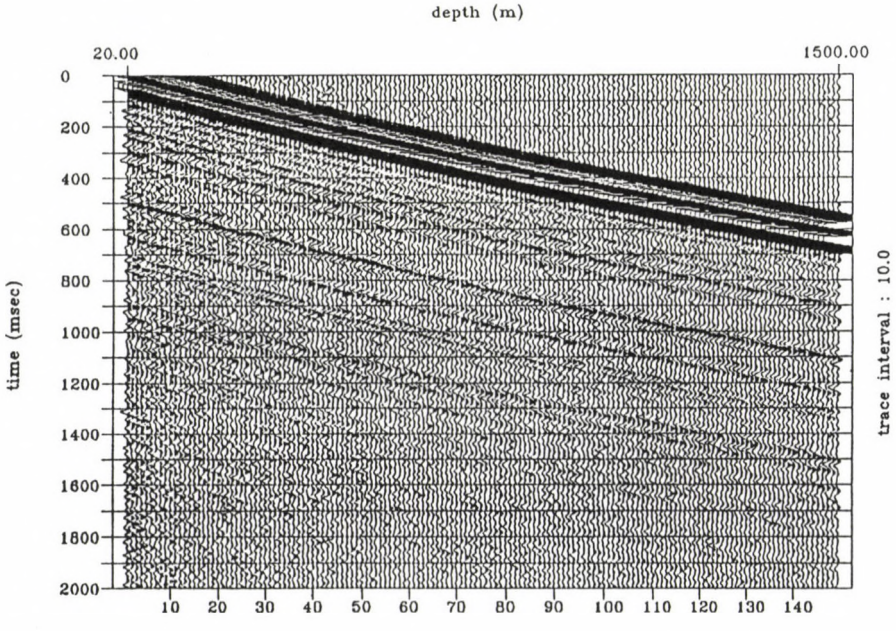


Fig. 8. The downgoing wavefield shown in Fig. 4 corrupted by 10 percent coloured noise

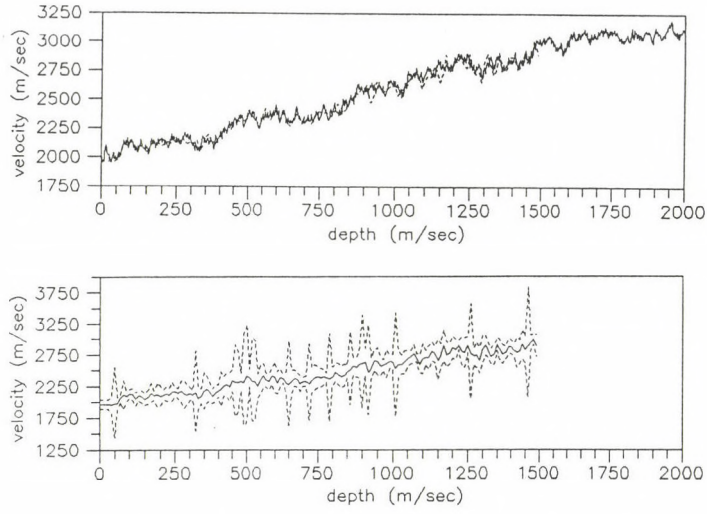


Fig. 9. Top: The true (solid line) and the estimated (dashed line) velocity functions in case of 5 percent coloured noise. Bottom: The estimation uncertainty

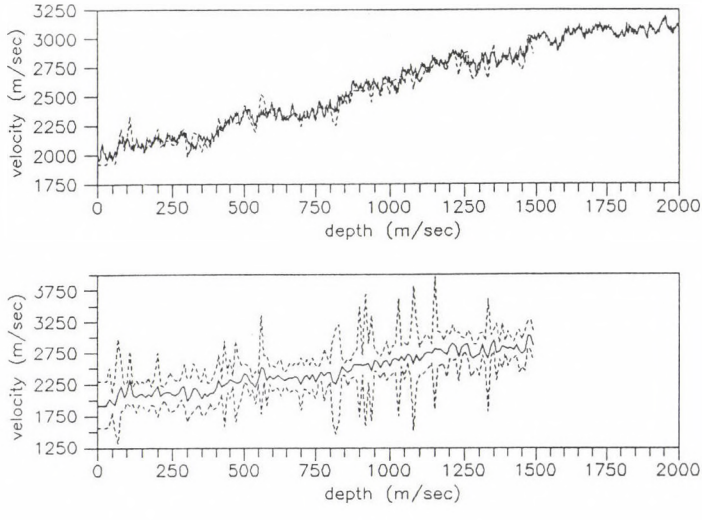


Fig. 10. Top: The true (solid line) and the estimated (dashed line) velocity functions in case of 10 percent coloured noise. Time gate: 200 msec. Bottom: The estimation uncertainty

estimation uncertainty originating from the deviation of the gradient data is not greater than ±10 percent in the main part of the depth interval. However, in the case of 10 percent noise the estimated velocity values differ considerably from the true velocity function (the mean difference is 2.4 percent) and the estimation uncertainty is between ±10 and ±20 percent in most layers. But this problem can be solved, if the time intervals which contain the gradient data used in velocity determination are chosen so short around the first breaks, that inside them the signal to noise ratio - because of the large amplitudes - is very high. If this time interval is only 50 msec long, the inversion of the VSP corrupted by 10 percent additive noise gives the fairly good final results shown in Fig. 11: the estimation uncertainty has considerably decreased and the mean difference between the estimated and the true velocity function has changed to 1.4 percent.

In the case of zero-offset VSP the gradient of the travel time curve of the first arrivals is equal to the wave velocity only if the medium is horizontally stratified. Thus it is

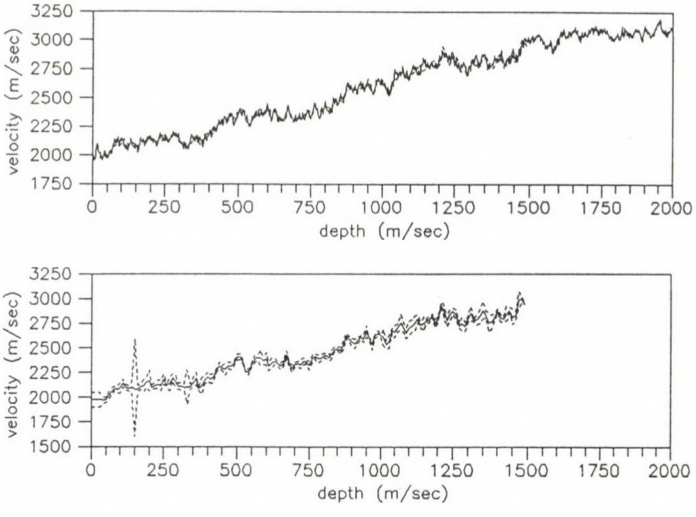


Fig. 11. Top: The true (solid line) and the estimated (dashed line) velocity functions in case of 10 percent coloured noise. Time gate: 50 msec. Bottom: The estimation uncertainty

desirable to investigate how the estimated velocity values differ from the true velocity function if the layer boundaries are not horizontal.

In order to carry out this investigation some two-dimensional geological models were defined with parallel, but not horizontal layers for which the velocity function shown in Fig. 12 is valid at the borehole. The synthetic VSPs were calculated by ray tracing, with the conditions that the offset is 10 metres, the depth of the uppermost geophone is 100 metres, geophone spacing is 10 metres and the wavelet is a damped sine function with the peak frequency of 30 Hz.

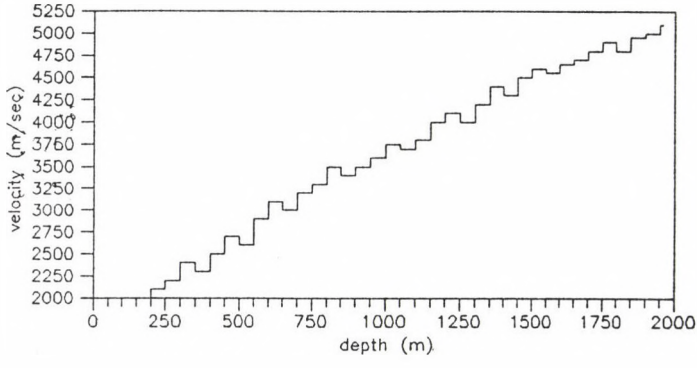


Fig. 12. Velocity function for two-dimensional geological model with parallel, but not horizontal layers

The inversion of synthetic VSPs calculated from the model of layers with dips of 10 and 20 degrees gives the final results shown in Figs 13 and 14, respectively. For the velocity calculation those gradient data were used which were found in 200 msec wide time intervals around the first breaks. In the case of layers with dips of 10 degrees the estimated velocity values - except for a few layers - are practically equal to the true velocity function (the mean difference is 0.9 percent) and the estimation uncertainty is only considerable at a few layers - usually where the estimation is inaccurate (Fig. 13). In the case of layers with dips of 20 degrees (Fig. 14), the difference between the estimated and the true velocity values

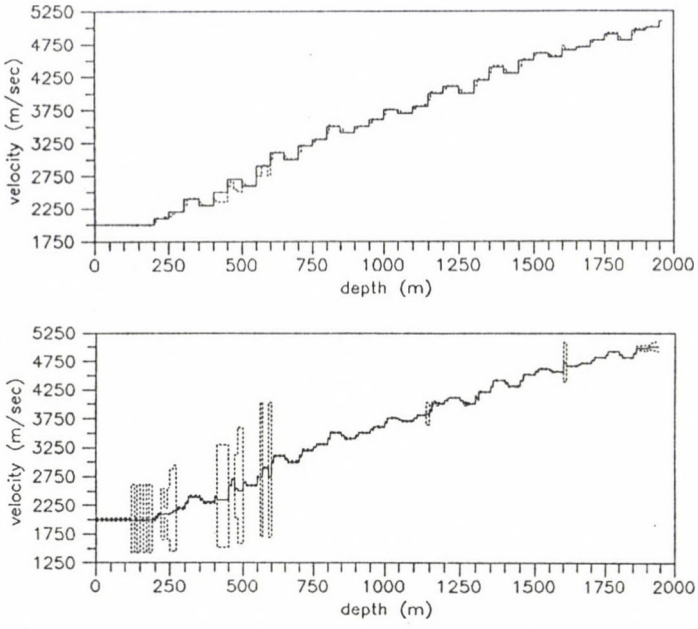


Fig. 13. Top: The true (solid line) and the estimated (dashed line) velocity functions in case of dipping layers. Dip = 10 degrees. Bottom: The estimation uncertainty

is greater than in the previous case, still its mean value is only 1.2 percent. The estimation uncertainty if again only considerable at a few layers, i.e. the error of the estimated velocity function which increases with depth can only be explained by the dip of the layers have not been taken into account during the inversion procedure. After all, the inversion method described in this paper gives reliable results when the dip of the layers is - from a practical view of point - large.

Remark: If the true velocity function consists of relatively thick layers, it can be observed that the estimated velocity function follows the true velocity changes (velocity "jumps") by insertion of one or two additional steps, even if the estimation is very good. It is because the edge detection operator, during the calculation of the necessary differences, determines weighted average of the adjacent instantaneous phase data (see Appendix): where the travel time curve of the first

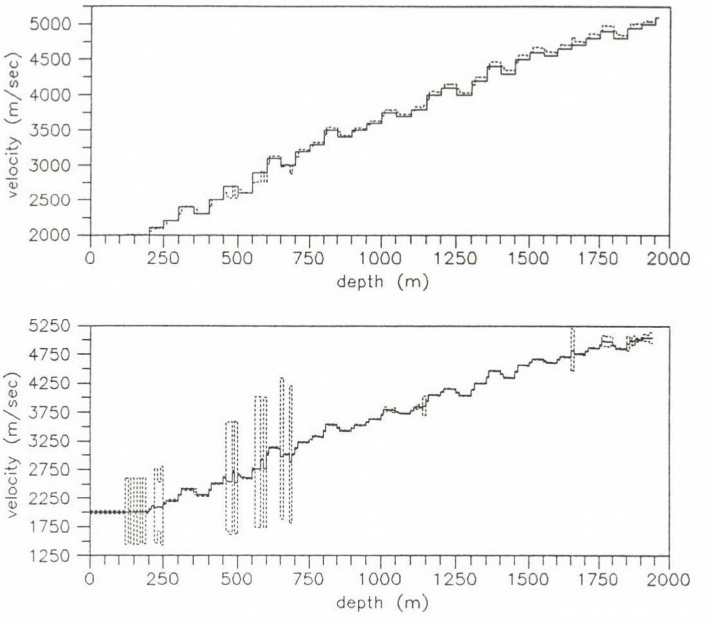


Fig. 14. Top: The true (solid line) and the estimated (dashed line) velocity functions in case of dipping layers. Dip = 20 degrees. Bottom: The estimation uncertainty

arrivals is broken - i.e. where the velocity changes rapidly - the edge detection operator rounds the travel time curve off. Because of this round-off, the gradient data do not change rapidly, so the estimated velocity function cannot change rapidly, either.

APPLICATION TO FIELD DATA

In this section the proposed inversion method is illustrated with a real example of zero-offset VSP measured in Hungary by the Geophysical Exploration Company (GKV). The GKV has made available for us the separated downgoing and upgoing wavefields, the velocity function estimated from the first break times and the acoustic log.

The downgoing wavefield separated from the zero-offset VSP measured in borehole Dévaványa D-1 and the corresponding instantaneous phase section are illustrated in Figs 15 and 16,

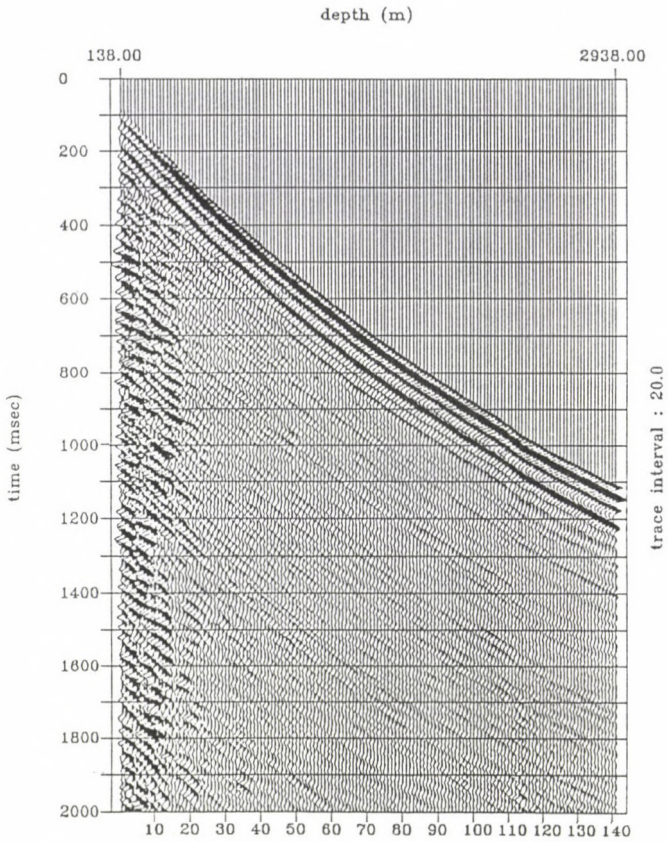


Fig. 15. Downgoing wavefield separated from the VSP measured in borehole Dévaványa D-1

respectively. The first 20-25 traces of the downgoing wavefield are fairly noisy, but on the instantaneous phase section sharp edges can be found in the vicinity of the first arrivals. Accordingly, in the velocity estimation those gradient data were used, which were found in 200 msec wide time intervals around the first breaks.

The velocity function calculated from the acoustic log (solid line) and that estimated from the first break times (dashed line) are shown in Fig. 17. The estimated velocity values differ considerably from the measured velocity data at many depths because of the erroneous estimation of the first

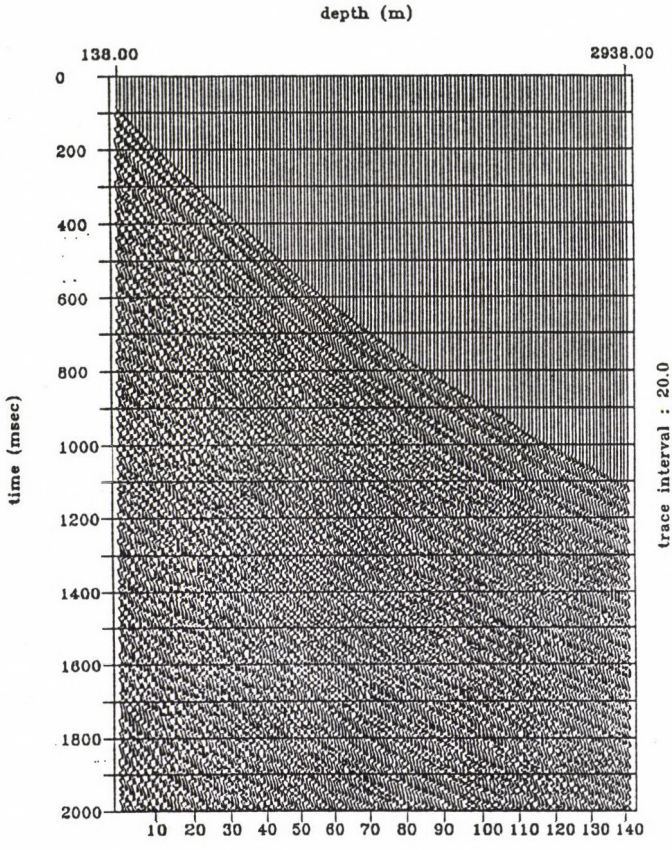


Fig. 16. Instantaneous phase section calculated from the downgoing wavefield shown in Fig. 15

break times: the mean difference between the two velocity functions is 11.7 percent, the largest difference is 104 percent.

The final results of the proposed inversion technique are shown in Figs 18 and 19. Figure 18 shows the velocity function calculated from the first break times (solid line) and that given by the inversion (dashed line). In Fig. 19 the estimated velocity values (dashed line) are compared with the measured ones (solid line): the mean difference between the two velocity functions is 7.7 percent, the largest difference is 33.2 percent. These values prove that the proposed inversion method

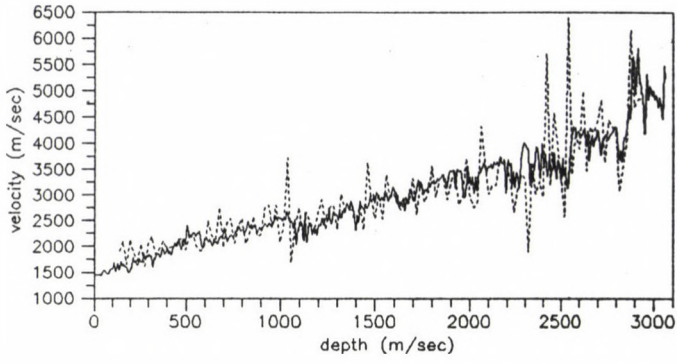


Fig. 17. The velocity log (solid line) and the velocity function estimated from the first break times (dashed line)

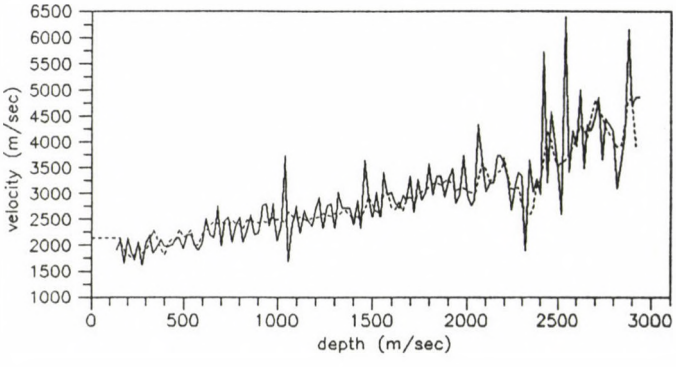


Fig. 18. The velocity function estimated from the first break times (solid line) and the inversion result (dashed line)

gives more accurate results than the velocity estimation procedure based on the first break times.

On the basis of the last two figures it can be observed that low frequency part of the inversion result and that of the velocity function estimated from the first break times are very similar. However, in the depth intervals of 2300-2400 m and 2700-2800 m the inversion results differ considerably from the measured velocity values. In other words, the two velocity functions estimated by two different methods from the same VSP

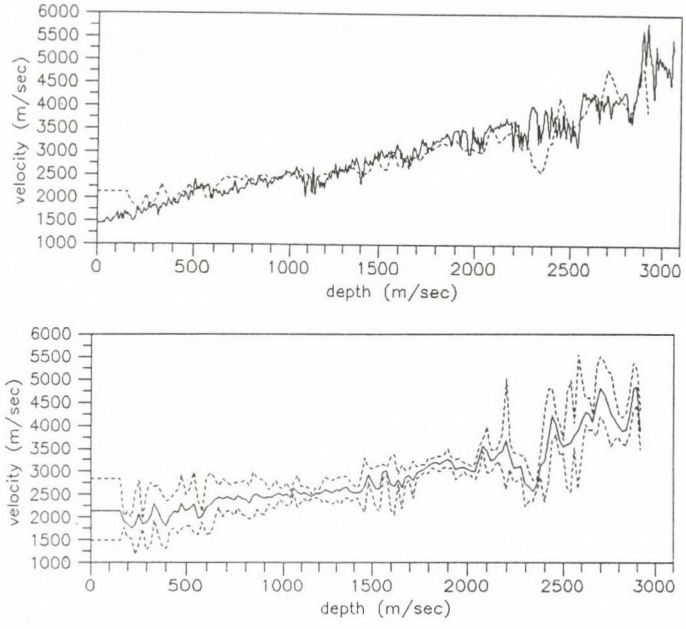


Fig. 19. Top: The velocity log (solid line) and the estimated velocity function (dashed line). Bottom: The estimation uncertainty

are similar, but - at least in some depth intervals - they are not in accordance with the acoustic log. For lack of sufficient information the reason of this observation cannot be explained.

SUMMARY

In this paper a new method for VSP inversion is described. After careful wavelet filtering and wavefield separation the gradient of the travel time curve of the first arrivals as a function of depth is estimated, on the basis of which the velocity function can also be estimated. The gradient of the travel time curve is determined from the instantaneous phase section calculated from the downgoing wavefield by means of edge detection well known from image processing. This new inversion procedure gives more accurate results then the method based on the first break times and needs much less computing time than the GLI: it takes only a few minutes on IBM PC.

APPENDIX

Edge detection

In image processing difference operators are used for sharpening the pictures. These operators give large values at those places where the gray levels are changing fast, while their response is close to zero at areas with slowly changing gray levels. The difference operators can be considered highpass filters that amplify not only the edges but the high frequency noises as well, thus it is advisable to smooth the picture before applying them.

An edge is defined as a local property of an image, where the gray level is relatively consistent in each of two adjacent, extensive regions, and changes abruptly as the border between the regions is crossed (Rosenfeld and Kak 1982). It involves that an edge detector applied to the instantaneous phase section will find the zero crossings of the signals.

In an image $g(x,y)$, if the rate of change of the gray levels in two perpendicular directions are $\partial g/\partial x$ and $\partial g/\partial y$, then the rate of change in an appropriate ϑ direction is

$$\frac{\partial g}{\partial x'} = \frac{\partial g}{\partial x}\cos\vartheta + \frac{\partial g}{\partial y}\sin\vartheta .$$

The rate of change is maximal in the

$$\vartheta = \arctan\frac{\partial g/\partial y}{\partial g/\partial x}$$

direction, and its magnitude is given by

$$m = \sqrt{(\partial g/\partial x)^2 + (\partial g/\partial y)^2} .$$

The vector whose direction is ϑ, and whose magnitude is m, is called the gradient of g at the location of (x,y).

In digital pictures we compute the first differences instead of the first derivatives:

$$\Delta_x g(x,y) = g(x,y) - g(x-1,y)$$

$$\Delta_y g(x,y) = g(x,y) - g(x,y-1)$$

which define the

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

convolution masks. Its should be remarked, however, that the first difference operator is not symmetric, and it is very sensitive to the high frequency noises to which it responds approximately as strong as to the edges.

The undesired responses to noise can be reduced if the picture is locally smoothed before taking the differences, i.e. if we apply an

operator which computes the difference of local averages (Rosenfeld and Kak 1982). the simplest operator having these properties is the Prewitt operator whose response is given by the convolution of the image with the two symmetric digital filters:

$$s_H(x,y) = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad s_V(x,y) = \frac{1}{3}\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

where s_H is sensitive to the horizontal edges and s_V perceives vertical edges on the image. The transfer functions of the Prewitt masks are given by

$$S_H(f_x,f_y) = \frac{2}{3}\cdot j\cdot \sin 2\pi f_y\cdot(1 + 2\cdot\cos 2\pi f_x)$$
$$S_V(f_x,f_y) = \frac{2}{3}\cdot j\ \sin 2\pi f_x\cdot(1 + 2\ \cos 2\pi f_y)\cdot$$

The transfer function of the first Prewitt mask is shown in Fig. 20a. Note that the transfer function does not disappear on the borders of the Nyquist interval.

We can apply weighted averaging in the local smoothing, too. This is what the most commonly used Sobel operator realizes which is defined by the convolution masks:

$$s_H(x,y) = \frac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad \text{and} \quad s_V(x,y) = \frac{1}{4}\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}\cdot$$

The filter renders larger weights to the points nearest to the center of the mask, therefore its response is not weakened so much as that of the Prewitt operator. The transfer functions of the Sobel filters are given by

$$S_H(f_x,f_y) = j\cdot\sin 2\pi f_y\cdot(1 + \cos 2\pi f_x)$$
$$S_V(f_x,f_y) = j\cdot\sin 2\pi f_x\cdot(1 + \cos 2\pi f_y)\cdot$$

The transfer function of the first Sobel mask - which disappears on the borders of the Nyquist interval - is shown in Fig. 20b.

The magnitudes and the directions of the edges on the image can be determined from the responses of the filters by the following formulas:

$$E_{magn}(x,y) = \sqrt{R_H^2(x,y) + R_V^2(x,y)}$$

$$E_{dir}(x,y) = \operatorname{atan}\frac{R_V(x,y)}{R_H(x,y)}$$

where $R_H(x,y)$ and $R_V(x,y)$ denote the response of the Sobel operator at the location (x,y) on the image, and E_{magn} and E_{dir} are the magnitude and the direction of the edge, respectively.

Using these two edge properties, two different images can be constructed; one for the edge magnitudes and the other for the edge

directions.

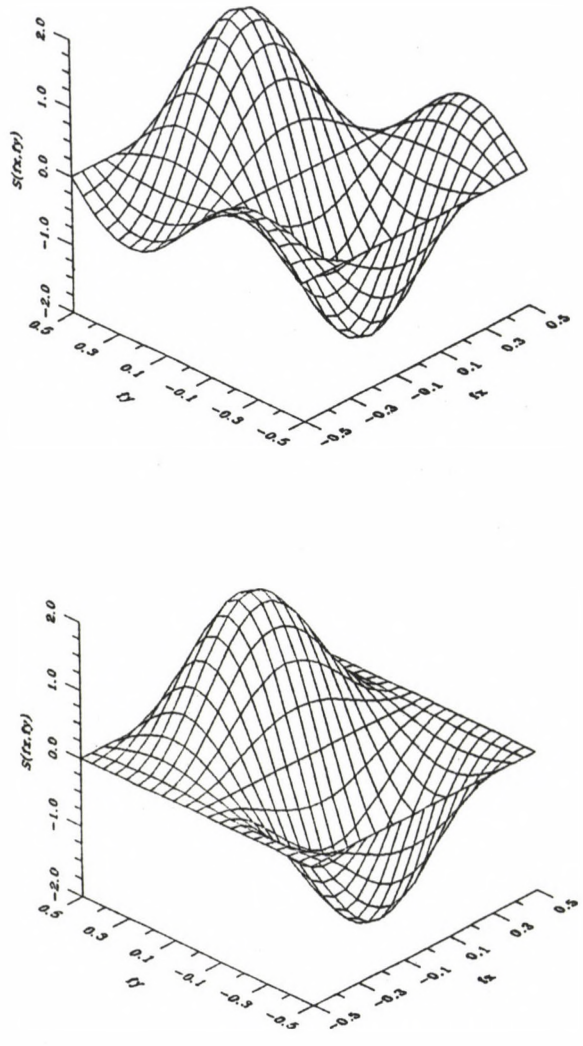


Fig. 20. Top: The transfer function of the Prewitt operator. Bottom: The transfer function of the Sobel operator

REFERENCES

Balch A H, Lee M W 1984: Vertical seismic profiling: Techniques, applications and case histories; IHRDC

Ferenczy L 1988: Magyar Geofizika, 29, 83-94.

Ferenczy L, Hajagos B, Steiner F 1988: Magyar Geofizika, 29, 104-112.

Grivelet P A 1985: Geophysics, 50, 924-930.

Mace D, Lailly P 1986: Geophysical Prospecting, 34, 1002-1021.

Rosenfeld A, Kak A C 1982: Digital picture processing. Academic Press, New York

Steiner F 1988: Geophysical Transactions, 34, 139-260.

Stewart R R 1984: Geophysical Prospecting, 32, 608-628.

Taner M T, Koehler F, Sheriff R E 1979: Geophysics, 44, 1041-1063.

Ursin B 1986: Geophysical Prospecting, 34, 1213-1218.

Vermes M 1984: Magyar Geofizika, 25, 186-193.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 323-338 (1991)

PLANNING AND PROCESSING OF MULTIPLE COVERAGE OVSP DATA

Z Wéber

Department of Geophysics, Eötvös University, H-1083 Budapest,
Kun Béla tér 2, Hungary

[Manuscript received January 8, 1991]

In offset vertical seismic profiling (OVSP) usually one-fold coverage is achieved since during data acquisition the source position is essentially not changed. If the source offset is large compared to the depth interval to be investigated, the one-fold coverage results in laterally limited resolution during OVSP migration. It is obvious that in order to increase the resolution, as well as the signal to noise ratio, a multiple coverage method should be used. Multiple coverage can be obtained if in due course of data acquisition two or more OVSPs are measured in the same borehole with different source positions.

In this paper the advantages of using multioffset VSP data in VSP-CDP mapping are illustrated on synthetic examples. Sets of curves have been constructed for the approximate determination of the source offsets before the multioffset VSP experiment. The interactively iterative VSP-CDP mapping, used during the preparation of the examples, is also described.

Keywords: multioffset VSP; multiple coverage method; sets of curves; VSP-CDP mapping

INTRODUCTION

In surface seismic exploration the multiple coverage method has become standard practice. Multiple coverage is obtained when consecutive spreads are closer than half of the total spread length. Several reflections originating from the same - or approximately the same - reflecting point are detected. It implies that this method gives more information about the subsurface than the continuous coverage method. Signal-to-noise ratio can be significantly enhanced and additional information, such as velocities can be extracted.

In offset vertical seismic profiling (OVSP) usually one-fold coverage is achieved since during data acquisition the

source position is essentially not changed. If the source offset is large compared to the depth interval to be investigated, the one-fold coverage results in laterally limited resolution during OVSP migration (Chiu et al. 1987).

It is obvious that in order to increase the resolution, as well as to improve the signal to noise ratio, a multiple coverage method should be used. Multiple coverage can be obtained if in due course of data acquisition two or more OVSPs are detected in the same borehole with different source positions. Multioffset VSP data are useful in delineation of complicated structure, too (Chang and McMechan 1986), since it may happen that some part of the structure cannot be imaged with only one source position.

CRP traces can be added after dynamic and static corrections. In the case of VSP data a more sophisticated preparation called VSP-CDP mapping is necessary.

In this paper the advantages of using multioffset VSP data in VSP-CDP mapping are illustrated on synthetic examples. The planning of multioffset VSP field geometry - first of all the source positions - is also considered. The imaging method used during the preparation of the examples is shortly also described.

VSP-CDP MAPPING

VSP-CDP mapping is defined as a process that transforms the VSP traces into a vertically oriented seismic section that begins at or near the borehole and extends laterally to the farthest reflection point.

Early versions of the VSP-CDP mapping algorithms move the data of the upgoing wavefield to the reflection points whose coordinates are calculated by assuming straight raypaths and horizontally layered earth (Dillon and Thomson 1984). The result of the procedure is the image of the subsurface in the horizontal distance - depth (x-z) domain. In order to get uniformly spaced, vertically oriented image traces, a grid of stacking bins should be introduced into the x-z domain.

The interpretational value of an image, created by the procedure described above, is limited primarily by the fact that the image construction is based on the assumptions of horizontally layered earth and straight raypaths.

The VSP-CDP mapping algorithm, used during the preparation of the examples shown in this paper, overcomes many of the limitations, mentioned above. The interactively iterative modeling technique allows the user to define a complex, two-dimensional geological model; to calculate the reflected raypaths between the source and the geophones in this assumed model by ray trace modeling (Wéber 1990); and then to use the calculated arrival times - together with the real OVSP data - to construct an image of the subsurface. Since the earth model usually consists of only a few layers, the calculated travel times and reflection point coordinates are interpolated in such a way that they can be dealt with as continuous functions of time and space. Thus all data points in the real OVSP traces can be transformed into the x-z domain. The geological model then should be updated until satisfactory match is achieved between the real and the model-determined arrival times (Lines et al. 1984, Hardage 1985) or between the model and the image. Because of its greater accuracy, the latter method is used in this paper.

The flow chart of the interactively iterative VSP-CDP mapping procedure is shown in Fig. 1.

VSP-CDP MAPPING OF MULTIOFFSET VSP DATA

In order to illustrate the importance of multioffset VSPs in VSP-CDP mapping, three synthetic OVSPs have been calculated by ray trace modeling for the simple two-dimensional geological model shown in Fig. 2, and then the three sections have been transformed into the x-z domain by the method described in the previous section. Since the aim of this investigation is not to prove that the procedure can recover the geological structure, for the sake of simplicity the true geological model was used during the mapping procedure.

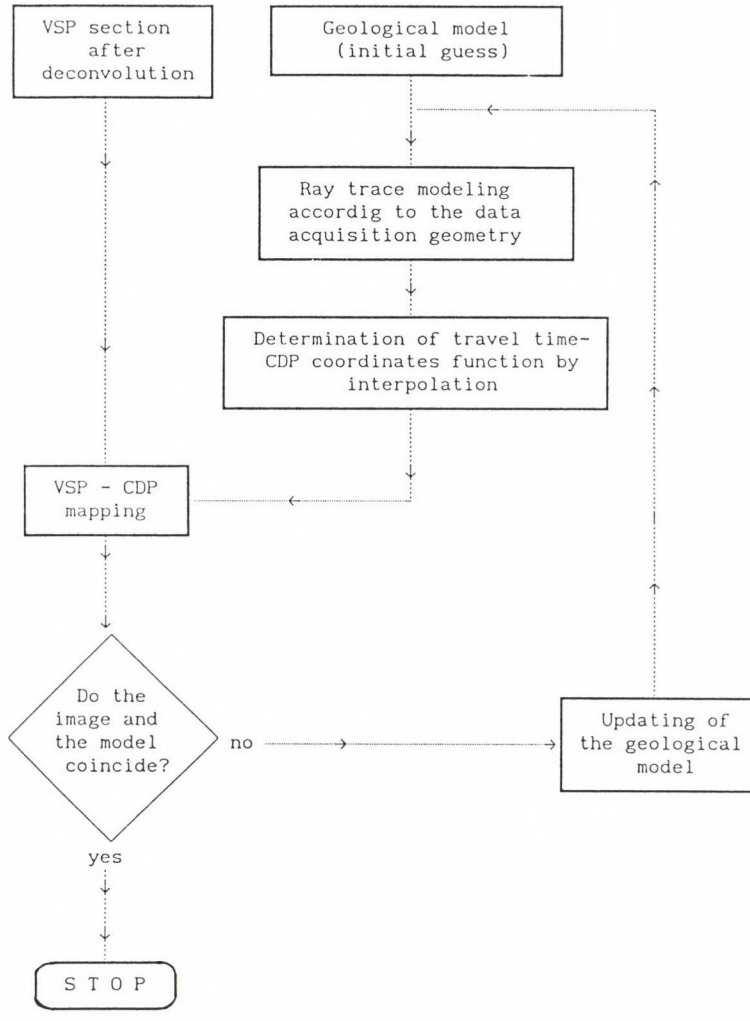


Fig. 1. Flow chart of the interactively iterative VSP-CDP mapping

Table I. Layer parameters of the model shown in Fig. 2

Layer No.	Velocity (m/sec)	Density (g/ccm)	Q
1	2000	2.0	100
2	2100	2.1	100
3	2200	2.2	100
4	2300	2.3	100
5	2400	2.4	100
6	2500	2.5	100
7	2600	2.6	100
8	2700	2.7	100

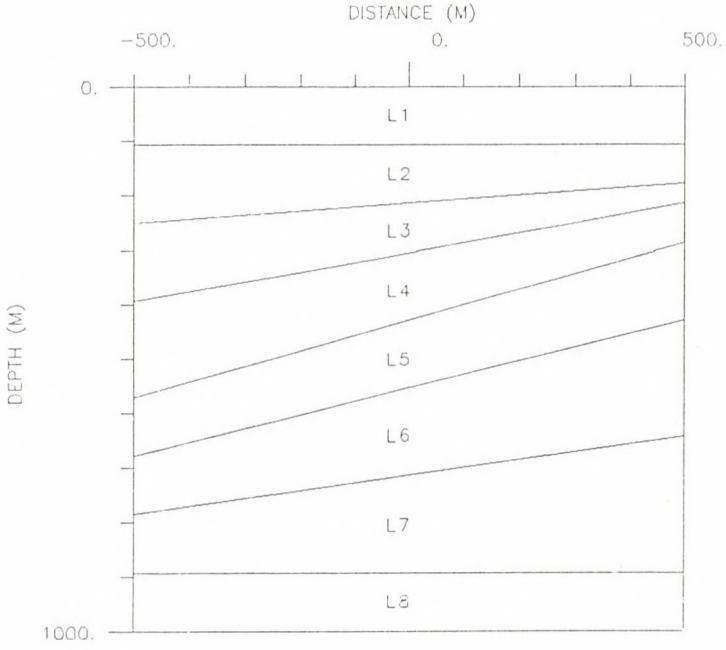


Fig. 2. A simple two-dimensional geological model. The layer parameters are shown in Table I

The three synthetics have been calculated for source offsets of 80, 200 and 300 metres. The reflection points calculated during the OVSP simulation and the VSP-CDP image for the source offset of 300 m can be seen in Fig. 3.

The results illustrate that the density of the reflection points is far from being uniform: for fixed-offset VSP it increases with depth and with the distance from the borehole. It is also obvious that the smaller the source offset, the higher the density of the reflection points in the vicinity of the well. As a consequence the stacking fold in a VSP-CDP mapping varies accross the subsurface image, since it is in direct proportion to the density of the reflection points. On the other hand, the quality of the VSP-CDP image is poor where the stacking fold is small, as illustrated in Fig. 4. The figure is based on the same input data as those used in creating the VSP-CDP image of Fig. 3, but the data in those bins, in which the stacking fold is less than a previously defined threshold value, were substituted by zeros. In Fig. 4 the threshold value is 3. The figure proves that the quality of the image, the density of the reflection points and the stacking fold are all closely related.

The above discussed features of the VSP mapping involve that the larger the source offset, the larger the horizontal extension of the image, but at the same time the stacking fold becomes smaller and smaller near the well resulting in poor image quality in this region. If the source offset is decreased, the image quality becomes acceptable everywhere, but the extension of the image becomes smaller.

In order to achieve both the desired extension and quality, VSP-CDP mapping of multioffset VSPs is necessary. The reflection points for the source offsets of 80, 200 and 300 metres can be seen together in the upper part of Fig. 5. Now the density of the collected reflecting points is acceptable both in the vicinity of and far from the well. It suggests that the superposition of the three different images gives an image with high resolution and good signal to noise ratio. The lower part of Fig. 5 shows the result of this superposition which is

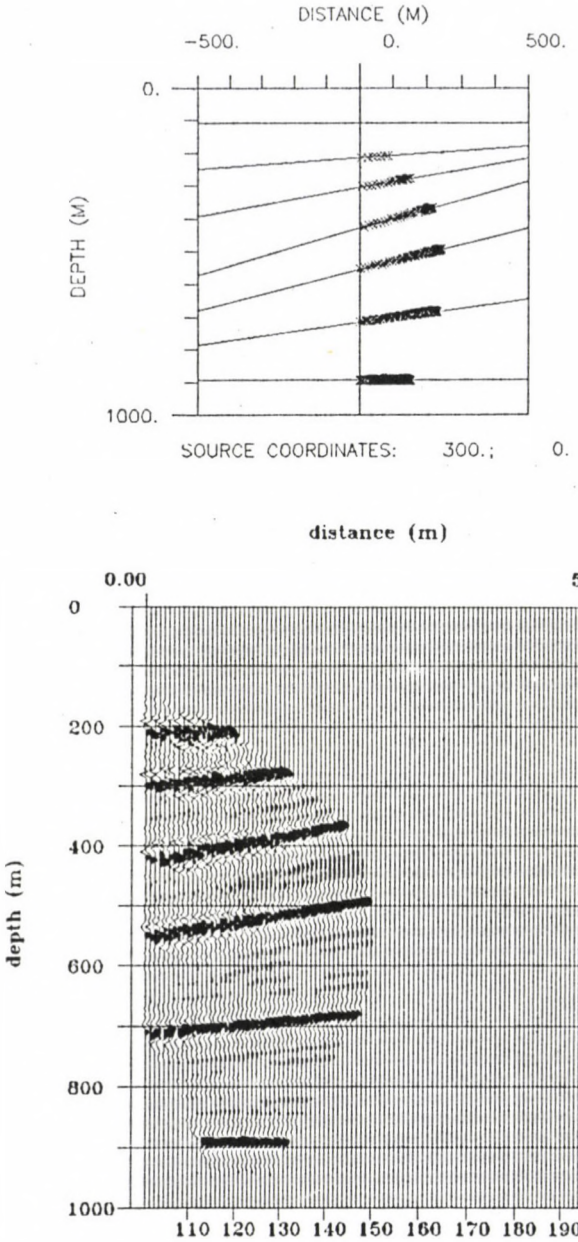


Fig. 3. Top: Reflecting points corresponding to the source offset of 300 m and geophone distance of 10 m. Bottom: VSP-CDP image calculated from the synthetic OVSP

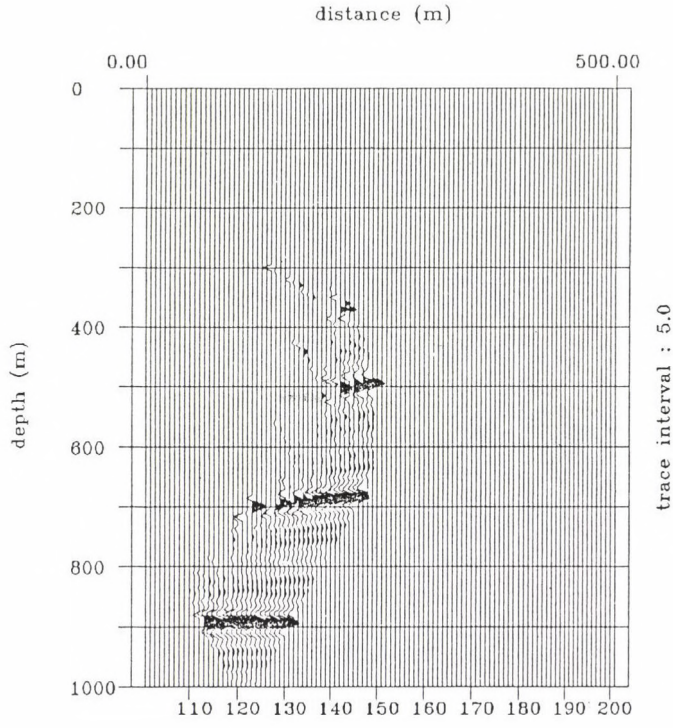


Fig. 4. As Fig. 3, but the data in those stacking bins, in which the stacking fold is less than 3, are substituted by zeros

really a high quality image of the subsurface.

Since the multioffset VSP method shows close similarity with the multiple coverage method used in surface seismic exploration - both of them detect reflections originated from the same region with different source positions - it may be called multiple coverage OVSP method.

DETERMINATION OF THE OPTIMUM MULTIOFFSET VSP FIELD GEOMETRY

The determination of the proper source offsets before multioffset VSP measurements is very important both from the economical and research point of view. In this section we investigate how to position the seismic sources in order for the stacking fold in the VSP-CDP mapping not to be less than a

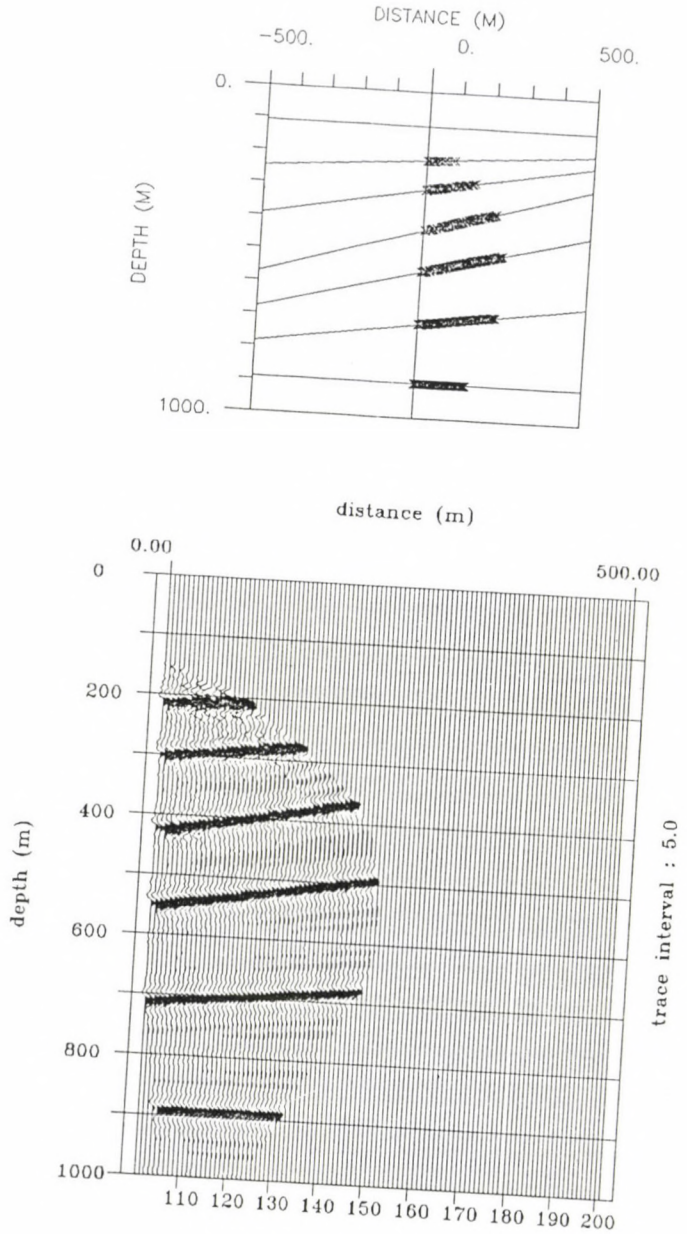


Fig. 5. Top: The reflecting points for the source offsets of 80, 200 and 300 m shown together. Bottom: The superposition of the three VSP-CDP images

previously defined threshold value.

Let us determine first the connection between the offset (L), the geophone depth (g) and the coordinates of the reflecting point (x and z). Assuming horizontal layering and straight raypaths (Fig. 6) we get that

$$x = \frac{z-g}{2z-g} L \ . \qquad (1)$$

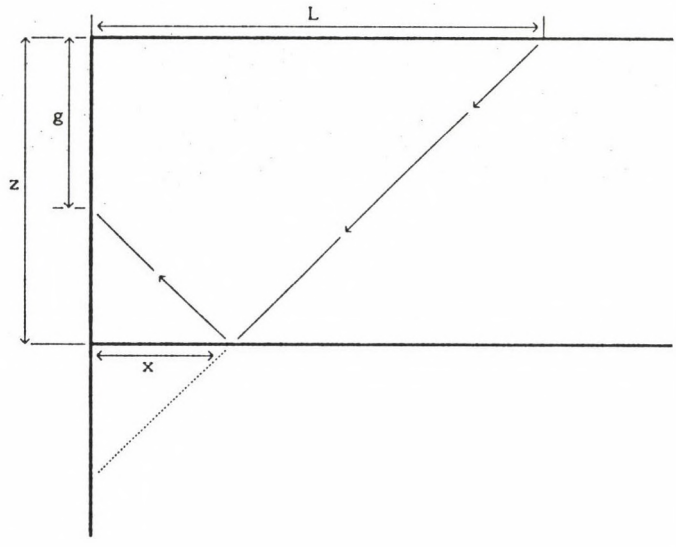


Fig. 6. OVSP data acquisition geometry

For fixed offset and depth Eq. (1) defines the x coordinate of the reflecting point as a function of geophone depth. If the geophone depth changes with Δg, then the absolute value of the variation of x is approximately

$$\Delta x = \left| \frac{dx}{dg} \right| \Delta g = \frac{zL}{(2z-g)^2} \Delta g \ . \qquad (2)$$

Expressing g from Eq. (1) and substituting it into Eq. (2), we obtain

$$\Delta x = \frac{(L-x)^2}{Lz} \Delta g \ . \qquad (3)$$

If inside the geological section to be imaged the distance between two adjacent reflecting points must not be greater than a previously defined value Δx_{max}, then we have the inequality

$$\Delta x_{max} \geq \frac{(L-x)^2}{Lz} \Delta g. \qquad (4)$$

Eq. (4) yields a quadratic inequality for the source offset L:

$$L^2 \Delta g - L(2x\Delta g + z\Delta x_{max}) + x^2 \Delta g \leq 0 \ . \qquad (5)$$

The equality in Eq. (5) holds if L is equal to either of the two roots of the polynomial in Eq. (5), while the inequality holds, if L is between these two roots.

It results from the above discussion that if inside the geological section to be imaged the distance between two adjacent reflecting points must not exceed Δx_{max}, then the source offset should be between the two roots of Eq. (5) and, at the same time, greater than

$$L_o = \frac{2z-g_{min}}{z-g_{min}} x_{max},$$

where g_{min} and x_{max} denote the depth of the uppermost geophone and the distance of the farthest image point from the borehole, respectively. The latter condition is necessary in order to get reflection from the farthest point to be imaged.

The analytical procedure described above is difficult to use in practice, because the alteration of only a single parameter requires the repeated solution of the quadratic inequality (5) and, on the other hand, in the case of multioffset VSPs Δx_{max} is the largest permissable distance between the reflecting points of all the OVSPs and, as a consequence, its value cannot be used directly in Eq. (5). In order to remove the above mentioned difficulties, sets of curves have been constructed.

Let us denote the grid size used during the VSP-CDP

mapping and the stacking fold at coordinates (x,z) by ΔX and N_{xz}, respectively. Taking Eq. (3) into consideration, it is obvious that

$$N_{xz} = \frac{\Delta X}{\Delta x} = \frac{Lz}{(L-x)^2} \frac{\Delta X}{\Delta g} . \qquad (6)$$

As it was found in the previous section, the larger the value of N_{xz}, the better the quality of the image.

Based on the Eq. (6) - assuming that $\Delta X/\Delta g = 1$ - sets of curves can be constructed. Each set refers to a given value of the parameter L. Such sets of curves are shown in Figs 7 to 9. The vertical and the horizontal coordinates denote depth and the lateral distance from the borehole, respectively. The solid lines connect the points of equal N_{xz}. The actual value of the stacking fold is equal to the number written beside the curve in question multiplied by the value of $\Delta X/\Delta g$. The dotted lines define the reflecting points corresponding to those geophone depths, at which the curves start.

In order to illustrate how to use the sets of curves during the determination of the source offsets before multioffset VSP experiment, let us set the aim to image the domain $x \leqslant 200$ m and $z \geqslant 200$ m on condition that the stacking fold is nowhere less than 2. Let us assume that the depth of the uppermost geophone is 50 m, the geophone distance and the grid size used during the mapping procedure are equal to 10 m. As a consequence, $\Delta X/\Delta g = 1$, i.e. the N_{xz} values written beside the curves can be used without modification.

According to the x coordinate of the farthest point to be imaged, the largest source offset should be 400 m. As it can be seen in Fig. 9, for this source offset the stacking fold is less than 2 in the main part of the region to be imaged and, at the depth of 200 m, the values of N_{xz} are between 0.5 and 1.5.

Since the stacking fold does not reach the predetermined minimum value anywhere at the depth of 200 m, the second offset should not be much less than 400 m, otherwise at the upper edge of the image domain a too wide zone with $N_{xz} < 2$ will occur. So let the next offset be 350 m. It can be read from the set of

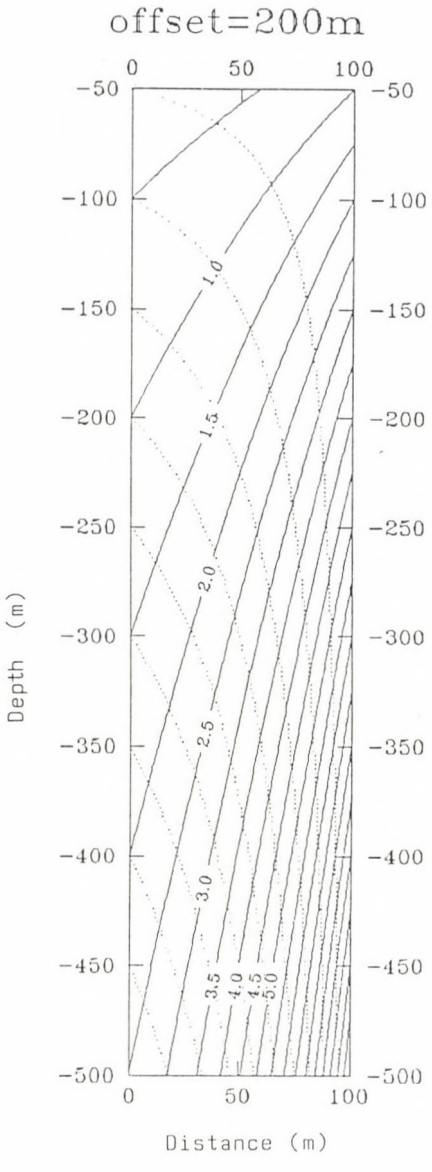


Fig. 7. Set of curves to determine N_{xz}. Source offset = 200 m

curves with the offset = 350 m (Fig. 8) that at the depth of 200 m the stacking fold is between 0.6 and 1.75. Because the N_{xz} values corresponding to two different source offsets should

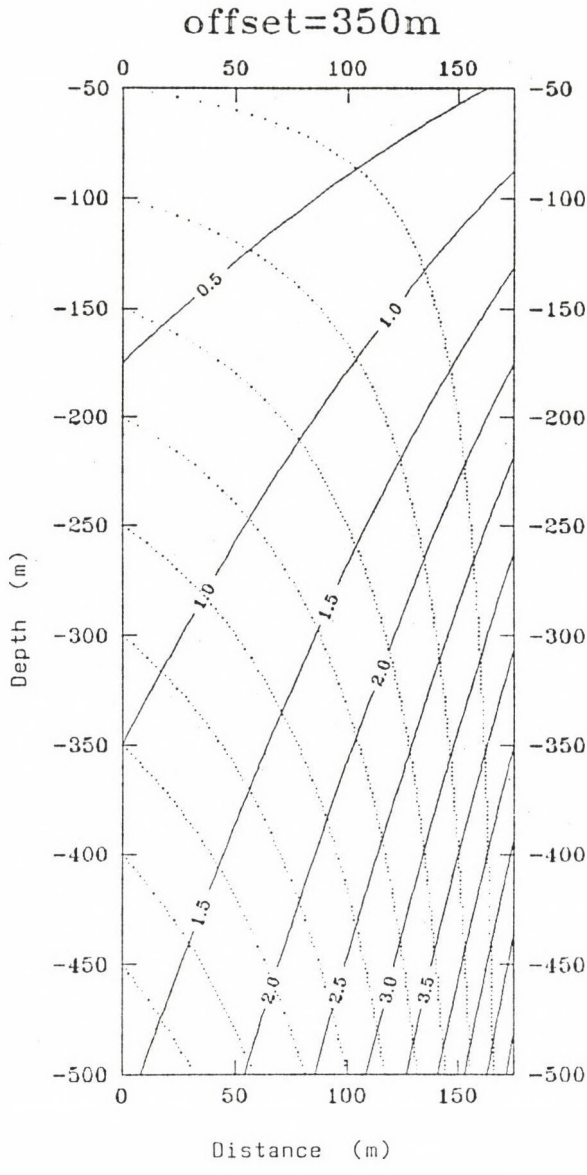


Fig. 8. Set of curves to determine N_{xz}. Source offset = 350 m

be summed up, the summarized stacking fold is between 1.1 and 3 at the depth of 200 m and is less than 2 for the x coordinates less than 90-100 m. This fact involves that a third offset

should also be used.

The third offset should be chosen so, that the farthest image point at the depth of 200 m is about 90-100 m far from

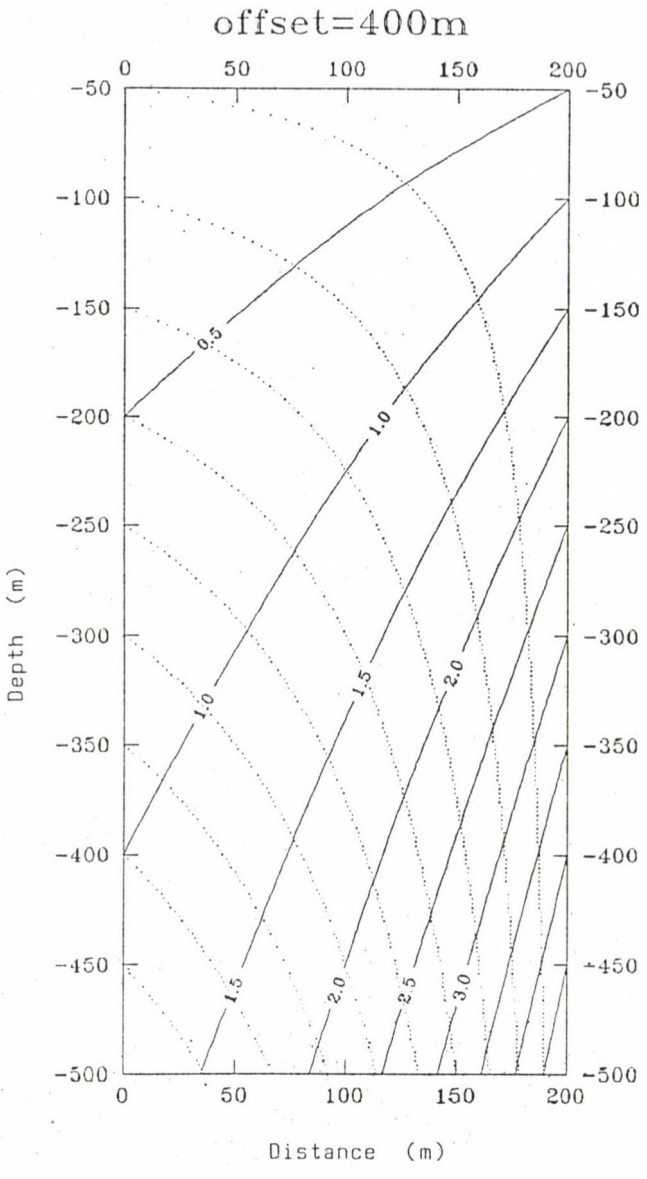


Fig. 9. Set of curves to determine N_{xz}. Source offset = 400 m

the borehole and the stacking fold is at least 0.9 near the borehole. The source offset of 200 m meets these requirements (Fig. 7).

Finally, it should be mentioned that the above described method yields approximate offset values only, because during the calculations horizontal layering and straight raypaths were assumed. If according to a priori information these approximations do not hold, the source offsets should be estimated by ray trace modeling using the trial-and-error method.

SUMMARY

In this paper the advantages of using multioffset VSP data in VSP-CDP mapping have been illustrated on synthetic examples. The examples have demonstrated that if the source offset is large compared to the depth interval to be investigated, high resolution can be obtained by VSP-CDP mapping of multioffset VSP data, only.

For the approximate determination of the source offsets before multioffset VSP experiment the use of sets of curves have been proposed and illustrated on a simple example. The example has proved that the sets of curves yield a useful tool for fast estimation of the offsets.

REFERENCES

Chang W, McMechan G A 1986: Geophysics, 51, 67-84.

Chiu S K L, Stewart R R 1987: Geophysics, 52, 1085-1098.

Dillon P B, Thomson R C 1984: Geophysical Prospecting, 32, 790-811.

Hardage B A 1985: Vertical seismic profiling, Part A: Principles, Handbook of Geophysical Exploration, Section I, Vol. 14A, Geophysical Press, London - Amsterdam

Lines L R, Bourgeois A, Covey J D 1984: Geophysics, 49, 250-264.

Wéber Z 1990: Ray trace modeling in seismic exploration. Geophysical Transaction (in press)

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 339-350 (1991)

A COMPARISON BETWEEN HORIZONTAL MAGNETIC AND VERTICAL ELECTRIC FIELDS DUE TO A VERTICAL ELECTRIC DIPOLE SITUATED WITHIN A COAL SEAM

E Takács

University Miskolc, Department of Geophysics, H-3515 Miskolc, Egyetemváros, Hungary

[Manuscript received January 13, 1991]

For the exploration of a horizontal coal seam a vertically oriented electric dipole produces the most favourable electromagnetic field. Due to the vertical dipole orientation both the vertical electric (E_z) and the horizontal magnetic (H_φ) components can be applied for the exploration of lateral inhomogeneities.

The goal of the study is to determine the role of the seam parameters and the adjacent layers in E_z and H_φ as a function of the frequency and the transmitter-receiver distance.

The most typical feature of the sounding curves is the increase of $|E_z|$ due to the seam which is much more significant than that of $|H_\varphi|$. Therefore $|E_z|$ is the component most recommended for the transillumination of the seam.

Keywords: horizontal magnetic field; mining geophysics; seam; underground frequency sounding; vertical electric dipole; vertical electric field

1. INTRODUCTION

Effects of the coal seam on the vertical electric field (E_z) due a vertical electric dipole (VED) situated within the seam have been studied by Takács et al. (1986), Takács (1988a, 1988b). Some features of E_z proved to be very promising to the detection of the inhomogeneities in the seam by a multifrequency transillumination method. These theoretical results have been justified by field measurement carried out in Hungarian coal mines.

A question arises: what is the situation with the magnetic field-component H_φ, which is also connected to the current-system in the seam. The question will be answered in this paper by a comparison between E_z and H_φ. In such a way their

advantages or disadvantages can be cleared, too.

The numerical investigations were carried out for an elementary VED having a unit moment, but the possibility of using a finite dipole instead of the infinitely small one will be presented, too.

2. HORIZONTAL MAGNETIC FIELD IN THE EQUATORIAL PLANE OF A VED IN HOMOGENEOUS SPACE

The dependence of the horizontal magnetic field on the transmitter-receiver distance (R), frequency (f) and resistivity (ϱ) can be comprehensively presented by the so called induction number (B). B turns out to be the ratio of the transmitter-receiver distance R to the skin-depth, which is determined by f and ϱ. In Fig. 1 $|E_z|$ and $|H_\varphi|$ fields are shown normalized to the DC field values as a function of B. The

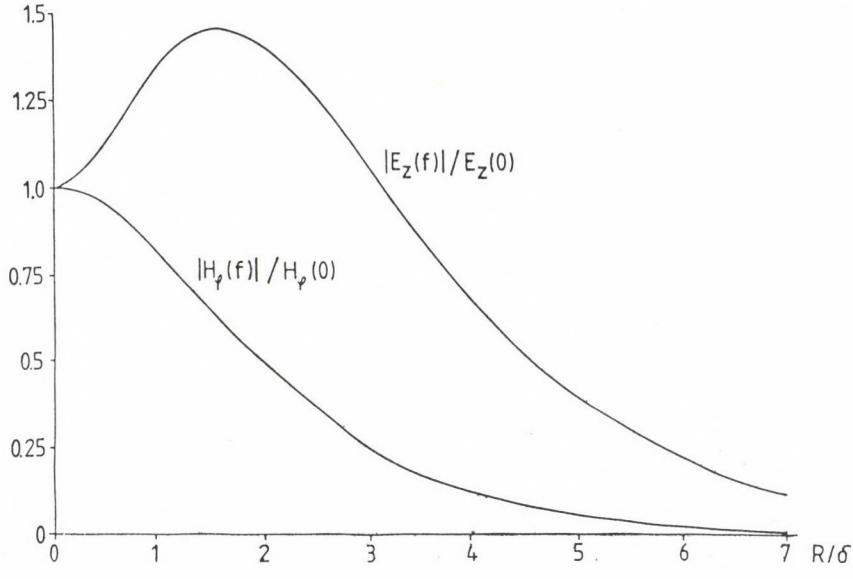


Fig. 1. AC amplitudes of the vertical electric and horizontal magnetic components normalized to their DC ones as a function of the induction number in a uniform whole space in the equatorial plane of a VED. $B = R/\tilde{\delta}$ (where R is the transmitter-receiver distance and $\tilde{\delta}$ is the skin depth)

vertical electric field is more influenced by the induction effects than H_φ.

Frequency sounding curves measured in a layered space shown in Fig. 1 will be used in the following for apparent resistivity computation. This procedure is based on the position of characteristic points: extreme values and zero-crossings along the frequency-axis. (Similarly to the case of E_z, for apparent resistivity computation the H_φ phase proved to be more convenient (Takács 1988a).)

3. HORIZONTAL MAGNETIC FIELD IN THE EQUATORIAL PLANE OF A VED WITHIN A COAL SEAM SITUATED IN A LAYERED SPACE

The horizontal magnetic field is presented for typical Hungarian three-, and four-layered coal mine sections. In the three-layered section the coal seam is the second one; the four-layered section corresponds to geological situations, where the Triassic limestone or dolomite basement is close to the coal seam. The transmitter and the receiver are situated in the middle part of the seam (ϱ_2 = 250 ohmm, h_2 = 5 m).

In Figs 2 and 3 amplitude and phase frequency sounding curves for H_φ and E_z are shown. In order to estimate the effect of the seam, sounding curves corresponding to the 20 ohmm homogeneous whole space are also presented.

According to Fig. 2 in the three-layered section the quasi-stationary values of $|H_\varphi|$ increased approximately by one order due to the effect of the seam. With increasing frequency the effect of the seam becomes more and more significant. However this increase is for $|H_\varphi|$ much less, than for $|E_z|$. (For $|E_z|$ this increase reaches about two orders in the quasi-stationary range.)

The $|H_\varphi|$ curves do not have - even in the presence of the seam - any maxima. However such maxima can be always found on the $|E_z|$ curves. Such a maximum is a very favourable feature, since it makes $|E_z|$ more measurable. The situation is similar on the descending part of the curves, because in the three-layered section the $|E_z|$ curve becomes less steep than that in

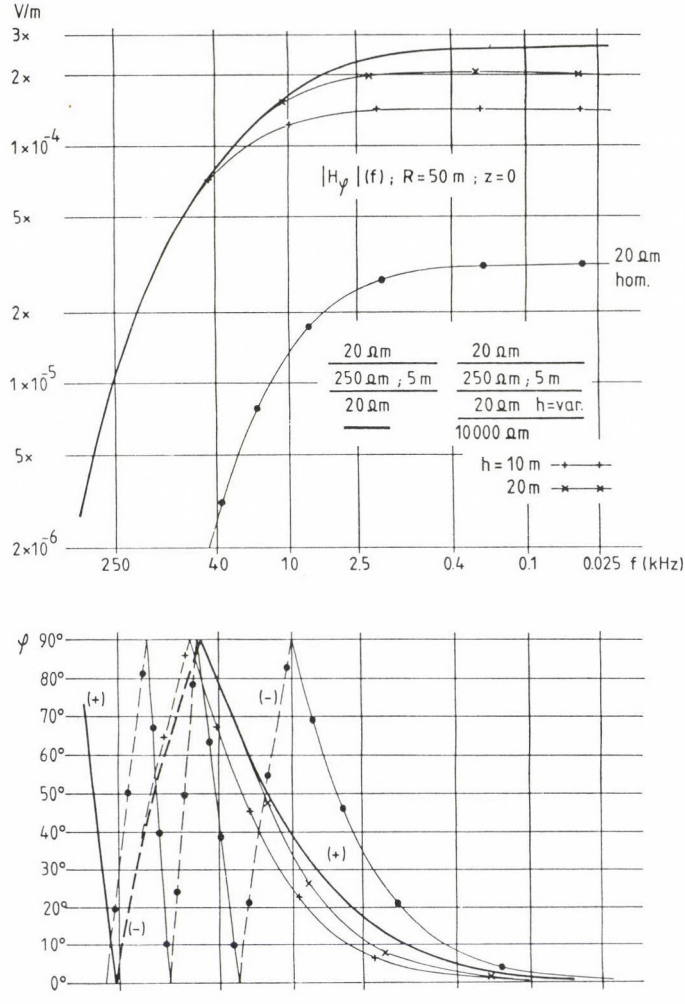


Fig. 2. Amplitude and phase frequency sounding curves of the horizontal magnetic field H_φ in the equatorial plane of a VED situated in the middle of the seam at a distance of 50 m

the homogeneous space. For example, at 40 kHz $|E_z|$ turns out to be higher, than at DC, while $|H_\varphi|$ is only about one-third of its DC value.

However, the situation with H_φ is more favourable regarding the phase. For example the E_z phase values are very small below 10 kHz and therefore the seam has an effect of only

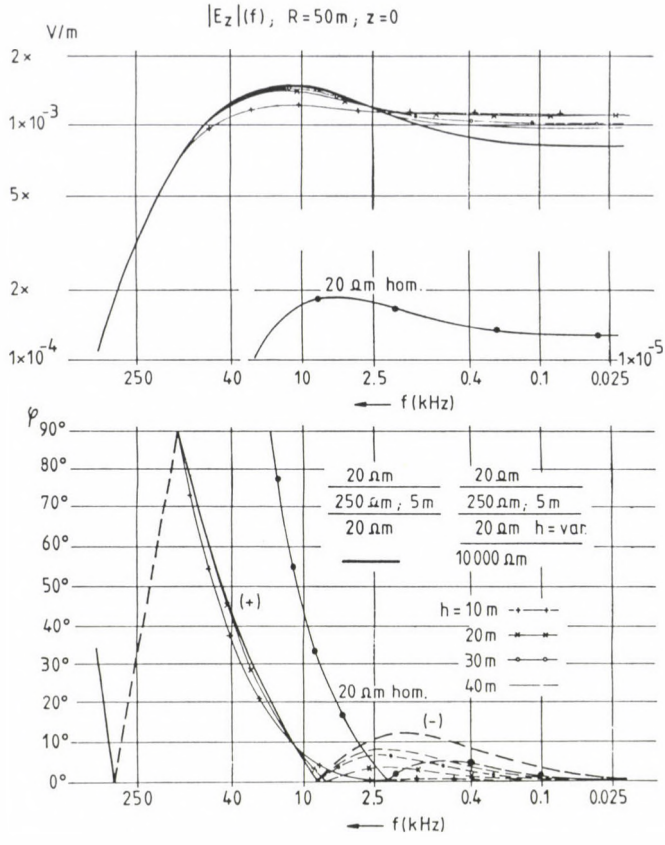


Fig. 3. Amplitude and phase frequency sounding curves of the vertical electric field E_z in the equatorial plane of a VED situated in the middle of the seam at a distance of 50 m

a few degrees. In the same frequency range the normal phase of H_φ has higher values and the seam produces a well detectable phase decrease.

On the basis of frequency sounding curves shown in Fig. 3 the effect of the fourth layer (a resistive basement) can be estimated, too. It is very peculiar that the maximum of the $|E_z(f)|$ curve is reduced due to the presence of the basement. This phenomenon is considered as an indication of the basement's closeness to the seam. At smaller frequencies this effect becomes more and more significant, since the current

will be distributed within a larger and larger volume. The low-frequency zone is shifted towards higher values, resulting in an amplitude decrease in the maximum zone. The response of $|H_\varphi|$ to the depth-changes of the basement is somewhat more pronounced.

Above 40 kHz practically there is a coincidence between the three- and four-layered curves. It means that the field-values at this R are influenced merely by 10 m thick stripes in the roof, and in the floor.

Due to geometrical sounding effects the response of the seam depends on R, too. Figure 4 shows the relative increase of the quasi-stationary $|H_\varphi|$ and $|E_z|$ values due to the seam situated in a 20 ohmm homogeneous whole space as a function of R.

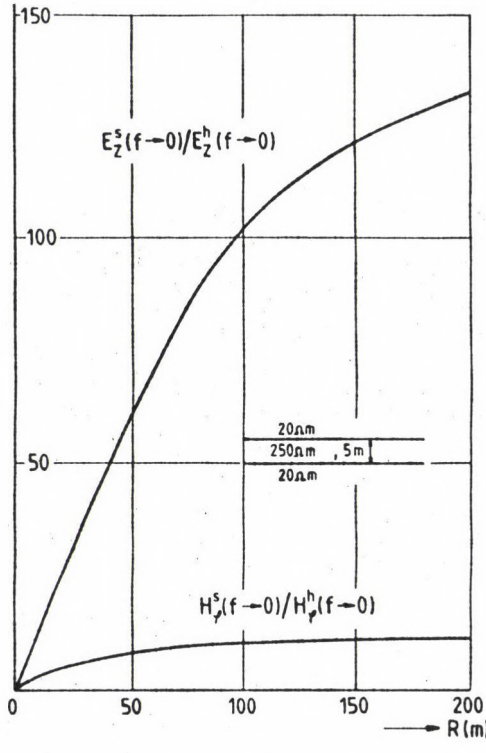


Fig. 4. Relative increase in the amplitude of the horizontal magnetic and vertical electric components in the equatorial plane of a VED in the middle of the seam due to the presence of the seam

From practical points of view it is an essential problem to study the effect of changes in layer parameters on the different field components. In order to get such an information the parameters of the three-layered section were alternatively increased by 10-10 p.c. and the amplitude-changes were calculated for different frequencies and transmitter-receiver distances. The result is given for $|H_\varphi|$ in Fig. 5 and for $|E_z|$ in Fig. 6. The absolute value is affected mostly by the seam resistivity, and this effect increases with increasing frequency and transmitter-receiver distance. An increase in the

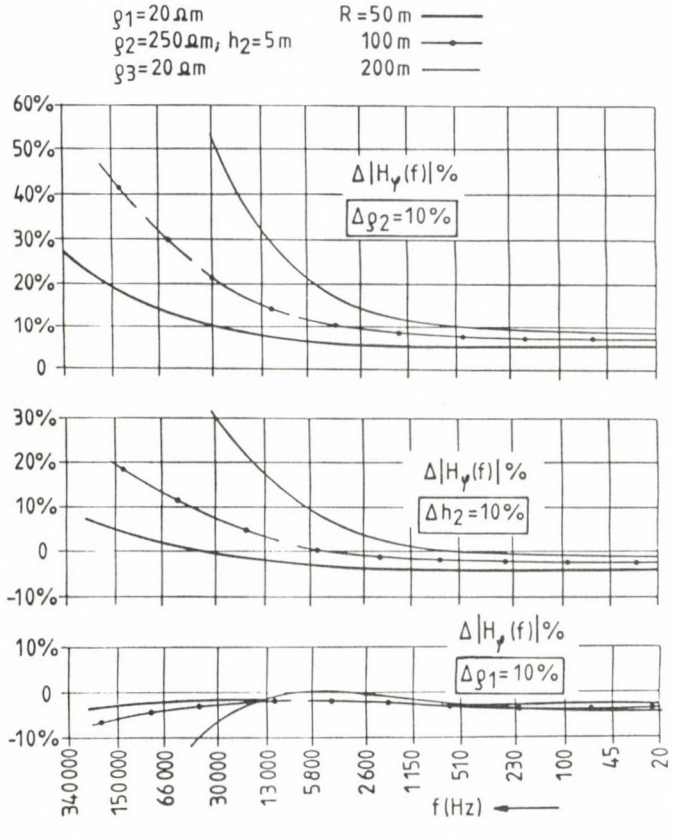


Fig. 5. Amplitude response of the horizontal magnetic component to a 10 p.c. change in different parameters of the given three-layered section, in the equatorial plane of a VED situated in the middle of the seam as a function of the frequency and the transmitter-receiver distance, in percents

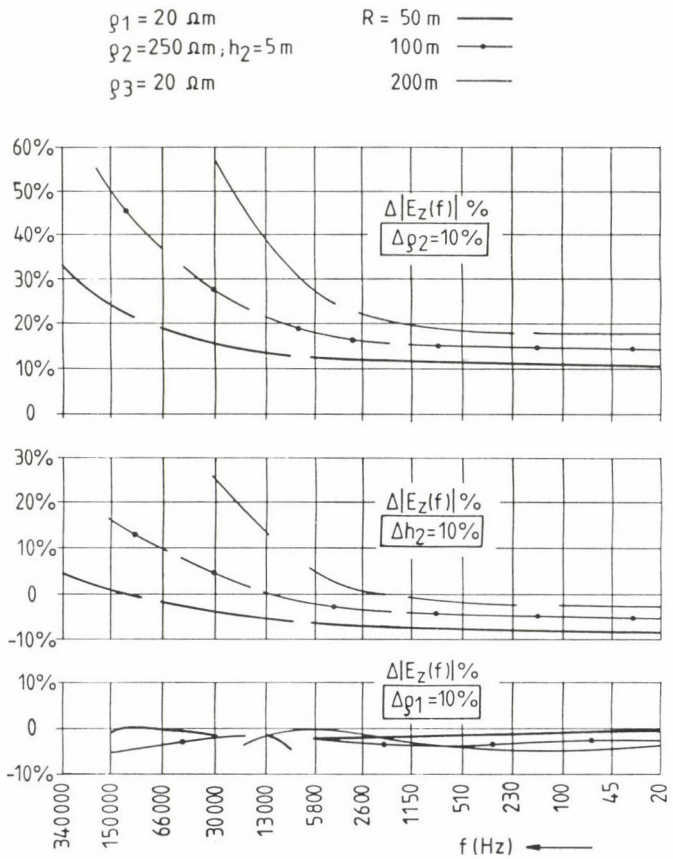


Fig. 6. Amplitude response of the vertical electric component to a 10 p.c. change in the parameters of the given three-layered section, in the equatorial plane of a VED situated in the middle of the seam as a function of the frequency and the transmitter-receiver distance, in percents

seam resistivity ϱ_2 causes an increase in both amplitudes. The changes in $|H_\varphi|$ and in $|E_z|$ are similar, however, in $|E_z|$ it is especially significant in the range 2.5-40 Hz. At low frequencies an increase in the seam thickness results in a decrease in both field intensities, while at high frequencies the effect is reversed.

Any increase in the roof or floor resistivity decreases somewhat the absolute values. In a certain frequency range depending on the layer parameters the effect due to the roof

and the floor is a minimum. This remarkable fact could be applied for the monitoring of physical processes within the high-resistivity seam.

As it has been mentioned, apparent resistivities can be derived on basis of Fig. 1 from the frequency sounding curves measured in a layered space. The apparent resistivity can be considered as an effective resistivity of that part of the layered space, which contributes to the electromagnetic field at a given frequency and transmitter-receiver distance. In Figs 7 and 8 the frequency-dependence of apparent resistivities calculated by using the curves in Fig. 1 for the three-layered section for several transmitter-receiver distances are shown. Increasing the frequency the current-system due to the VED will be contracted towards the equatorial plane. In this way the

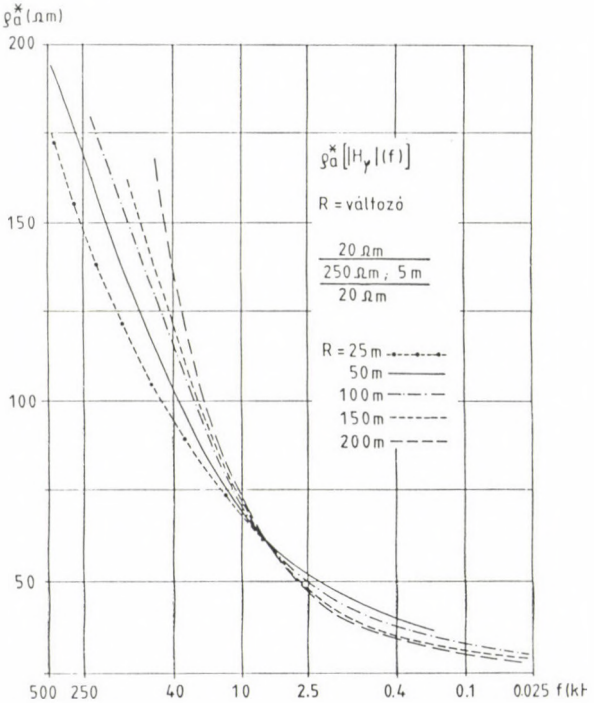


Fig. 7. Frequency dependence of the apparent resistivity for several transmitter-receiver distances calculated from the phase of the horizontal magnetic field in the equatorial plane of a VED situated in the middle of the seam

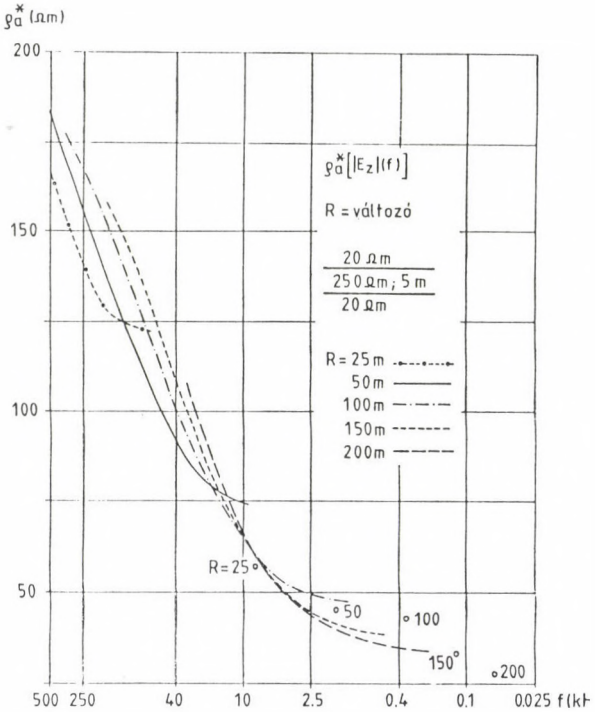


Fig. 8. Frequency dependence of the apparent resistivity for several transmitter-receiver distances calculated from the phase of the vertical electric field in the equatorial plane of a VED situated in the middle of the seam. Values denoted by circles belong to the amplitude maxima

relative effect of the seam in the apparent resistivity is more and more emphasized, leading to an R-dependent dynamic increase of the apparent resistivity toward higher frequencies shown in Figs 7 and 8. At larger R the frequency dependence is more significant, especially in the case of H_φ. It is a remarkable feature, that the apparent resistivity in a section-determined frequency range - in our case at 4-9 kHz - hardly depends on R. In exploration of seam inhomogeneities by the crossed-ray transillumination method, when R depends on the ray direction, this phenomenon can be well applied.

4. HORIZONTAL MAGNETIC FIELD DUE TO A VED HAVING A FINITE LENGTH

In the practice only dipoles having finite lengths can be realised. In order to have sufficient current intensity, the transmitter is usually grounded on the roof and the floor to electrodes placed in the more conductive adjacent layers. According to the present investigations the horizontal magnetic field due to a VED reaching to a small extent into the more conductive layers can be derived by the approximation method, used already for the E_z component (Takács 1988a).

The integral along the finite dipole (that is the integrated effect of the constituent elementary dipoles) can be substituted by a summation of the fields due to segments within the seam and out of it. It has been found, that the field in the middle part of the seam depends only slightly on the vertical position of an elementary VED within the seam. Out of the seam the field is reduced by a factor of $\varrho_{adjacent}/\varrho_{seam}$ and up to a moderate vertical distance - e.g. 0.5 m - it remains nearly constant. For this reason the horizontal magnetic field in the middle of the seam due to a finite VED (H^f_φ) can be derived from the horizontal magnetic field at the same site due to an elementary VED placed in the middle of the seam (H^e_φ). The horizontal magnetic field of a finite VED is as follows

$$|H^f_\varphi| \approx h_2|H^e_\varphi| + \Delta h_1(\varrho_1/\varrho_{seam})|H^e_\varphi| + \Delta h_3(\varrho_3/\varrho_{seam})|H^e_\varphi|$$

where - h_2 is the dimensionless thickness of the seam
- Δh_1, Δh_2 are dimensionless segments in the host rocks.

The smaller Δh_1, Δh_3, h_2, as well as the frequency, are and the larger R is, the more acceptable is this approximation.

5. CONCLUSIONS

A dipole-generated electromagnetic field is influenced to

the highest degree by the coal seam in case of a vertically oriented dipole. Around a VED both its vertical electric (E_z) and horizontal magnetic (H_φ) components are suitable for the exploration of lateral inhomogeneities. A morphological difference in their frequency-dependent amplitudes appears due to formation of a maximum on the $|E_z|$ curve between the quasi-stationary and the high frequency zones. This maximum increases with the increase of the seam resistivity and of the seam thickness. On the E_z phase curve a similar maximum appears at low frequencies, which is missing on the H_φ phase curve.

The seam parameters have much larger effect to both components than the parameters of the adjacent layers. With increasing frequency the effect of the seam gets more and more dominant.

Apparent resistivities calculated from the frequency-sounding curves increase considerably with the frequency, since the currents get more and more concentrated towards the equatorial plane, that is towards the seam.

For the two components there are some differences in the frequency- and R-dependence of the parameter-sensitivity and of the apparent resistivity, however, the trends in the case of both components are the same.

For determination of apparent resistivity values in the whole frequency-range the H_φ amplitude-curves seem to be more suitable, because - on the contrary to the $|E_z|$ curves - they do not have any maximum. The phase of H_φ is also measurable at lower frequencies, and the H_φ phase is characteristic to the layering. In spite of the above mentioned facts nevertheless $|E_z|$ is recommended for the lateral transilluminating exploration since its increased field intensity in the seam.

RFREFERENCES

Takács E 1988a: Geophysical Transactions, 34, 343-359.

Takács E 1988b: Acta Geod. Geoph. Mont. Hung., 24, 453-469.

Takács E, Nagy J, Mádai F 1986: Geophysical Transactions, 32, 43-56.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 351-363 (1991)

EXPLORATION OF ROOF AND FLOOR LAYERS BY THE HORIZONTAL ELECTRIC COMPONENT DUE TO A VERTICAL ELECTRIC DIPOLE SITUATED WITHIN A COAL SEAM

E Takács

University Miskolc, Department of Geophysics, H-3515 Miskolc, Egyetemváros, Hungary

[Manuscript received January 13, 1991]

According to one-dimensional numerical model computation the horizontal electric field E_x around a VED transmitter situated within the seam is measurable on its boundaries due to the signal-enhancing effect of the resistive seam. The E_x frequency-sounding curves are determined mainly by the parameters of the adjacent layers. The vertical extension of the section having influence on the horizontal electric field is a function of the frequency and the transmitter-receiver distance. This zone is broader at lower frequencies. Layers with higher resistivity contrast appear on the curves with local extrema. Therefore a possibility arises by measuring E_x frequency sounding curves at the upper and lower boundaries of the coal seam for the separate investigation of layers in the floor and the roof.

Keywords: coal seam; floor; horizontal electric field; layering; mining geophysics; roof; vertical electric dipole

1. INTRODUCTION

The tasks of geophysical measurements in a coal mine are twofold:

1. lateral exploration of the coal seam inhomogeneities and
2. vertically up- and downwards oriented exploration of layer inhomogeneities in the roof and the floor.

Vertical electric (E_z) and horizontal magnetic (H_φ) field components of a vertical electric dipole (VED) situated in the coal seam proved to be very sensitive indicators for lateral inhomogeneities in the coal seam (Takács 1991). At the same time these components are much less sensitive to inhomogeneities below or over the coal seam. Taking into account that E_z and H_φ determine a horizontal Poynting vector component, the horizontal electric field - since it contributes

in a vertical Poynting field component - could be applied in the vertically - up- and downwards - oriented exploration.

In a homogeneous whole space the E_x component is zero in the equatorial plane, and it remains also very small in its vicinity. For this reason is questionable whether this component is measurable or not. However in a layered medium the E_x field is considerably modified by the high resistivity coal seam. At the boundaries of the seam - or in their vicinity - the horizontal electric component increases to a well measurable value, so it becomes usable in exploration.

Results shown in this paper are based on 1-D numerical investigation. The numerical method was described by Takács et al. (1986).

2. HORIZONTAL ELECTRIC FIELD DUE TO A VED IN A THREE-LAYERED SECTION

Let the elementary VED having a unit moment be situated in the middle part of the most resistive second layer, the coal seam (its resistivity is ϱ_2 = 250 ohmm and its thickness is H_2 = 5 m). In the case when the section is symmetrical to the equatorial plane, the radial electric components at the upper and lower boundaries are considerably greater due to the presence of the seam, and at the same time these two radial field components have the same absolute values with opposite signs. The zero-crossing takes place just in the equatorial plane of the VED (and the coal seam). In the case of an asymmetric resistivity section, when the resistivities of the upper and lower half-spaces are not the same, the situations are concerning the absolute values and the zero-crossing site different.

In Figs 1 and 2 the change of the horizontal electric component is shown along a vertical line 50 m apart from the VED for five different frequencies, if ϱ_1 = 10 ohmm, ϱ_2 = 250 ohmm, h_2 = 5 m and ϱ_3 = 30 ohmm. In Fig. 1 the real and imaginary parts are presented. The zero-crossings of the real part are shifted from the equatorial plane towards the

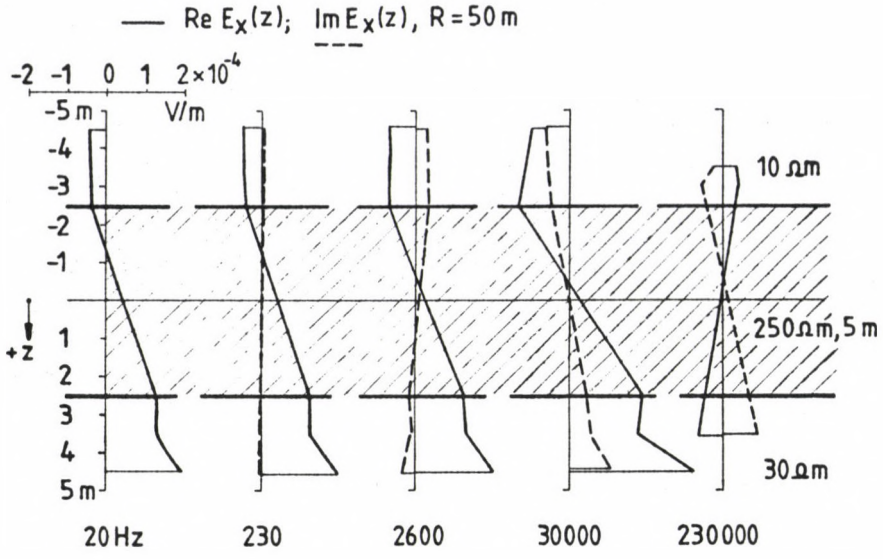


Fig. 1. Vertical variation of the real and imaginary parts of the horizontal electric component due to a VED situated in the middle part of the coal seam at a distance of 50 m, at different frequencies

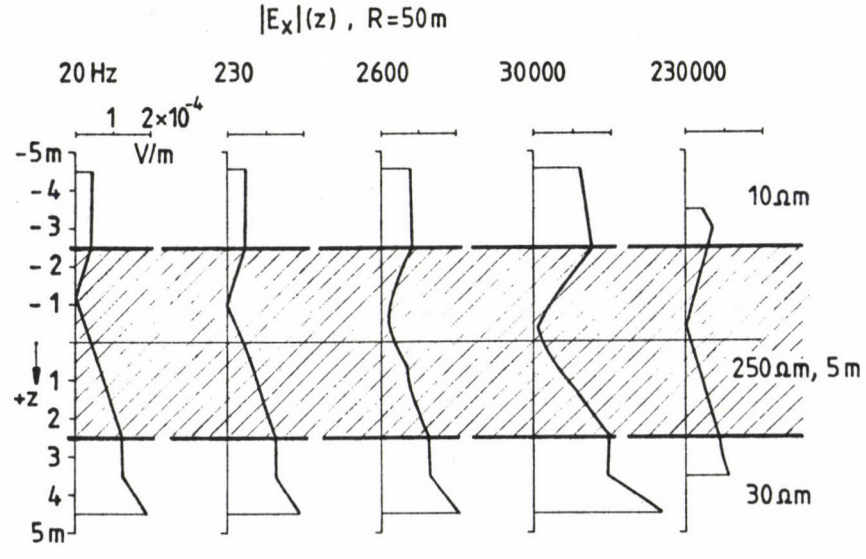


Fig. 2. Vertical variation of the amplitude of the horizontal electric component due to a VED situated at a distance of 50 m, at different frequencies

lower resistivity half space especially at lower frequencies. Increasing the frequency the zero-crossings approximate the equatorial plane. The horizontal field-component is smaller at the more conductive side. The behaviour of the imaginary part is just the opposite regarding both its zero-crossings and absolute values. At higher frequencies both parts - disregarding the sign - have the same behaviour due to their inductive origin. There are sign reversals in the real part at 230 Hz and in the imaginary part at 30 kHz. The sign reversal is connected to the fact, that at increasing frequencies the two opposite-phase current-systems are gradually shifted toward the transmitter. (Consequently a phase-boundary crossed the site 50 m apart from the VED.) Figure 2 shows the variation of the E_x amplitude along the vertical axis. Its character results from the above mentioned facts.

Let us see now the ratio of the horizontal electric field values on the lower (z=+2.5) and the upper (z=-2.5) boundaries.

Frequency	20	230	2600	30000	230000 Hz
Re $E_x(2.5)$/Re $E_x(-2.5)$	3.0	2.9	1.8	1.4	1.5
Im $E_x(2.5)$/Im $E_x(-2.5)$	0.06	0.11	0.34	0.96	1.6
$\|E_x(2.5)\|/\|E_x(-2.5)\|$	3.0	2.9	1.6	1.36	1.17

At low frequencies the current volume is larger than at high frequencies and therefore at low frequencies the contribution of the upper and lower half-spaces are much more significant to the horizontal electric field than that of the seam. For this reason the ratio of the real parts and the ratio of the absolute values practically coincide with the lower/upper half space (floor/roof) resistivity ratio. With increasing frequency the ratio of the electric fields decreases due to the increasing relative proportion of the coal seam in the whole current volume.

The volume - from which the information originates - depends on the transmitter-receiver distance R. Consequently the ratio of the horizontal electric fields changes with R. In the present case the situation is as follows:

R	25	50	100	150	200 m
$\|E_x(2.5)\|/\|E_x(-2.5)\|$	3.0	3.0	3.0	2.9	2.6

Due to the fact that a vertically oriented exploration usually needs only small R transmitter-receiver distances these results can be regarded satisfactory.

In Fig. 3 four frequency-sounding curves are shown: the z=-2.5 m, z=+2.5 m curves and for comparison two additional curves belonging to the two symmetric sections corresponding to the 10 ohmm and the 30 ohmm adjacent half-space resistivities. In the latter cases there are no differences between the curves at the z=-2.5 m and z=+2.5 boundaries. The curves become different only in case of the asymmetric section.

On the E_x amplitude curve after a quasi-stationary part a maximum appears, due to the high resistivity coal seam. (On the curves of the horizontal magnetic component there is not any maximum in spite of the fact that in the homogeneous space the $|E_x|$ and $|H_y|$ frequency sounding curves have the same form.) The curves belonging to the upper and lower boundaries are separated from each other sharply beginning from the maximum towards the quasi-stationary part according to the resistivities of the roof and the floor. However they do not coincide with the curves belonging to the two corresponding symmetric sections. The difference between the curves increases with the resistivity asymmetry. At higher frequency the effect of the roof and the floor is less significant, resulting in a decreasing difference between the curves. However, there is not any coincidence between the two curves, since even at high frequencies some effect of the adjacent half-spaces remains and

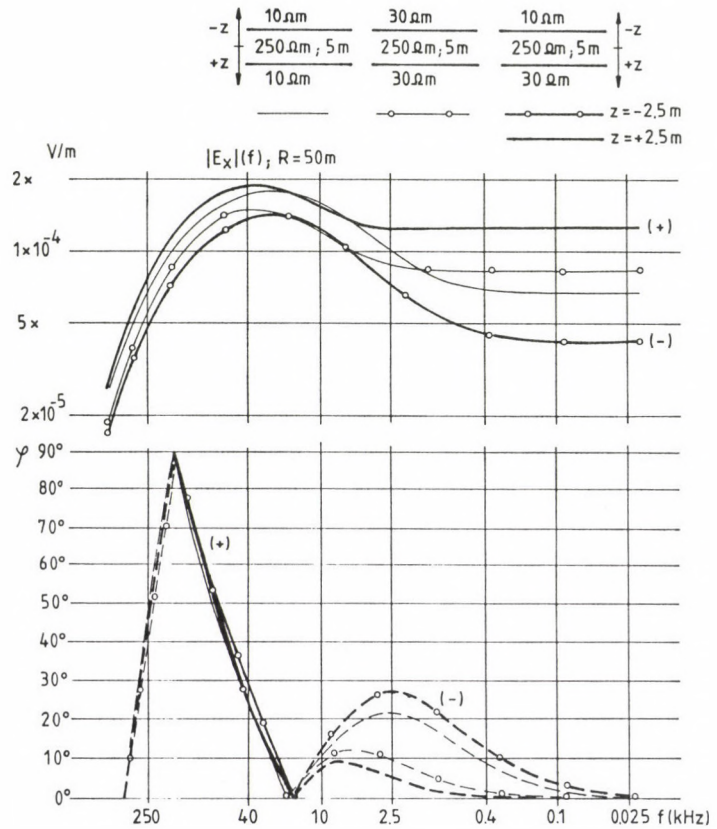


Fig. 3. Amplitude and phase frequency-sounding curves of the horizontal electric field at the seam-boundaries in symmetric and asymmetric three-layered sections. The VED is situated in the middle part of the seam at a distance of 50 m

moreover the zero-crossing points do not coincide with the equatorial plane of the seam. The maximum sites along the frequency-axis are very near to those belonging to the corresponding symmetric sections. It is evident that sounding curves belonging to the boundary of the 10 ohmm half-space must have their maximum at a lower frequency, than the curves measured at the boundary of the 30 ohmm half-space.

The phase-curve differences give information about the roof and the floor at lower frequencies: from the zero-crossing towards the DC zone. At the boundary of the more conductive

half-space the phase-values are higher and the maximum appears at a lower frequency than at the other side. The maximum site of the phase curve is very close to that of the corresponding symmetric situation.

The form of the frequency-sounding curves bears largely separate information on the roof and floor layers. However the horizontal electric field at a boundary depends on both the roof and bottom parameters. It means that any change in one of the half-spaces influences the horizontal electric field at both seam boundaries.

The regularity of this change in the quasi-stationary range - at the greatest penetration depth upwards and downwards - is as follows. In the case of a three-layered section a resistivity change in the upper or in the lower half-space influences the field intensity on the corresponding boundary. The change in the field intensity is in a close relation with the half of the ratio of the changed to the original half-space resistivity values. At the other boundary - where the resistivity contrast remained unchanged - the change of the horizontal electric component is connected with the half of the reciprocal of the same ratio. To this value (that is to the half of the ratio or of its inverse) a correction is to be added depending on the layer parameters.

This correction is due to the size difference of the rock volumes influencing the field intensity in the roof and in the floor and the zero values of E_x do not coincide any more with the equatorial plane of the seam.

In order to have some idea about the magnitude of this correction the curves in Fig. 3 (symmetric section with 10 ohmm upper- and lower half-space resistivity and an asymmetric section with 10 and 30 ohmm roof and floor resistivities) are investigated. In the case of a unit moment transmitter at 20 Hz the field intensities are as follows:

ϱ_1, ϱ_3	10, 10 ohmm	10, 30 ohmm
ϱ_2, h_2	250 ohmm, 5 m	250 ohmm, 5 m
$\|E_x(z=+2.5)\|$	$6.71\ 10^{-5}$ V/m	$4.13\ 10^{-5}$ V/m
$\|E_x(z=-2.5)\|$	$6.71\ 10^{-5}$ V/m	$1.25\ 10^{-4}$ V/m

At 20 Hz a resistivity change at the lower boundary from 10 to 30 ohmm causes a 1.86 = (30/10)/2 +0.36 time growth in $|E_x|$. At the upper, unchanged boundary $|E_x|$ decreases with a factor of 0.615 = $[(30/10)/2]^{-1}$-0.051. At the same time the ratio of E_x amplitudes belonging to the two boundaries becomes 3, which value corresponds to the floor/roof resistivity ratio.

The thicker is the seam the nearer is the ratio of the field intensities to the floor/roof resistivity ratio. The effect of the seam thickness is shown at unchanged resistivity values, at R = 50 m and f = 20 Hz by the following results:

h_2 (m)	2.5	5.0	10.0
$\|E_x(z=+2.5)\|/\|E_x(z=-2.5)\|$	3.3	3.0	3.0

3. HORIZONTAL ELECTRIC FIELD DUE TO A VED IN THE CASE OF MULTILAYERED ROOF AND FLOOR

In a three-layered section the ratio of the horizontal electric field intensities at the two seam-boundaries reflects the floor/roof resistivity ratio. A question arises, whether it is possible to conclude from the E_x frequency-sounding curves measured at the seam-boundaries to the roof- and floor layerings.

The roof and floor layering differences cause a frequency-dependent difference in the apparent resistivities at the upper

and lower seam boundaries.

Let us see at first the effect of a high-resistivity basement, close to the seam. In Figs 4 and 5 $|E_x|$ frequency sounding curves measured at the upper and lower seam boundaries are shown. For comparison the curves belonging to the three-layered section (a section without high-resistivity basement) are also presented.

At the lower boundary the $|E_x(f)|$ curves (see Fig. 4) have

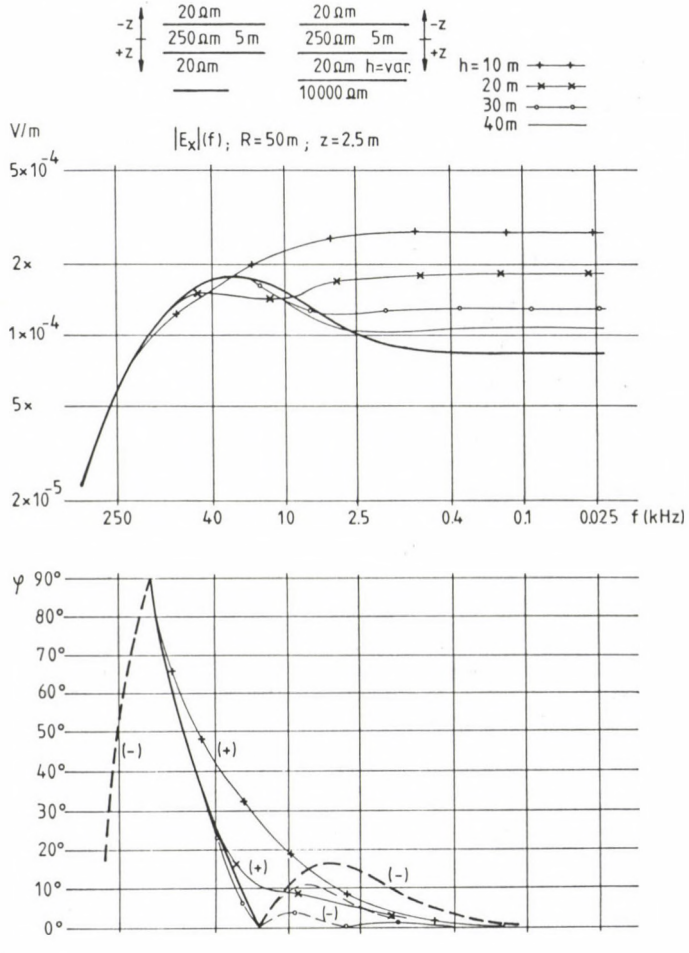


Fig. 4. Effect of the resistive basement near to the seam on the amplitude and phase frequency-sounding curves of the horizontal electric field at the lower seam boundary. The VED is situated in the middle of the seam at a distance of 50 m

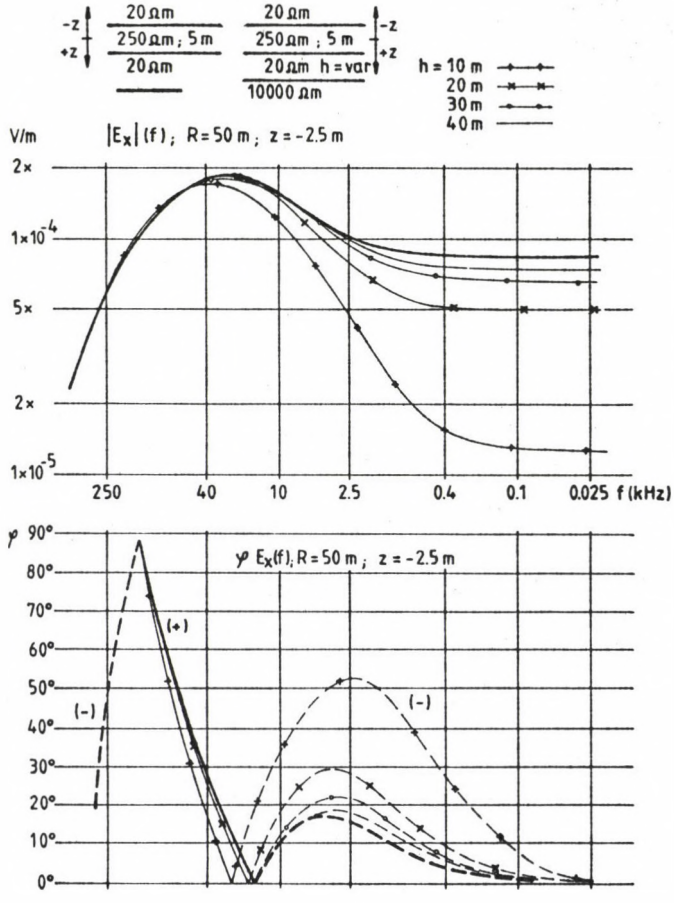


Fig. 5. Effect of the resistive basement near to the seam on the amplitude and phase frequency-sounding curves of the horizontal electric field on the upper seam boundary. The VED is situated in the middle of the seam at a distance of 50 m

a quasi-stationary asymptote. This asymptote depends on R due to the so-called geometric sounding effect. The $|E_x|$-increase is the higher, the nearer is the basement to the seam. In the intermediate frequency range a layer between the seam and the basement having 20 ohmm resistivity is clearly indicated by a relative amplitude decrease.

The basement has a characteristic indication on the phase curve, too. Beginning from h_3=30 m at low frequencies a positive undulation appears, which is more emphasized at

smaller h_3 values, resulting finally in a change of sign and in a significant change of the form of the curve. However the phase changes due to the basement remain rather small.

At the upper boundary (see Fig. 5) the basement causes a quasi-stationary decrease in $|E_x(f)|$ as it is expected from the three-layered case. The basement has an especially drastic effect at its smaller depth, resulting in a dynamically changing curve. The formation of the maximum in the $|E_x(f)|$ curves is due to the high resistivity seam. The E_x phase at the upper boundary becomes considerably higher due to the presence of the basement in the bottom. The upper half-space above the seam is homogeneous, consequently the curves there do not have any indication of the layering. The transition from the quasi-stationary part towards higher frequencies - where already the effect of the seam dominates - is regular without any local, relative extrema.

Figure 6 demonstrates that the horizontal electric component bears much more information about the roof- and floor layering than the vertical electric component. The transmitter and the receiver are placed in the case of sections 1-3 in the layer of 250 ohmm; in the case of section 4 they are placed in the layer of 20 ohmm resistivity, in each case at a distance 50 m from each other. There is a significant difference in the vertical electric field component $|E_z(f)|$ at higher frequencies only in the case of section 4. This difference is connected to changes in the resistivity and thickness of the seam. In the quasi-stationary frequency-range the difference between the $|E_z|$ curves is small in spite of considerable changes in the parameters of the sections. Consequently the vertical electric field is sensitive first of all to the inhomogeneities within the seam.

At the same time the amplitude-curves of the horizontal electric component considerably deviate at the seam-boundaries from each other especially towards the lower frequencies because of the difference in the roof and floor layerings. E.g. a 20 ohmm layer having a thickness 15-20 m is already clearly indicated at the lower seam boundary by a relative minimum.

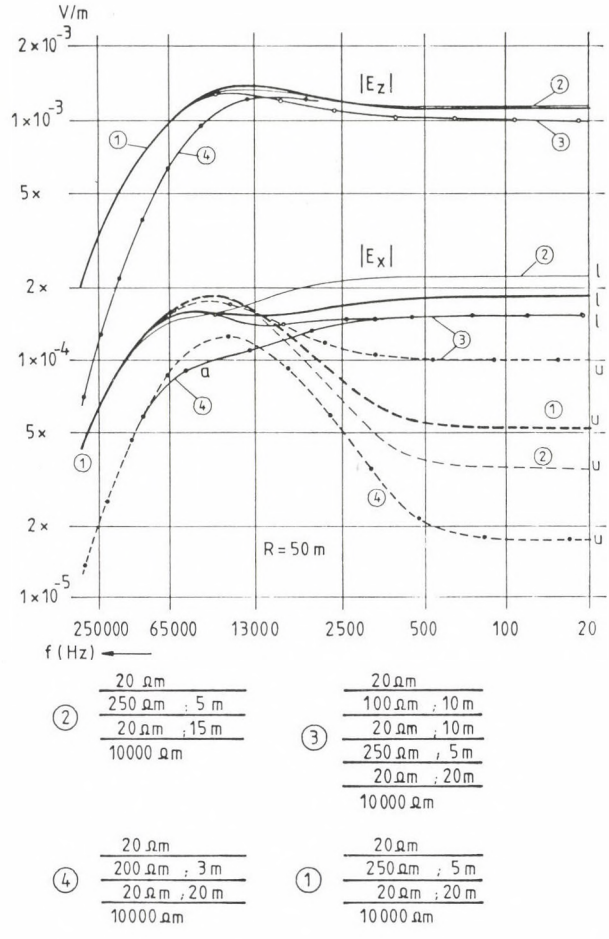


Fig. 6. Amplitude frequency-sounding curves of the horizontal (E_x) and vertical (E_z) electric components in different sections. The VED is situated in the middle part of the seam at a distance of 50 m. The $|E_x|$ curves belong to the lower (l) and upper (u) seam-boundaries while the $|E_z|$ curves refer to the middle part of the seam

4. CONCLUSIONS

The vertical electric field in the equatorial plane of a VED situated within the seam depends fundamentally on the parameters of the seam, especially at higher frequencies. The horizontal electric fields at the seam boundaries, however, are mainly influenced - especially at lower frequencies - by the

parameters of the adjacent layers.

In the case of an asymmetric section the zero-crossing of the horizontal electric component is shifted along a vertical axis from the equatorial plane towards the half-space of lower resistivity. In a three-layered section at low frequencies the ratio of the absolute values of the horizontal electric field intensities measured at the two boundaries is the same as the resistivity ratio of the corresponding adjacent layers.

The frequency sounding curves of the horizontal electric field at the upper and lower seam boundaries have the smallest difference at higher frequencies, when they are determined mainly by the coal seam and by the close vicinity of the seam. Decreasing the frequency the deviation between the two curves becomes larger according to the asymmetry in the roof and floor parameters. Layers with characteristic resistivity contrast and sufficient thickness appear with local extreme values. Therefore the frequency-sounding curves measured at the seam boundaries can be really applied for exploration of layers in the roof and in the floor separately.

REREFERENCES

Takács E 1991: Acta Geod. Geoph. Mont. Hung. (present issue)

Takács E, Nagy J, Mádai F 1986: Geophysical Transactions, 32, 43-56.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 365-383 (1991)

TWO-DIMENSIONAL MAGNETOTELLURIC INVERSION WITH MODELS OF VARIABLE GEOMETRY

T Steiner

Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H-9401 Sopron, POB 5, Hungary

[Manuscript received January 18, 1991]

The presented algorithm is a two-dimensional magnetotelluric inversion program. The crucial step in the method is the rapid calculation of modeling data with respect to the model parameters. The applied model includes resistivity as well as geometric parameters. Synthetic inversion results are given for two simple models.

Keywords: magnetotellurics; two-dimensional modelling; two-dimensional inversion

1. INTRODUCTION

The development of computational algorithms that easily extract useful information about the conductivity structure of the earth has been a long time goal of geophysics. In the case when the conductivity structure is only a function of depth a great number of methods have been proposed, implemented numerically, and applied in every-day practice. On the other hand, few algorithms have been developed that are able to use two-dimensional data and fit a two-dimensional model to them.

One remarkable approach to 2d inversion consists of a computer workstation with high resolution graphics and great computational power. An interpreter can use a trial and error method to find a subsurface geoelectrical model that fits the data and any other information at a required accuracy. A computer program solving the direct problem can be applied by the interpreter, then he can measure the "distance" between measured and computed data, and decide which model parameters he should change in order to obtain a better fit to the observed data and to any other information. This might be

labour-consuming for the interpreter. Anyway, this approach is currently feasible and useful.

There are a number of "standard methods" for 2d inversion which imply the following basic steps: (i) linearise the 2d MT direct model around an initial guess to the solution model (ii) find derivatives of observed data with respect to small changes of parameters (iii) find a better guess to the last guess (iv) repeat (i)-(iii) until a satisfactory fit to the observed data has been reached.

Standard methods have the advantage of being simple. This is a superior goal for algorithms to be used in every-day practice. The main disadvantage is their high time consumption, especially in calculating partial derivatives. This may be prohibitive in three dimensions.

An important thing that cannot be circumvented is the parameterization of the model. A common approach is to divide the earth structure into individual homogeneous blocks. Then conductivities of these blocks are only changed. A reason for this approach might be that it is quite easy to calculate partial derivatives with respect to conductivity parameters. On the other hand the geometric configuration of the model is unchanged. This restriction might be eliminated to some extent by making the blocks smaller, thus resulting in a greater number of conductivity parameters. In using this approach one must bear in mind that a fine subdivision (too many parameters) may lead to unstable interpretation (Cerv 1981). Too many parameters lead to an overinterpretation of data (e.g. revealing non-existing structures).

Presently available algorithms that allow the inclusion of changes in geometric parameters of models are based on various principles (Ku 1976, Rodi 1976, Pek 1987).

A drawback of the approach of mixing conductivity and geometric parameters is that the obtained derivatives mix two effects together: the effect of changing position of the geometric boundary of the structure and the effect of changing parameters of the mesh. A change in one of the parameters can easily lead to a change in a remote parameter (Pek 1987).

A great number of iteration methods different from the

standard ones have also been developed. The crucial step of the so-called rapid inversion schemes is to approximate the lateral gradients of the electric and magnetic fields in the model by their values from the previous iteration (Smith, Booker 1989). The EMAP (Electromagnetic Array Profiling) intends to overcome the effect of static shifts caused by near-surface conductors (Bostick 1986). One more interesting approach is the aim of generating smooth models (deGroot-Hedlin 1989), that is a model the resistivity function of which is as smooth as possible.

The algorithm to be presented here is the two-dimensional version of the 1d MT inversion program I have developed (Steiner 1989). It is a standard method that aims to determine a model with variable conductivity and geometric parameters.

2. BACKGROUND OF THE METHOD

A basic step is each inversion algorithm is the computation of the direct problem. In this case, the direct problem is the calculation of the electric and magnetic fields for a given 2d resistivity distribution below the earth surface.

Many 2d MT modeling programs have been developed over the past years. For this method to be applicable, it is not necessary to know much about the direct problem calculation. Thus, I will not mention much about the direct problem calculation method I used.

The calculation of the direct problem consists of setting up a system of linear equation $B(i) * x(i) = c(i)$ for the i-th period (with $B(i)$ as matrices, $x(i)$ as response modeling results, and $c(i)$ as the right hand side of equations). The $B(i)$-s can be set up e.g. by the discretisation of Maxwell's equations on a rectangle with the finite element (FEM) or the finite diffrence (FDM) or by any other method.

The developed iterative standard method (based on Tarantola 1987) applicable to two-dimensional magnetotelluric data consists of the following basic steps:

- The real earth structure is parametrized, thus we obtain a number of resistivity and geometric parameters - blocky model
- A starting model is guessed, a priori information about the wanted solution model is collected.

- Matrices B(i) describing the direct problem for the i-th frequency are calculated by using the finite element method (triangular elements).
- Matrices B(i) are decomposed into B(i) = L(i) * tran(L(i)) with Cholesky's method (matrices L(i) and tran (L(i)) are lower and upper triangular, respectively).
  Equations B(i) * x(i) = c(i) are solved, electric E(i) and magnetic fields H(i) are computed from x(i)-s.Thus, the initial guess responses are computed.
- The partial derivatives of the data (apparent resistivities or complex impedances) with respect to small changes in resistivity and geometric parameters are calculated. The matrix A of partial derivatives contains as many lines as many data there are (e.g. the number of measuring sites multiplied by the number of frequencies multiplied by the number of polarisation modes) and as many columns as many model parameters there are in the model.
- Resistivity and geometric perturbations are predicted that should best improve the fit to the observed data (least-squares or least-absolute-values approach). The function S to be minimized in each step is:

$$S = \Sigma \mid A * m - d_{obs} \mid^{q} / w_d + \Sigma \mid m - m_{prior} \mid^{q} / w_m$$

A : matrix of partial derivatives of the observed data with respect to the model parameters (linearised direct problem around the last approximation to the solution)

d_{obs} : observed data
m : model parameters
m_{prior} : a priori model
w_d : data weights
w_m : model parameter weights

where q = 1 for the least-absolute-values and q = 2 for the least-squares approach, respectively. The summation is done for all data and all parameters. The 1st part of the sum expresses that the modeling results (A * m - Remember that A is the locally best linear MT response function around m.) should be as close as possible to the observed data. The 2nd part,

which can be possibly omitted, prescribes that the model parameters should be close to the previous knowledge about the model. The inclusion of an a priori model can help to prevent the parameters from converging to uninteresting or unlikely ones. These two parts can be weighted by the data and parameter weights.

q = 2 leads to a system of linear equations which can be solved by any method. q = 1 leads to the standard problem of linear programming which is to be solved e.g. by the simplex method.

- The model perturbations are added to the initial guess in order to produce a better approximation to the solution model. The forward problem is solved once more again to calculate new data residuals (S). The process is repeated until a satisfactory fit to the data (and to the a priori model) has been reached.

A crucial step is the calculation of partial derivatives because this step might be the most time-consuming one. In a one-dimensional MT inversion, this calculation does not require much time since a 1d MT calculation is very rapid. However, in the 2d case, a faster method is required. This is based on the factorisation of matrices B(i) into the product of a lower and an upper triangular part. If the model parameters are only changed to a small extent, the new model response can be computed using the old and the new equations, the old solution and the old factorisation.

3. MESH GENERATION

An important part of a two-dimensional inversion algorithm is the inclusion of an automatic mesh generation routine. The utilized routine has the following features:

- The mesh sizes are the smallest at the model boundaries, then the sizes increase exponentially. This is a compromise between the limited number of nodes and the requirement of small (possibly equally spaced) meshes.
- The mesh size at boundaries is about one third of the penetration depth for the largest conductivity. The mesh is

big enough to ensure negligible influence of inhomogeneities on the response.

- At lower depths the decreasing amplitude of the excitation is taken into account. Thus there are smaller meshes at the upper part of the body than at its lower part (Figure 1).

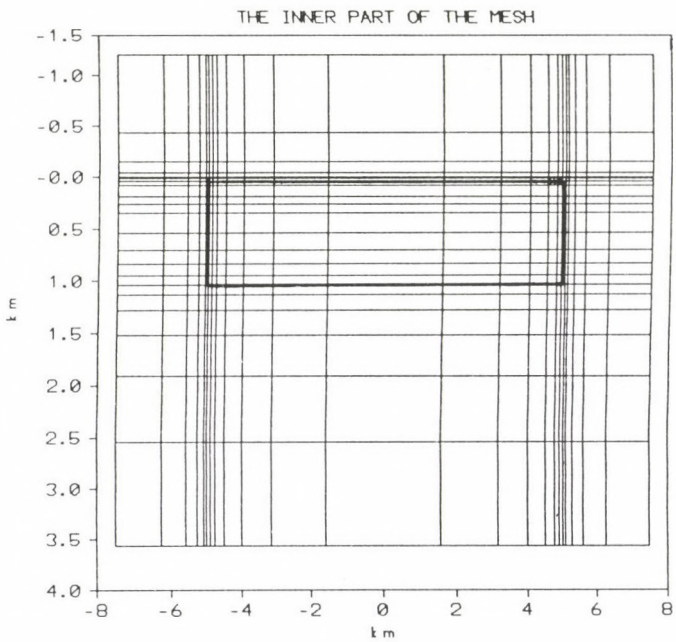


Fig. 1

Automatically created mesh
Horizontal size of body : 10 km
Depth of body : 0.04 km
Thickness of body : 1 km
Resistivity of host : 500 ohmm
Resistivity of body : 1 ohmm
Period : 0.3 sec

4. COMPUTATIONAL CONSIDERATIONS

The generally accepted method of measuring parameters and data has been used:

- log10(resistivities)
- log10(geometrical parameters)
- log10(complex impedances) or log10(apparent resistivities)

In the presented examples only log10(complex impedances) are

used. It is certain that the results would be different if the apparent resistivities were included.

The location of the model (the distance between the model centre and observation site centre) is measured either linearly or the third root of the distance is taken.

It would also be possible to use other measuring systems for data or parameters (e.g. impedance phases linearly measured, electric or magnetic field values logaritmically). The model parameters are not necessarily measured in such a simple way as given above. It might be adviseable to fix some of the model parameters or some funtion of them. E.g. one only wants to determine the ratio of some layer thicknesses instead of their actual values. This might reduce the number of parameters (the computation time needed) and enable the interpreter to include a part of his knowledge about the searched model.

The algorithm has been implemented on an IBM-AT with 640 K core memory. The run time depends on the complexity of the model and the number of periods used. A typical simple 5-6 parameter model with about 4 periods requires 2 hours.

The time consumption depends on the number of model parameters (complexity) because the number of partial derivatives increases. This is one more reason to keep the number of parameters as low as possible. The time consumption also depends on the number of frequencies (but not on the number of measuring sites) since the response of the model must be calculated at each frequency. Finally, it depends on the number of iterations. The latter is in concord with the closeness of the initial model to the solution.

This time consumption can be reduced by a network (even by a PC based one). Due to the structure of the algorithm, it is possible to assign the computation of direct problems at different periods to different PCs. A co-ordinating "main" PC can be in charge of collecting the modeling results (and the derivatives) and of calculating the next guess.

5. SYNTHETIC MODEL STUDIES

Three simple models have been used to test the algorithm.

Two of them consist of a body embedded in a host medium while the third one consists of two bodies in a host medium.

The measure used to indicate difference between the observed data and the theoretical response is a least squares measure (RMS: root mean square error):

$$RMS = \Sigma \mid r_{th} - d_{obs} \mid^2 / w_d$$

r_{th} : theoretically computed response of the last approximating model

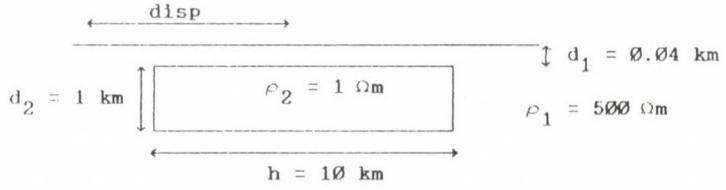
d_{obs}: log10 of observed data (impedances or apparent resistivities)

w_d : data weights (in the presented examples = 1)

Figure 2 shows test model 1 (model used to generate data), the initial guess, a single intermediate result (iteration 1) and the final guess (iteration 3) which numerically coincides with the test model. In this case, only 3 iterations are necessary to reach the solution. Nevertheless, unbiased data are used, the starting model is not very far from the solution, and the algorithm is only allowed to change four parameters: the depth and thickness of the body, the resistivity of the host and that of the body.

Figure 3 is a more complex TE-mode study because more parameters can change. It can be seen that the convergence of a single parameter (e.g. the thickness of the body) is not always straightforward. At first, this parameter increases then it begins to tend towards its true value. A similar study of the same data with some Gaussian noise added practically yields the same results. That means errors do not necessarily prevent the method from converging to the "true" solution. Note that the RMS error at the final step is quite large compared to the model in Figure 2. This is because of the Gaussian noise added to the data. On the other hand, it should be pointed out that this RMS would still be small for field data.

A TM mode calculation for the same amount of data and the same model would show (not presented here) that a much closer initial guess is required to reach the same accuracy of model parameters. Thus, it is advisable to use only TE mode data in the first few steps of the iteration and to include TM mode data



disp : distance between model centre and centre of measuring sites (fixed)

d_1 : depth of body (can be changed)

d_2 : thickness of body (can be changed)

h : width of body (fixed)

ρ_1 : resistivity of host medium (can be changed)

ρ_2 : resistivity of bodies (can be changed)

Test model used to generate data

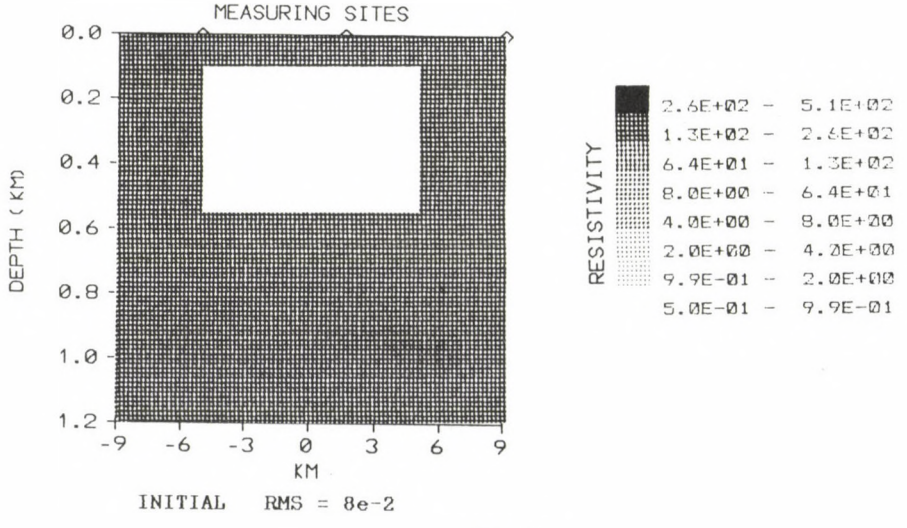


Fig 2

TE mode convergence for test model

Convergence for this simple case is rapid. Iteration 3 is practically the same as the model that was used to generate data.

Period : Ø.3 sec
Number of observation sites : 3

Part 1

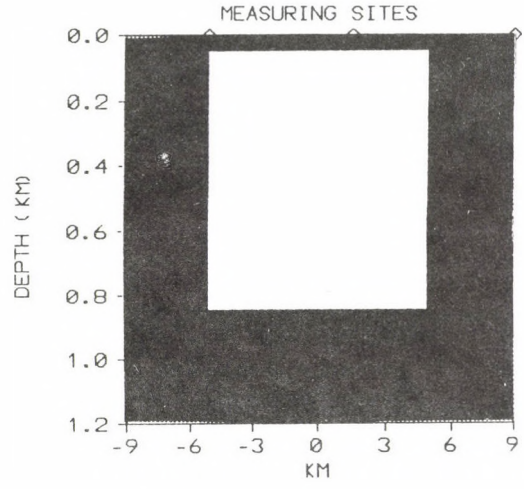
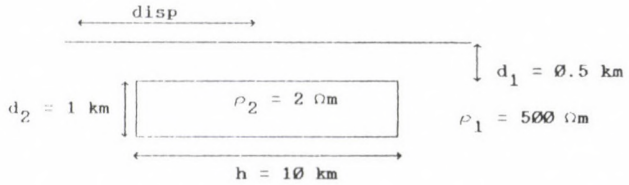


Fig. 2

TE mode convergence for test model

Part 2



disp : distance between model centre and centre of measuring sites (can be changed)

d_1 : depth of body (can be changed)

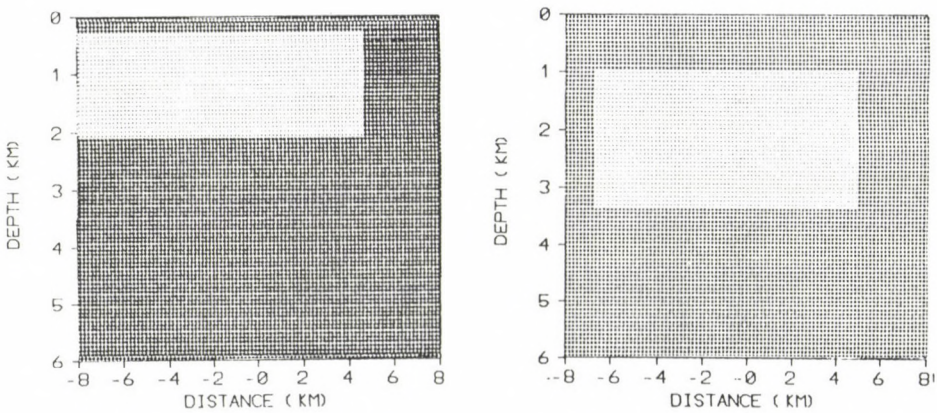
d_2 : thickness of body (can be changed)

h : width of body (can be changed)

ρ_1 : resistivity of host medium (can be changed)

ρ_2 : resistivity of body (can be changed)

Test model used to generate data



INITIAL RMS = .255908

disp = 2 km h = 13 km

d_1 = Ø.2 km d_2 = 2 km

ρ_1 = 2ØØ Ωm ρ_2 = 5 Ωm

ITER = 3 RMS = .145095

disp = Ø.855 km h = 11.29 km

d_1 = Ø.913 km d_2 = 2.489 km

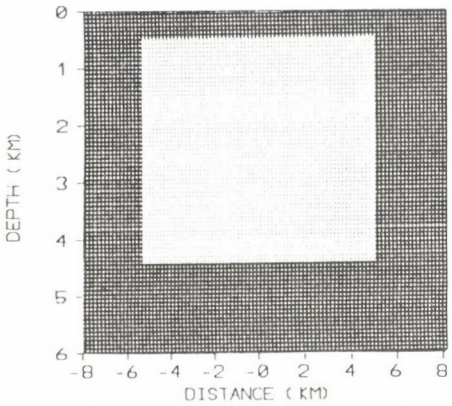
ρ_1 = 12Ø.7 Ωm ρ_2 = 4.18 Ωm

Fig 3

TE mode convergence for test model

Convergence for this case is slower than in Fig. 2.
Periods: 1Ø,1ØØ,1ØØØ,1ØØØ sec
Number of observation sites : 13
Gaussian noise added to the data

Part 1

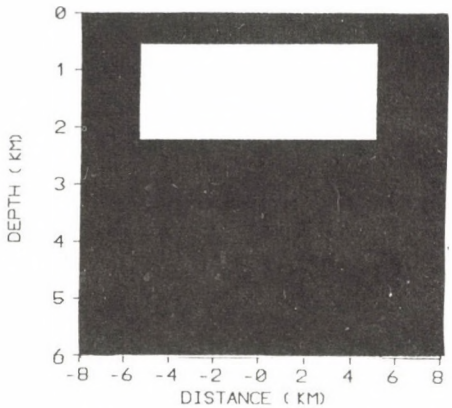


ITER = 5 RMS = .809281e-1

disp = 0.158 km h = 10.13 km

d_1 = 0.447 km d_2 = 3.973 km

ρ_1 = 314 Ωm ρ_2 = 4.57 Ωm

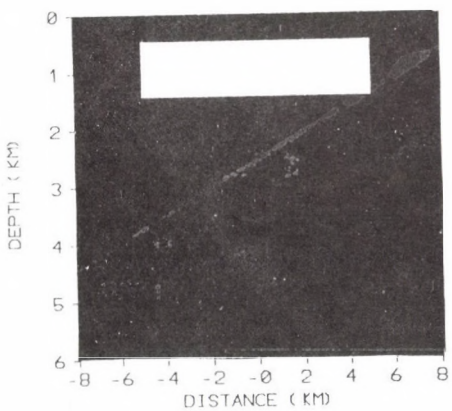


ITER = 7 RMS = .41444e-1

disp = 0.130 km h = 10.19 km

d_1 = 0.486 km d_2 = 1.712 km

ρ_1 = 383.3 Ωm ρ_2 = 3.14 Ωm

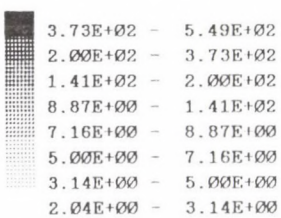


ITER = 14 RMS = .87092e-2
SOLUTION

disp = 0.040 km h = 10.02 km

d_1 = 0.504 km d_2 = 1.007 km

ρ_1 = 512.1 Ωm ρ_2 = 2.00 Ωm



RESISTIVITY CODES FOR MAPS

Fig. 3

TE mode convergence for test model

Part 2

afterwards.

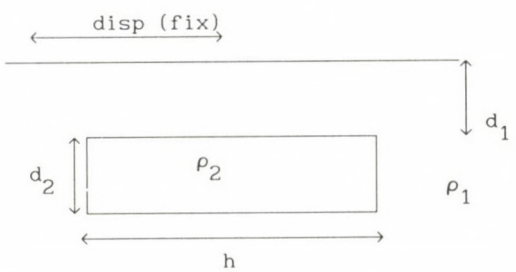
Figure 4 shows a more complicated test model, the initial guess, and some iteration steps. It is remarkable that a quite large number of iterations are required to reach the right fit to the data (without errors).

The correct parametrisation is a vital step in a 2d MT interpretation. If the data are not from the same model group as we assume results are unpredictable. E.g. if we use the data of two embedded bodies (Figure 4)

$$h_1 + h_2 + h_3 = 44 \text{ km} \qquad p_1 = 100\ \Omega m \qquad p_2 = 0.1\ \Omega m$$

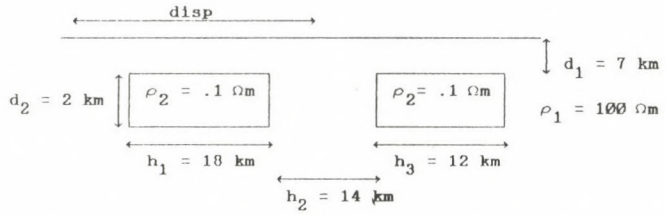
but another model for inversion (one embedded body - Figure 3), then it is not necessary to be able to determine the model parameters correctly. A study (Table I) shows that the host resistivity can be determined correctly but the thickness and the resistivity of the bodies cannot. (Remember the difference between the model used to generate data and the model used to invert them). At the same time, the algorithm converges to

Table I



	h	d_1	d_2	ρ_1	ρ_2	rms
initial guess	44	7	2	100	.1	.79e-1
iter = 1	44.8	8	10.4	95	.08	.39e-1
iter = 2	45.2	8.3	.18	95	.01	.39e-1
iter = 4	45.8	8.1	.04	94	6e-4	.35e-1

another model. The difference between the initial guess and the 1st iteration result is quite large while the differences between iter=1, iter=2 and iter=3 are small if we consider the rms error of field data. On the other hand, the difference in the models of iter=1, iter=2 and iter=3 is large. An idea to circumvent this equivalence problem is to involve an a priori



disp : distance between model centre and centre of measuring sites (fixed)

d_1 : depth of bodies (fixed)

d_2 : thickness of bodies (fixed)

h_1 : width of left body (can be changed)

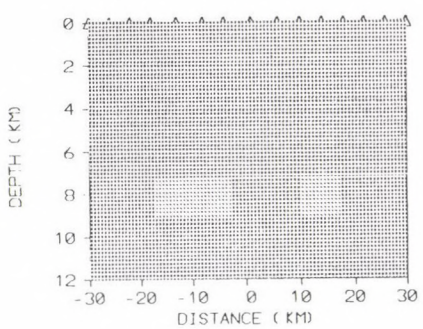
h_2 : distance of bodies (can be changed)

h_3 : width of right body (can be changed)

ρ_1 : resistivity of host medium (can be changed)

ρ_2 : resistivity of bodies (can be changed)

Test model used to generate data

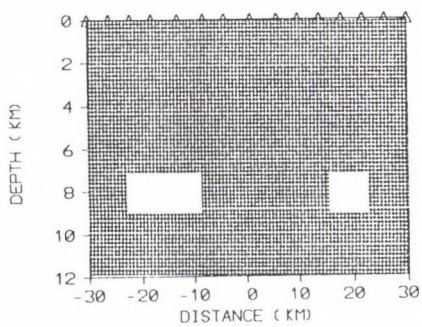


INITIAL RMS = .132425

h_1 = 15 km

h_2 = 12 km h_3 = 8 km

ρ_1 = 60 Ωm ρ_2 = .5 Ωm



ITER = 2 RMS = .523078e-1

h_1 = 14.034 km

h_2 = 24.07 km h_3 = 6.29 km

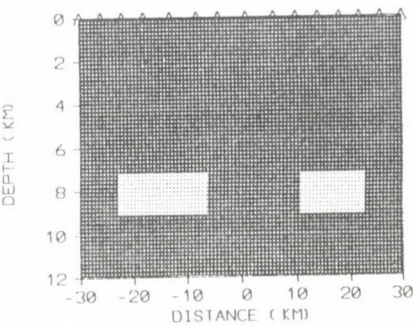
ρ_1 = 71.3 Ωm ρ_2 = 0.06 Ωm

Fig. 4

TE+TM mode convergence for test model

Note the small difference in the error between iteration 4 and iteration 6. In practice, results at iteration 4 could not be improved because of the small error.

Periods: 10, 100, 1000, 1000 sec
Number of observation sites : 15
Unbiased data are used.

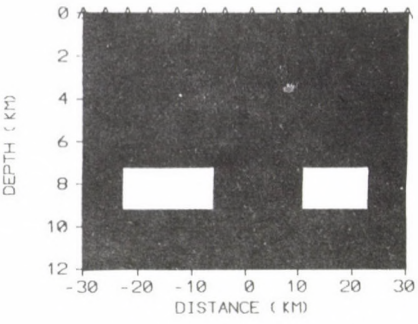


ITER = 3 RMS = .402193e-1

h_1 = 17.16 km

h_2 = 17.57 km h_3 = 11.30 km

ρ_1 = 90.3 Ωm ρ_2 = 0.10 Ωm

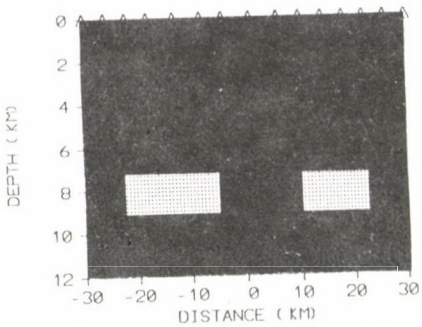


ITER = 4 RMS = .208829-1

h_1 = 17.50 km

h_2 = 15.71 km h_3 = 11.77 km

ρ_1 = 101.1 Ωm ρ_2 = 0.08 Ωm

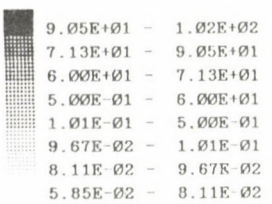


ITER = 6 RMS = .612053e-3
SOLUTION

h_1 = 17.93 km

h_2 = 14.07 km h_3 = 11.92 km

ρ_1 = 99.73 Ωm ρ_2 = .10 Ωm



RESISTIVITY CODES FOR MAPS

Fig. 4

TE+TM mode convergence for test model
Part 2

model in the inversion. This would prevent the convergence to a highly unlikely model (e.g. iter=4).

Another difficult problem may arise from the small number of nodes. It is clear that it would be useful to apply as many nodes as possible in calculating the direct problems and the derivatives. However, one is confined to the available memory. The following question arises: Is the result of the inversion influenced by the smaller than necessary number of nodes?

Table II contains some iteration steps for model of Figure 4 if the synthetic data are generated by a 18*18 mesh (18 nodes horizontally and 18 nodes vertically), and the inversion is done with the same number of nodes (the vertical parameters are fixed). In this case, the convergence is not surprising. The data in Table III are generated by a 31*26 mesh but the inversion is done with an 18*18 mesh. The latter involves that some systematic error is surely included in calculating the direct problems and the partial derivatives. It is clear that the same accuracy cannot be achieved but the main features of the model still remain.

Table II

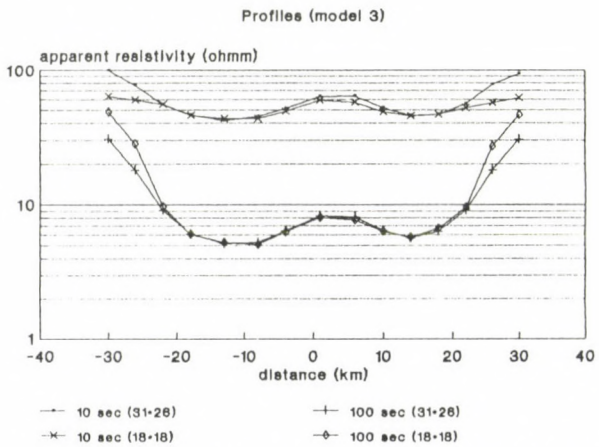
	h_1	h_2	h_3	ρ_1	ρ_2	rms
	(see model definition in Figure 4)					
test model (mesh: 18*18)	18	14	12	100	.1	
initial guess	15	12	8	60	.5	.173
iter = 1 (mesh size:18*18)	8.7	3.1	5.3	46	.05	.088
iter = 2	16.7	20.2	13.2	67	.13	.058
iter = 3	17	15.4	11.3	87	.10	.009
iter = 4	17.8	14.2	11.8	99	.10	.0014
iter = 5	18	14	12	100	.10	.0001

A higher accuracy with this limited number of nodes (18*18) cannot be achieved. In Table IV, the starting model is the same model that was used to generate the data with a 31*26 mesh. That means a better fit to the data can only be obtained with a greater number of nodes. From Table IV and Figure 5 it can be clearly seen that the algorithm tries to increase all horizontal

Table III

	h_1	h_2	h_3	ρ_1	ρ_2	rms
	(see model definition in Fig. 4)					
test model (mesh: 31*26)	18	14	12	100	.1	
initial guess	15	12	8	60	.5	.186
iter = 1 (mesh: 18*18)	8.4	3.8	5	42	.05	.118
iter = 2	14.4	25	10	55	.11	.071
iter = 3	18	17.3	11.6	76	.10	.03
iter = 4	19.3	15.1	12.9	94	.10	.0266
iter = 5	19.6	14.5	13.3	95	.10	.0264

sizes. The reason for that lies in the different mesh size used for computing the data and the inversion mesh size. The smaller than neccessary mesh size (and the systematic error in computing the direct problem - Figure 5) causes the algorithm to make the horizontal sizes larger.



2d modeling results of a model type 3 for two frequencies and two different meshes. The influence of the smaller than necessary number of nodes is larger at the higher frequency. The inversion algorithm uses a 18*18 mesh, the data are generated with a 31*26 mesh. Thus, the inversion algorithm enlarges the horizontal sizes of the model.

Fig. 5

Table IV

	h_1	h_2	h_3	ρ_1	ρ_2	rms
	(see model definition in Figure 4)					
test model (mesh: 31*26)	18	14	12	100	.1	
initial guess	18	14	12	100	.1	.0334
iter = 1 (mesh: 18*18)	19.5	14.5	13.1	95	.1	.0265
iter = 3	19.5	14.5	13.3	96	.11	.0264

6. CONCLUSIONS

We have seen a quite simple algorithm for solving 2d MT problems. The presented synthetic model results show that it is possible to use a PC for such complicated problems as is a 2d MT inversion which is generally tackled by large computers. Studies with small performance computers show interesting features of MT modeling, the resolution capacity, the required quality of initial guesses.

The main advantages of tackling the 2d MT inversion with an inversion algorithm of variable geometry and resitivities are:

- It is one of the simplest approaches. The simplicity is the more laudable because field data are usually blurred. It is good if at least the method is clear and simple.
- The possibility of the variable geometry enables to reduce the number of parameters searched thus the calculation time, too.

The drawbacks are also clear:

- The parametrization can mislead the interpreter. Overparametrisation can also introduce non-existing structures.
- The method is time-consuming, although there are approaches to keep the time consumption within acceptable limits.

The special advantages of the presented approach are:

- It is a simple PC-based method, thus generally usable.
- By using 1d and 2d methods, and gradually increasing node numbers, it is not difficult to produce results in sensible time.
- The inclusion of a priori models allows us to give a better

control over nearly equivalent models and the parameters generally.

REFERENCES

Bostick F X Jr1986: Electromagnetic array profiling. Soc.Expl. Geophys. 56'th Ann. Mtg.

Cerv V, Pek J 1981: Stud. Geophys. Geod., 15, 69-80.

deGroot C Constable S 1989: Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data. Geophysics (in press)

Ku C C 1976: Geophysics, 41, 276-286.

Pek J 1987: Physics of the Earth and Plan. Int., 45, 193-203.

Rodi W L 1976: J.R. Astron. Soc., 44, 183.

Smith J T Booker J R 1989: Rapid Inversion of Two and Three-Dimensional Data. (manuscript)

Steiner T 1989: MT inversion using least-absolute-values and least-squares. 34. International Geophysical Symp., 639-649.

Tarantola A 1987: Inverse problem theory. Elsevier

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 385-394 (1991)

APPLICATIONAL POSSIBILITY OF THE GEOSTATISTICAL AND GEOMATHEMATICAL METHODS IN THE FIELD OF ENVIRONMENT PROTECTION

J Kovács[1] and F Lesták[2]

[1]ELTE Applied Geological Department, H-1083 Budapest, Kun B. tér 2, Hungary
[2]Central Mining Development Institute, H-1525 Budapest, POB 83, Hungary

[Manuscript received January 25, 1991]

In geostatistical practice three dimensional (spatial) models are used. Time as a fourth dimension should be taken into consideration from the point of view of the environmental protection. It requires the development of the geostatistical theory and the extension of empirical investigations. The paper gives a summary of the geostatistical methods which can be applied in the environmental protection. It refers to the fields, where new theoretical and methodological developments are needed. Finally, experience of the investigations carried out so far is presented.

Keywords: environmental protection; geostatistics

We have to face more and more problems connected with the transformation of our environment. We have tried to apply a lot of methods so far to solve them. Beyond these methods there is a need for applying other ones, too. The applications of the probability theory and the mathematical statistics are favourable possibilities. One of the new models of the adaptation is the geostatistics. This discipline has become lately more and more popular among geologists and mining engineers. It should be noted, that hydraulic engineers have already used classical statistical methods for decades.

In connection with the possible application of geostatistics in the environmental protection we emphasize, that we meet here processes differing to a certain extent from those of other geostatistical applications. Namely, pollution and its impacts are processed having both time and space dimension. Research therefore should be extended.

A random variable is defined for the results of the investigations (considered as experimental results). This is

the theoretical basis of application of the methods. The variable may be equal to any value of a reasonable domain. The distribution, density function and parameters of this variable can be determined for a given time t by different estimation methods on the basis of the samples. The expected value, the standard deviation, the variation factor etc. are considered as parameters. Variability in time of the statistical indices can be described by functions referring to different times.

There are a lot of well-known distribution types in the special literature, but only a few of them are used in the problems of each special field. One of the research tasks is to determine the suitable distribution functions for the environmental protection and the environmental geology.

Not only the physical characteristics can be calculated with the help of the distribution functions, but the statistical properties, direct and indirect mechanisms of effects derived from these characteristics can be determined as well. If we know e.g. the distribution of it in some formation, the sorption properties derived from it or the statistical characteristics of the delayed migration can be also determined. The above methods can be used for identification of strata, but they are also suitable for examining spontaneous areal pollutions (caused by e.g. transportation).

The different physical and chemical parameters may have complicated connections with each other. Investigation of these connections may throw light on new laws. Measure and character of connections between two or more properties can be described by correlation analysis. In some cases, results of this analysis make it possible to reduce the number of the further laboratory tests. Connections between cause and effect can be often described by mathematical formulas in the environmental geology. Nevertheless, if the measured connected values are represented in a coordinate system, the coefficients of the corresponding function vary from sample to sample because of the restricted measurements precision and the inhomogeneity of the layer. This is also a correlate relationship. Let us consider what kind of difficulties may arise e.g. in the

determination of physical and chemical parameters of the hypergene zone, i.e. in the investigation of the interrelationship among the oxydation-reduction, hydration solution, colloidal and sorption processes. Anyway, they are the determinant factors of the element distribution. Obviously multivariate correlation analysis should be applied in the investigation of the above parameters. Method of principal components and factor analysis should be used after each other in order to determine the number of typical characteristics. Correlation analysis is applied in geology, too. For example experts of the Planning Office of 'Dorog' Collieres appplied it to describe the connection between the water level of the Danube and the cavern water flow.

Application of cross correlation plays an important role in the anomalies of water economics and the environmental problems. There are a lot of examples in the hydrological statistics - in connection with surface cases and underground water - which prove this. Cross correlation can be applied in the analysis of infiltration and water level change, lowering of cavern water level and remote effect, or in the investigation of plants influencing the water economics.

For example any kind of interference with the natural relations of rivers induce processes which cause changes in stream, sludging velocity and the decomposition mechanism in the water, therefore these processes influence the self-purification processes. It can be considered as another example, that terrain subsidences as a surface impact of the hydrocarbon mining may induce a raise of the underground water level and they may modify the fall of rivers as well. These cases can be considered as environmental damages, where correlation can be applied. We have already mentioned the role of time dimension in the pollution of the environment. If data series are analysed as a function of time, three components can be generally separated: the trend, a periodic component and a stochastic one as a random effect. The first component describes the long-term trend. The second one is the result of regularly repeating processes, where the length of period can

be determined. The random effect is derived from the process itself and the technical errors of sampling. A sufficient number of samples is the necessary condition of a reasonable calculation of the trend concerning a given time, whilst a sufficient number of samples is needed during a longer period for calculation of the long-term trend. Explanation of the trend may be either a natural or an artificial effect acting in a direction. Similar factors explain the periodicity, but their directions are cyclically changing in space and time. These changes are often in connection with meteorological phenomena. For example the daily and seasonal changes of the underground water can be connected with the changes of decomposition processes of pollution, as the presence of water is a determinant factor of the biological activity. Movement of water (waterflow) plays an important role in the spread of pollution, too. However the examination of the artifical effects has a greater importance in the environmental protection (periodically repeating pollution of agricultural activities of industrial and public institutions in towns).

For example phosphate minerals may appear in the soil as an effect of the regularly used artificial fertilizers. The process of their emergence can be examined with the help of time series. The 'memory' of each environmental element should also be investigated, because it plays an important role in the self-purification processes. In environmental protection there is a great possibility for observing and gathering data, if telesensing methods are applied. We recall the serial and cosmic photographs used frequently in detection of pollutants of the atmosphere and the oceans. Use of photographs made in appropriate time and wavelength intervals can be even extended. If the appropriate measures are carried out with the help of the photographs, time series of changes with respect to directions, areas and referring to the intensity can be determined.

A sufficient number of samples should be available for examining an area.

Archival data can be used, but they must not be accepted without any criticism. Namely, the unreliable data may

considerably bias the estimated parameter values, therefore one has to act with understanding when using them. The archival data have been collected for decades by different methods in different laboratories. The processing of this whole data set may be a source of error. It has to be emphasized that samples have to be representative for the data processing. It means that samples should reflect the assembly, and they have to be independent as well.

The above facts are illustrated by the following examples. Coastal filtering antennal wells have been constructed for the water supply of Budapest. However, at some intake places quantities of iron and manganese exceed the fixed standard values. As they are present in the water under reductive circumstances, a system with in situ high actual oxidation-reduction potential has to be artifically created for their precipitation. A way of it can be the so called aeration. Aering the experiment the process of aeration-production were repeated several times. The chemical data series were examined for every period and in an aggregated way as well. Fig. 1 presents the empirical and theoretical semivariogram of the manganese created on the base of the aggregated data series. One can see, that the influence time is cca. 6.5 days. At the given level of technology there is an important economic impact of this result from the research and production: it is sufficient to take samples only by 11-12 days.

During the experiment samples were examined from the point of view of the chlorine, too. It has to be emphasized, that the experiment has no influence on the quantity of the chlorine. Variability of chlorine content can be optimally presented by a spline interpolation (Fig. 2). The polynomial and the harmonic trends have a smoothing effect on the data series (Fig. 3). As the influence time is 10.47 days (Fig. 4), the last samples from this interval of the time series were used for the prediction. Prediction can be based either on polynomial (Fig. 5) or harmonic trend (Fig. 6). Sometimes their combination can be also used. The range of prediction error is illustrated by using a polynomial trend for the prediction (Fig. 7). It shows,

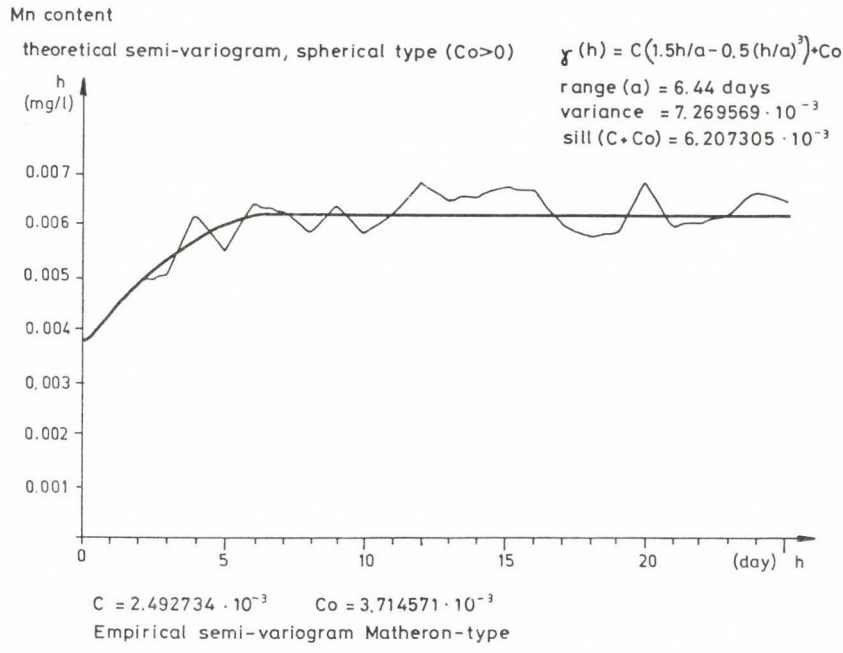


Fig. 1. Ráckeve, first well series, variogram analysis of the manganese data

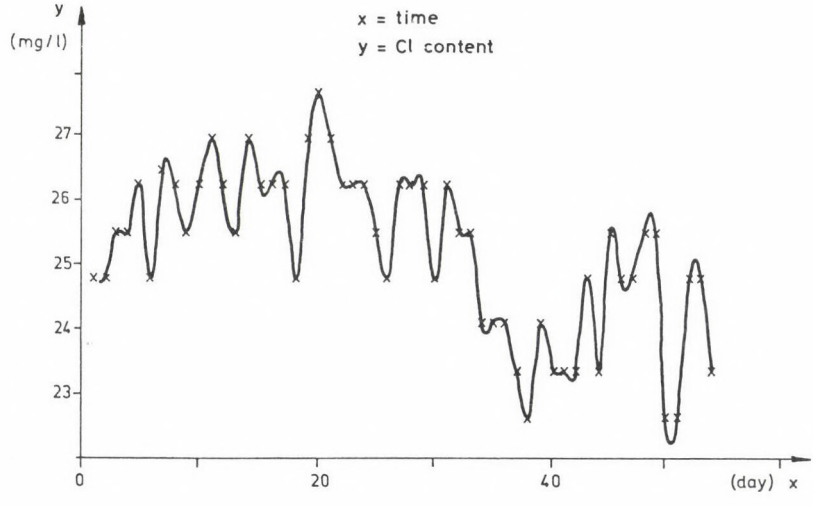


Fig. 2. Ráckeve, first well series, fourth period spline interpolation of the chlorine data

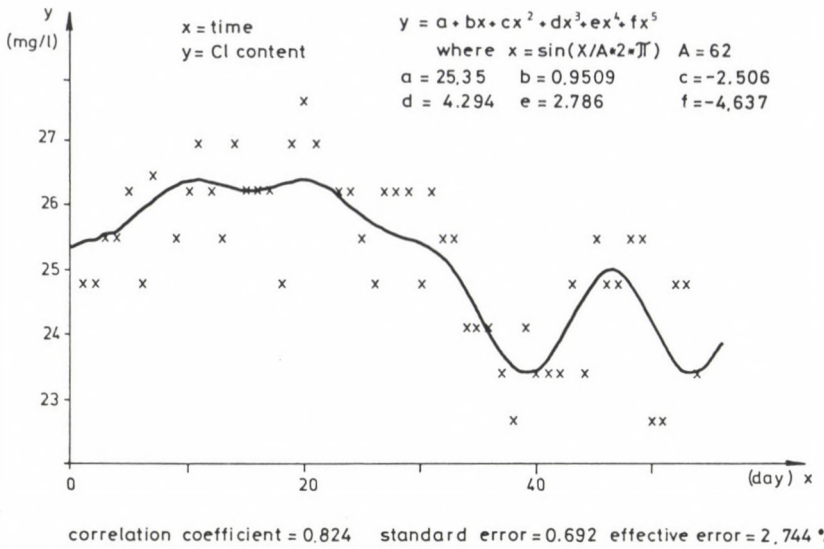


Fig. 3. Ráckeve, first well series, fourth period regression analysis of the chlorine data

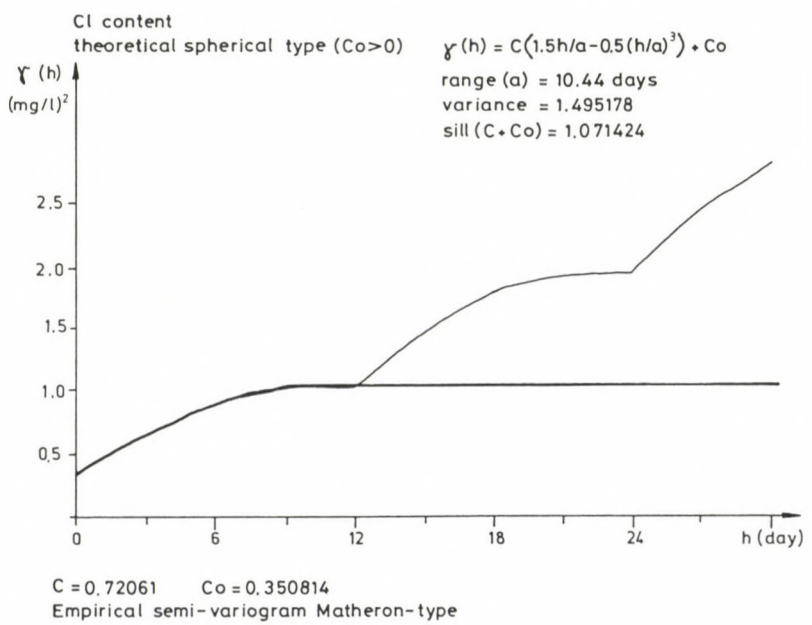


Fig. 4. Ráckeve, first well series, fourth period variogram analysis of the chlorine data

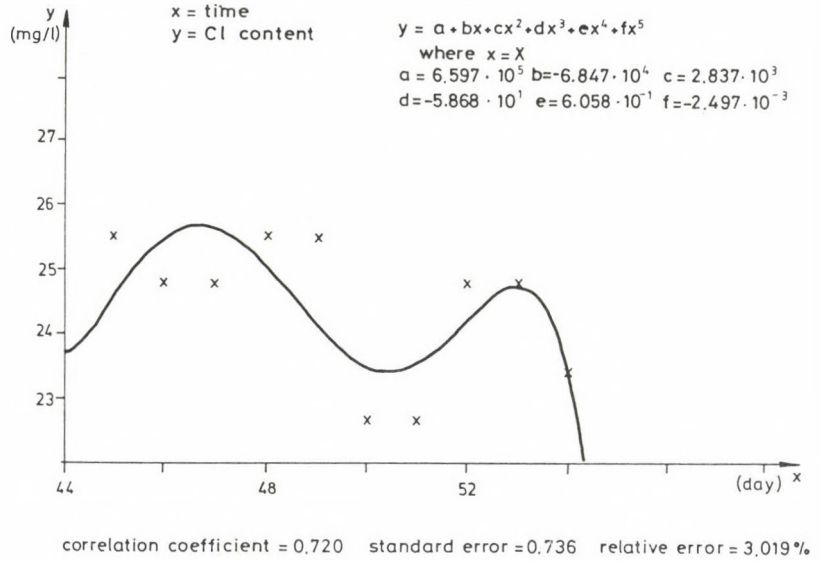


Fig. 5. Ráckeve, first well series, fourth period. Forecast by polynomial trend function of the chlorine data

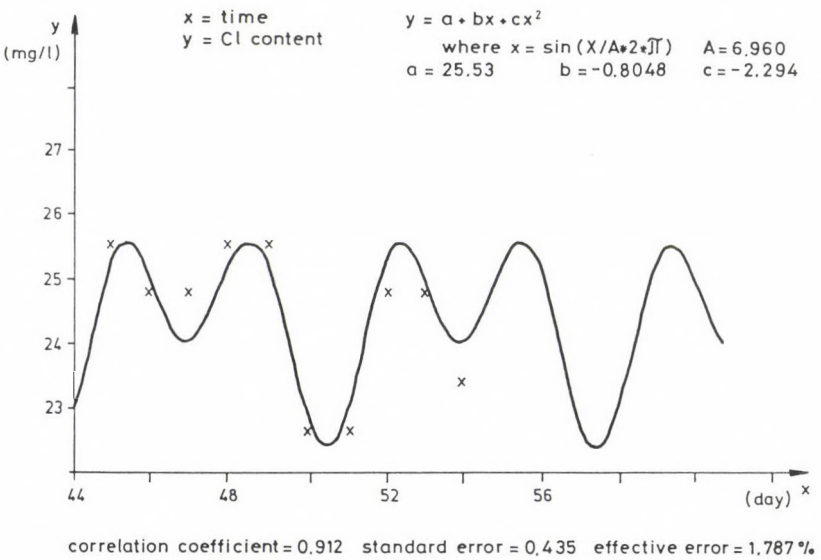


Fig. 6. Ráckeve, first well series, fourth period. Forecast by harmonic of the chlorine data

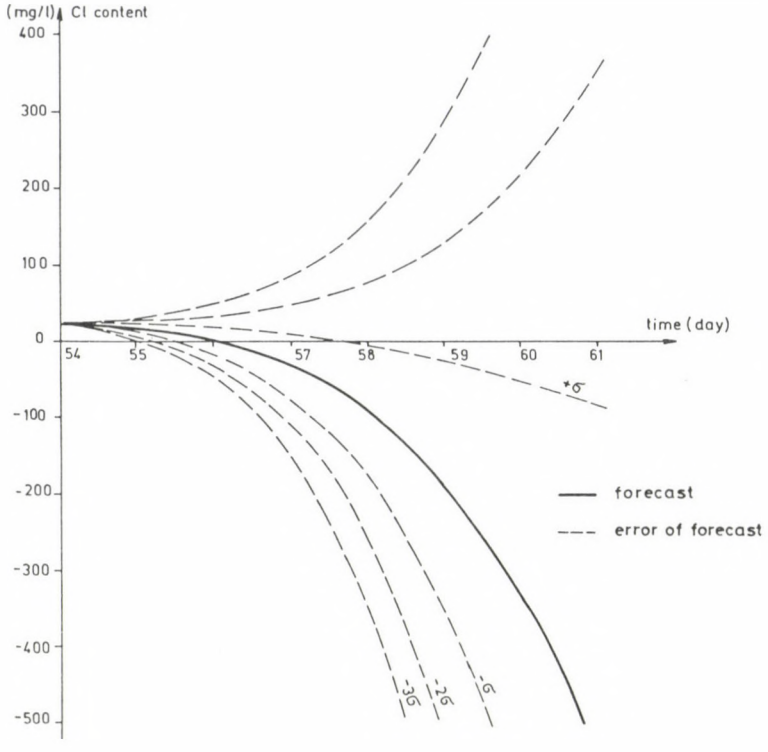


Fig. 7. Forecast and error of prediction in case of polynomial trend function

that if there is a trend in the prediction, the realistic prediction time may considerably shortened.

We hope, that we succeeded in highlighting the great possibilities of the mathematical and geostatistical applications in the environmental protection problems. At the same time we are aware of the diffulties, because we have not a data set yet satisfying the necessary conditions. Therefore the future examinations should be planned in such a way, so that they make possible the geostatistical approach of the above problems. Theory should be also developed for the application of geostatistics in the environment protection.

REFERENCES

Agterberg F P 1974: Geomathematics. Elsevier Scientific Publishing Company, Amsterdam, London, New York

Box G E P, Jenkins G M 1970: Time Series Analysis Forecasting and Control. Holden-Day, San Francisco, Cambridge, London, Amsterdam

David M 1977: Geostatistical Ore Reserve Estimation. Elsevier Scientific Publishing Company, Amsterdam, Oxford, New York

Guarascio M et al. (eds) 1976: Advanced Geostatistics in the Mining Industry. D. Reidel Publishing Company, Dordsecht-Holland/Boston-USA

Journel A G, Huijbregts Ch J 1978: Mining Geostatistics. Academic Press, London, New York, San Francisco

Marsal D 1987: Statistics for Geoscientists. Pergamon Press, Oxford, New York, Beijing, Frankfurt, Sao Paulo, Sydney, Tokyo, Toronto

Matheron G 1965: Les Variables Regionaliées et leur Estimation. Masson at Cie. Editeurs, Paris

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 395-405 (1991)

RESISTIVITY DATA INTERPRETATION IN THE VICINITY OF VERTICAL DISCONTINUITIES USING THE MATRIX METHOD

M Israil and Sri Niwas

Department of Geophysics, University of Kurukshetra, Kurukshetra-132 119, India

[Manuscript received February 28, 1991]

A matrix method is presented for the computation of apparent resistivity in the vicinity of vertical discontinuities. The method is also used to estimate the kernel function from the apparent resistivity data. A simple procedure is given for the estimation of parameters of the model from the computed kernel values. A few examples of interpretation using theoretical data sets are included which demonstrate the utility and limitations of the method.

Keywords: apparent resistivity; digital modeling; resistivity sounding; vertical discontinuity

INTRODUCTION

Logn (1954) used the spherical harmonic approach to deduce the potential as well as the apparent resistivity functions due to a point current electrode placed in the vicinity of vertical discontinuities. He presented a few types of curves using a rigorous computational procedure. Lee and Green (1973) developed a simple method for the direct interpretation of electrical sounding carried out either normal or parallel to the vertical discontinuities. For numerical computation of an infinite integral, Lee and Green (1973) used the Meinardus (1970) procedure. A critical analysis of the results by Lee and Green (1973) shows that the numerical values oscillate what brings subjectivity in the graphical method and undesired error in the automatic method of interpretation. A digital linear filter method for resistivity interpretation was first suggested by Kunetz (1966) and executed by Ghosh (1970) for horizontally stratified earth. Sri Niwas (1975) used the linear

filter method for vertical discontinuity problems. Patella (1980) has discussed a procedure similar to that of Lee and Green (1973) for graphical interpretation of the vertical fault problem.

Recently Sri Niwas and Israil (1986, 1987a, 1987b) developed a matrix method for the geoelectrical computation for horizontally stratified layered earth. The approach is extended here for the computation of apparent resistivities and kernel values for the vertical discontinuities. A few numerical results are presented as an example.

MATHEMATICAL FORMULATION

The expression for the potential as obtained by Logn (1954) due to a point current source of strength I placed at the origin in the vicinity of vertical dike or a fault is given by (Fig. 1)

$$V(r,x) = \frac{I\,\rho_1}{2\pi} \int_0^{\infty} \left[e^{-\lambda x} + K(\lambda)e^{\lambda x}\right] J_0(\lambda r)\,d\lambda \qquad (1)$$

$$-\infty < x < a \ .$$

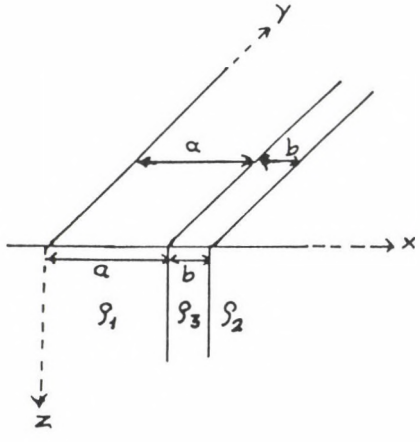


Fig. 1. Geometry of the vertical discontinuity problem

where r is the radial distance from the source on yz plane and $K(\lambda)$ is the kernel function given by $\varrho_{21}\, e^{-2\lambda a}$ and

$$\frac{\left[\varrho_{31}\, e^{-2\lambda a} + \varrho_{23}\, e^{-2(a+b)\lambda}\right]}{\left[1 - k e^{-2\lambda b}\right]}$$

for vertical fault and dike respectively, a is the distance from the first discontinuity, ϱ_1, ϱ_3 and ϱ_2 are the resistivities of the successive media, $\varrho_{mn} = (\varrho_m - \varrho_n)/(\varrho_m + \varrho_n)$, and b is the width of the dike. Here the X-axis and Z-axis are along the horizontal and vertical direction and the Y-axis is parallel to the strike of the vertical discontinuities.

The kernel can be written using an exponential approximation upto p term as (Sri Niwas and Israil 1986)

$$K(\lambda) = \sum_{i=1}^{p} f_i\, e^{\xi i \lambda} \quad . \tag{2}$$

Substituting Eq. (2) and by using the Lipschitz integral

$$\int_0^\infty e^{-\xi i \lambda}\, J_0(\lambda r) d\lambda = \frac{1}{\sqrt{(\xi^2 + r^2)}}$$

Eq. (1) can be written as

$$V(r,x) = \frac{I\, \varrho_1}{2\pi} \left[\frac{1}{(x^2+r^2)} + \sum_{i=1}^{p} f_i \frac{1}{(\xi_i - x)^2 + r^2} \right] \quad . \tag{3}$$

For sounding carried out parallel to the fault or dike (x=z=0) we get

$$V(Y) = \frac{I\, \varrho_1}{2\pi} \left[\frac{1}{Y} + \sum_{i=1}^{p} f_i \frac{1}{(\xi_i^2 + Y^2)} \right] . \tag{4}$$

This equation can be used to write the apparent resistivity expression for any desired electrode configuration. As an example the Schlumberger apparent resistivity expression can be written using the following equation

$$\varrho_{as} = \frac{2\pi}{I} Y^2 \left(- \frac{\partial V}{\partial Y} \right)$$

as

$$\varrho_{as} = \varrho_1 \left[1 + \sum_{i=1}^{p} f_i \left(\frac{Y}{(\xi_i^2 + Y^2)} \right)^3 \right] . \qquad (5)$$

Following the procedure outlined by Sri Niwas and Israil (1986) the matrix equation to compute the kernel function from measured data can be written as

$$K = E(GG^t)^{-1} G^t R . \qquad (6)$$

Where the elements of the matrix E are $e^{-\xi_i \lambda_j}$ and those of G are $\frac{Y_j^3}{(\xi_i^2 + Y_j^2)^{3/2}}$; i = 1, 2, .., p; j = 1, 2, .., q. The column vector R is assembled from the tabular values of the left hand side of Eq. (5) at q electrode separations y_d. Once the kernel function is obtained it can be subsequently analysed in terms of various parameters of the fault or dike.

DERIVATION OF THE PARAMETERS

For large values of λ, the expression of the kernel function for a vertical dike can be expanded as a power series in the exponential function $e^{-2\lambda b}$ as

$$K(\lambda) = \left[\varrho_{31} e^{-2\lambda a} + \varrho_{23} e^{-2\lambda(a+b)} \right] (1 - k e^{-2\lambda b})^{-1} \qquad (7)$$

the above equation can be written as,

$$K(\lambda) = \left[\varrho_{31} + \varrho_{23} e^{-2\lambda b} \right] e^{-2\lambda a} \sum_{n=0}^{\infty} k^n e^{-2\lambda b n} \qquad (8)$$

for a large value of λ , Eq. (8) may be approximated for a vertical dike as,

$$K(\lambda) \simeq \varrho_{31} e^{-2\lambda a} . \quad (9)$$

By substituting $\varrho_{23} = 0$, $b = 0$ and $\varrho_{31} = \varrho_{21}$ in Eq. (7), the kernel function for the vertical fault can be written as,

$$K(\lambda) = \varrho_{21} e^{-2\lambda a} . \quad (10)$$

If $K(\lambda)$ is plotted on a log scale against λ on a linear scale, a straight line is obtained in the case of a vertical fault. The intercept of the straight line on the ordinate gives ϱ_{21} and its slope the value of a. For the case of vertical dike, values of $K(\lambda)$ for large values of λ fall on a straight line whose intercept on the ordinate gives ϱ_{31} and whose slope gives the value of a. To determine the value of b and ϱ_{23} we have to calculate a function $F(\lambda)$ derived from $K(\lambda)$ given by

$$F(\lambda) = \frac{K(\lambda) - \varrho'_{31} e^{-2\lambda a'}}{e^{-2\lambda a'} + \varrho'_{31} K(\lambda)} \quad (11)$$

where a' and ϱ'_{31} are the interpreted values.

The above expression reduces to the following equation

$$F(\lambda) = \varrho_{23} e^{-2\lambda b} . \quad (12)$$

If we plot $F(\lambda)$ on a log scale against λ on a linear scale, a straight line is obtained whose intercept on the ordinate gives ϱ_{23} and the slope b.

ACCURACY OF NUMERICAL RESULTS

The computational procedure has been discussed in earlier works (Sri Niwas and Israil 1986, Israil 1988). Matrix G is constituted using the appropriate value of electrode separation and a set of ξ_i values increase in geometrical progression

with a leading term 0.2 and a common ratio 2. Schlumberger sounding carried out parallel to a vertical fault with parameters: a = 1.0 and ϱ_{21} = 0.435 is taken initially. The kernel values are evaluated using Eq. (6) and plotted in Fig. 2 with the values obtained by Lee and Green (1973) for the same model. The values obtained by Eq. (6) are consistently lying on a straight line. Interpreted values are a = 1.0 and ϱ_{21} = 0.434 (0.4), values in bracket are values interpeted by Lee and Green (1973). Another example of Schlumberger sounding carried out parallel to the vertical dike with parameters: a = 1.0, b = 0.5, ϱ_{31} = 0.435 and ϱ_{23} = -0.635 is also

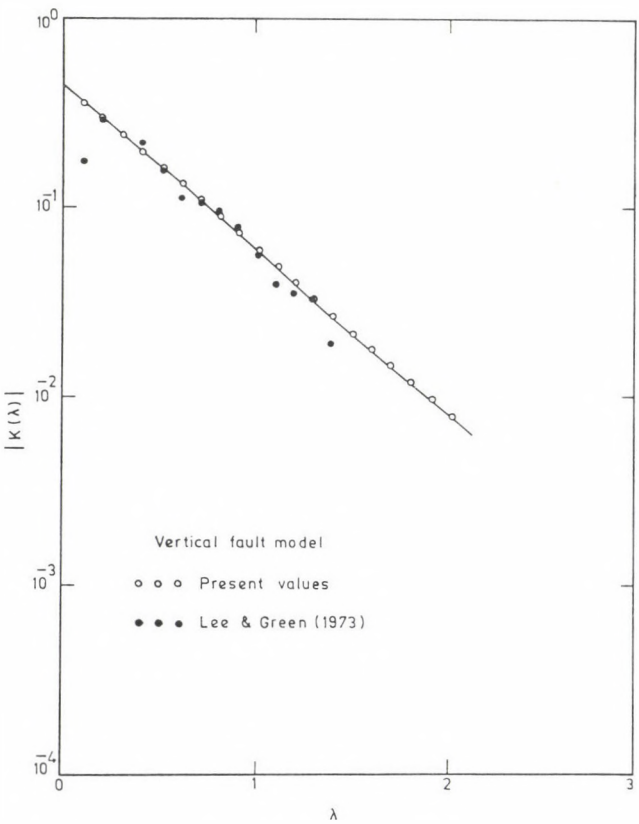


Fig. 2. Interpretation of vertical fault model with parameters: a = 1.0, ϱ_{21} = 0.435

interpreted and results are presented in Fig. 3. The first stage of interpretation is again compared with the values obtained by Lee and Green (1973). It may be mentioned here that a different expression has been used by Lee and Green for the second stage of interpretation. The first stage of interpretation yields a = 0.985 and ϱ_{31} = 0.44, these values have been used to evaluate F(λ) for the second stage of interpretation which gives b = 0.45 (0.6) and ϱ_{23} = -0.62 (-0.9). From Figs 2 and 3 it can be seen that oscillation has been reduced drastically in the present computations which leads to a considerable improvement in the interpreted results.

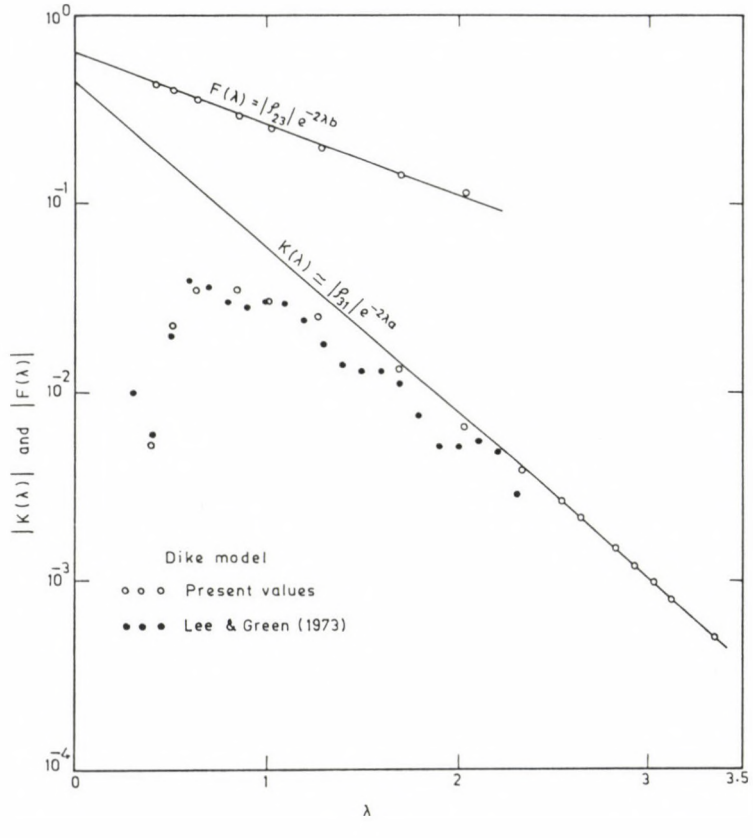


Fig. 3. Interpretation of dike model with parameters: a = 1.0, b = 0.5, ϱ_{31} = 0.435, ϱ_{23} = -0.635

To further test the limitations of the present method following models with variable dike thicknesses have been interpreted.

i) a = 1.0, b = 0.75; ϱ_{31} = 0.435, ϱ_{23} = -0.435
ii) a = 1.0, b = 0.5 ; ϱ_{31} = 0.435, ϱ_{23} = -0.435
iii) a = 1.0, b = 0.1 ; ϱ_{31} = 0.435, ϱ_{23} = -0.435 .

Interpretations of above models are shown in Fig. 4. The obtained values of the parameters are:

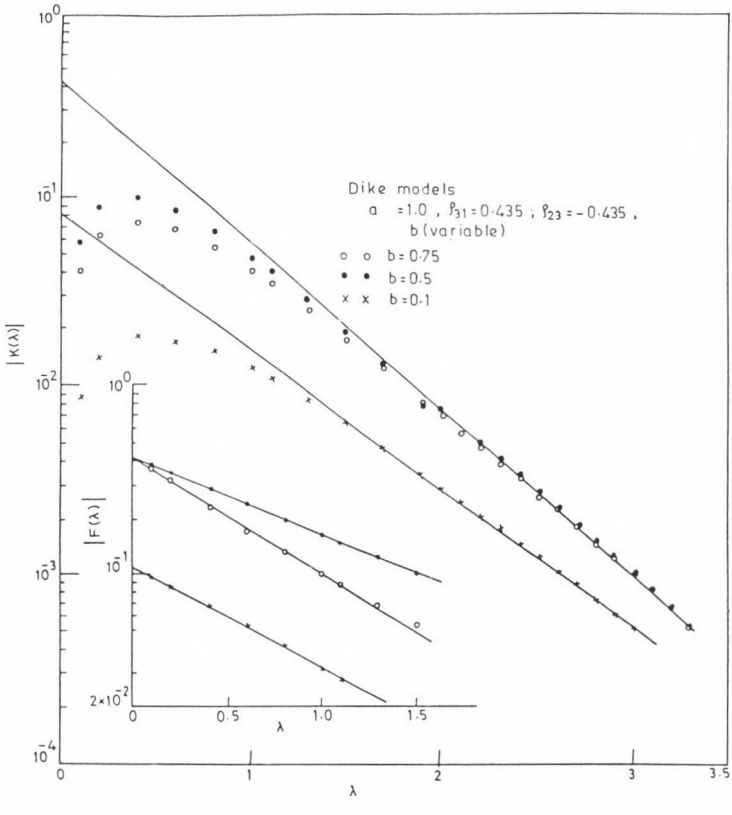


Fig. 4. Interpretation of three dike models with varying dike thickness b, parameters are: a = 1.0, ϱ_{31} = 0.435, ϱ_{23} = -0.435 and b(varying) = 0.75, 0.5, 0.1

i) a = 0.99, b = 0.72; ϱ_{31} = 0.43, ϱ_{23} = -0.43
ii) a = 0.99, b = 0.48; ϱ_{31} = 0.43, ϱ_{23} = -0.43
iii) a = 0.88, b = 0.62; ϱ_{31} = 0.105, ϱ_{23} = -0.11 .

For model iii), where the thickness of the dike is very small (0.1), a large error has been obtained in the interpretation. Thus it is not possible to determine a very small value of b ($\leqslant$ 0.1) accurately, using the graphical method. In such cases a direct inversion method can be applied. We have carried out an iterative least square inversion (Lines and Treitel 1984) for five dike models with various dike thicknesses. In this scheme kernel values are obtained by the method discussed in the present paper. Initial model parameters are obtained using the graphical method. Figure 5 shows the convergence for these models. In all cases good results are obtained with small error (less than 0.1 percent) in each parameter.

CONCLUSION

The utility of matrix method has been investigated for the problems of vertical discontinuities. The main contribution lies in simplifying the computation of apparent resistivity and kernel values from the known apparent resistivities. The procedure is analogous to the horizontally stratified earth. Once the kernel values are obtained, derivation of model parameters becomes easier. The oscillation in the computed kernel values have been reduced what leads to an improvement of the results. The models discussed above also show the limitations of the method for a very thin dike.

ACKNOWLEDGEMENT

The authors are indebted to Prof. V K Gaur and Prof K N Khattri who have been source of inspiration throughout the work. We are also thankful to Dr. P K Gupta for fruitful discussions. Financial support provided by CSIR is thankfully acknowledged.

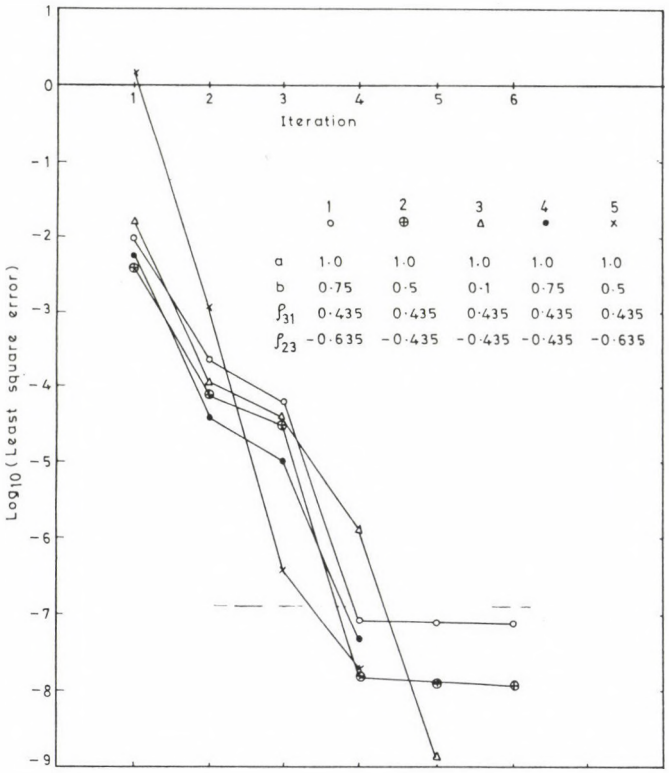


Fig. 5. Convergence of the least squares error for the various dike models

REFERENCES

Ghosh D P 1970: The application of linear filter theory to the direct interpretation of geoelectrical resistivity measurements. Ph.D. Thesis, Delft

Israil M 1988: Matrix method for the interpretation of direct current resistivity data. Ph.D.Thesis, Roorkee University

Kunetz G 1966: Principles of direct current resistivity prospecting. Borntraeger, Berlin

Lee T, Green R 1973: Geophysics, 38, 762-777.

Lines L R, Treitel S 1984: Geophys. Prospecting, 32, 159-186.

Logn O 1954: Geophysics, 19, 739-760.

Meinardus H A 1970: Geophysics, 18, 415-433.

Patella D 1980: Geophys. Prospecting, 28, 961-966.

Sri Niwas 1975: Geophysics, 40, 121-122.

Sri Niwas, Israil M 1986: Geophysics, 51, 1594-1602.
Sri Niwas, Israil M 1987a: Geophys. Prospecting, 35, 548-567.
Sri Niwas, Israil M 1987b: Geophysics, 52, 1412-1417.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 407-422 (1991)

ELECTROMAGNETIC DISTORTIONS IN TERMS OF POTENTIALS IN TWO-DIMENSIONAL MAGNETOTELLURICS

L Szarka

Geodetic and Geophysical Institute of the Hungarian Academy of Sciences,
H-9401 Sopron, POB 5, Hungary

[Manuscript received March 4, 1991]

Electric and magnetic field components inside a conducting lower half space exhibiting arbitrary resistivity inhomogeneities have been directly expressed by Szarka and Fischer (1990) in terms of subsurface current and charge distributions. Comparing these equations with the definitions of the potentials, significant new integral formulas can be found in the basic magnetotelluric situations for the components of the vector potential and the scalar potential.

A clear separation of the relations into normal and anomalous parts would perhaps provide a reliable basis for an exact elimination of the magnetotelluric field distortions.

The relation between different vector potential components and the scalar potential and its effect on the electromagnetic field components and their combinations are also analysed.

Keywords: current field; electromagnetic potentials; free charges; magnetotellurics; two-dimensional problems

INTRODUCTION

The classification of distorted magnetotelluric sounding curves given by Berdichevsky and Dmitriev (1976) has been the starting point of several recent investigations, however, the terminology used in Western countries is different from that used by Soviet authors (e.g. Jones 1983).

The usual method of determining magnetotelluric or other electromagnetic field distortions is to calculate apparent resistivity curves from the geoelectric parameters (usually the conductivities and the geometrical structure) and then to study how these curves depend on particular subsurface model parameters. In the present paper the investigation will be conducted starting from the real, effective source of

electromagnetic anomalies, i.e., from the subsurface current and charge distributions.

A need for such an approach was clearly expressed at the Sochi IAGA electromagnetic induction workshop (e.g. Jiracek 1988, Oldenburg 1988 and Spichak 1988). However, instead of limiting the considerations to apparent resistivity and phase, along the lines by, e.g. Menvielle (1988), we shall derive the electromagnetic potentials for the two basic two-dimensional magnetotelluric situations.

The aim of this paper is just to present close physical connections between the potentials and subsurface current- and charge distributions in order to demonstrate that in each particular situation which potential component is responsible for the observed electromagnetic field distortion.

The relatively simple derivation along the lines of Szarka and Fischer (1989, 1990) given here may perhaps contribute to establish a clear potential-oriented analysis of the electromagnetic methods.

PROBLEMS IN DETERMINATION OF ELECTROMAGNETIC POTENTIALS IN 2-D MAGNETOTELLURICS

The general solution for the vector potential $\underset{\sim}{A}$, where $\underset{\sim}{j}$ means the current density in the volume V', is well known:

$$\underset{\sim}{A}(\underset{\sim}{r},t) = \frac{\mu}{4\pi} \int_{V'} \frac{j(r',t)}{(\underset{\sim}{r}-\underset{\sim}{r}')} \, dV' \, . \qquad (1)$$

The scalar potential ϕ is obtained with a similar formula:

$$\phi(\underset{\sim}{r},t) = \frac{1}{4\pi\varepsilon} \int_{V'} \frac{(\underset{\sim}{r}',t)}{(\underset{\sim}{r}-\underset{\sim}{r}')} \, dV' \qquad (2)$$

where δ is the volume charge distribution within V', j and δ are assumed to be zero outside V' and the magnetic permeability

μ and the electric permittivity ε are both assumed to be uniform and constant.

Equations (1) and (2) show close connection between $\underset{\sim}{j}$ and $\underset{\sim}{A}$ also between δ and ϕ, but they are too general to demonstrate any close $\underset{\sim}{j}$ or δ-dependence in 2-D magnetotellurics.

The electric and magnetic fields can be obtained from the scalar and vector potentials by the following defining formulas:

$$\underset{\sim}{E} = - \frac{\partial \underset{\sim}{A}}{\partial t} - \text{grad } \phi \quad , \tag{3}$$

$$\underset{\sim}{H} = \frac{1}{\mu} \text{ rot } \underset{\sim}{A} \quad . \tag{4}$$

There are two possible ways to find 2-D relations between the electromagnetic potentials and the current or charge distributions:

a) by carrying out an integration of the general solutions (1) and (2) for some $\underset{\sim}{j}$ and δ distributions in a 2-D volume V';
b) under suitable conditions the 2-D electromagnetic field components can be determined first as a function of $\underset{\sim}{j}$ and δ and definitions (3) and (4) can then be used to find 2-D expressions for $\underset{\sim}{A}$ and ϕ.

Equations (1) and (2) cannot be used for a 2-D derivation because of divergence problems in the integration of the general functions $\underset{\sim}{j}$ and δ. Method (b) consists of two steps:

1. determination of $\underset{\sim}{E}$ and $\underset{\sim}{H}$ as functions of $\underset{\sim}{j}$ and δ,
2. application of Eqs (3) and (4) to derive $\underset{\sim}{A}$ and ϕ.

The main problem in method (b) is to find the missing relations between the electromagnetic field components and the current and charge systems.

SUBSURFACE ELECTROMAGNETIC FIELD IN TERMS OF CURRENT DISTRIBUTION AT DETPH

The missing equations needed to determine the potentials have been derived by Szarka and Fischer (1989, 1990). The

general fomulas are also valid for three-dimensional problems, but in Table I only the three simplest magnetotelluric configurations are considered, i.e., 1-D, 2-D/E- or H-polarization.

All equations will be expressed in a right-hand coordinate system x, y, z, where x is the strike direction, y is along the profile and z is directed downwards.

According to Table I in the 1-D situation the horizontal magnetic components are determined exclusively by currents which flow beneath the observation point, assumed to be located at a depth z. The corresponding electric field is determined by a linear moment of currents flowing below the observation point. It is a remarkable fact that currents above the observation point do not play any explicit role in these equations, as it was shown by Szarka and Fischer (1989, 1990). (Any current deflection above the observation point would obviously have an influence on the current distribution below z.)

It is a crucial point in the derivation that in both polarizations of the electromagnetic field components are always given as a sum of their normal (one-dimensional) and anomalous parts, where the 1-D component varies from place to place according to the actual geoelectric parameters.

For example if C is an arbitrary 2-D field component, in Table I the following representation is given:

$$C(y,z) = C_{1\text{-}D}(y,z) + C_a(y,z) \ , \qquad (5)$$

where $C_{1\text{-}D}$ is the corresponding 1-D field component and C_a means any 2-D anomalous term.

The E-polarization relations are similar to the one-dimensional ones, the only difference being that for the current density an expression $\left(j_{x_a} - \frac{\partial H_z}{\partial y}\right)$ is to be used instead of j_{x_a}.

In H-polarization the magnetic field is given by the sum of its one-dimensional value and the amount of anomalous horizontal current flowing below the depth z. However, when

$z = 0$ the total integrated anomalous current vanishes which leads to the well-known constancy of H_x. The horizontal electric field in H-polarization consists of three parts: a normal part and two anomalous ones.

Table I is a summary of the relations needed to carry out the first step toward determination of the 2-D potentials.

DETERMINATION OF THE 2-D ELECTROMAGNETIC POTENTIAL COMPONENTS

Table II shows the relations based on Eqs (3) and (4) between the electromagnetic field components and the 2-D potentials. In all of the three basic magnetotelluric situations the components actually excited are only listed.

It is quite remarkable that the components of the 2-D vector potential and the scalar potential can be directly determined by simple comparison of Tables I and II. The results are summarized in Table III.

The one-dimensional case

In the one-dimensional geometry it is only necessary to consider one horizontal vector potential component which has the same direction as the electric field. In the parentheses we give the vertical derivatives, obtained from the magnetic field components. (The necessary equivalence between these two apparently different formulas can be shown by a simple derivation.)

E-polarization configuration

The E-polarization equations for the vector potential are shown in the central part of Table III.

From Tables I and II the vector potential which has a single component, is obtained at once:

$$A_{x_E} = A_{x_{1\text{-}D}} - \mu \int_z^\infty (z-z') \left(j_{x_a} - \frac{\partial H_z}{\partial y} \right) dz' \,. \qquad (6)$$

H-polarization configuration

The H-polarization problem requires a more detailed discussion.

In Table I we see that the horizontal electric field is given by:

$$E_{y_H} = E_{y_{1-D}} + i\omega\mu \int_z^\infty (z-z')\, j_{a_y}\, dz' - \frac{\partial}{\partial y} \int_z^\infty E_z\, dz' \,. \qquad (7)$$

According to Table II the following relation is valid between E_{y_H} and the potentials:

$$E_{y_H} = E_{y_{1-D}} - i\omega\, A_{y_a} - \frac{\partial \phi}{\partial y} \,. \qquad (8)$$

Solving Eqs (7) and (8) for A_{y_a} and ϕ we obtain the following results:

$$A_{y_a} = \int_z^\infty (z-z')\, j_{y_a}\, dz' \qquad (9)$$

$$\phi = \int_z^\infty E_z\, dz' + \text{constant}, \qquad E_z = - \frac{\partial \phi}{\partial z} \,. \qquad (10)$$

Equations (9) and (10) have a simple but important physical meaning: the anomalous vector potential component A_{y_a} behaves like a linear moment of anomalous currents flowing beneath the observation point; the scalar potential ϕ represents the work needed to transport a unit charge from infinite depth to the depth z.

If $E_z = - \frac{\partial \phi}{\partial z}$, then A_z must be zero in the following equation

$$E_z = - i\omega A_z - \frac{\partial \phi}{\partial z} \,. \qquad (11)$$

It means that the vector potential has only a single component A_y, and at the same time E_z is simply the vertical Coulombian field of the boundary charges.

For the magnetic field the following equations are to be compared:

$$H_x = H_{x_{1-D}} - \int_z^\infty j_{y_a}\, dz \tag{12}$$

$$H_x = H_{x_{1-D}} + \frac{1}{\mu}\frac{\partial A_z}{\partial y} - \frac{1}{\mu}\frac{\partial A_{y_a}}{\partial z}\,. \tag{13}$$

If $A_z = 0$ then $\frac{\partial A_z}{\partial y}$ is zero, too, and from equations (12) and (13)

$$\frac{\partial A_{y_a}}{\partial z} = \mu \int_z^\infty j_{y_a}\, dz' \tag{14}$$

must hold. At the $z = 0$ surface the integrated anomalous currents vanish, therefore $\frac{\partial A_{y_a}}{\partial z} = 0$ for $z = 0$.

Now in Eq. (8) we have two apparently independent anomalous terms for the horizontal electric field, but A_{y_a} and ϕ are closely related, since, according to the so-called Lorentz condition,

$$\operatorname{div} \underset{\sim}{A} + \varepsilon\mu \frac{\partial \phi}{\partial t} = 0. \tag{15}$$

About the Lorentz condition

As a consequence, for the field components excited under H-polarization induction at the $z = 0$ surface we have:

$$\frac{\partial A_{y_a}}{\partial y} = - i\omega\varepsilon\mu\phi\,. \tag{16}$$

According to Eq. (16) the lateral variation of the anomalous part of the vector potential is determined by a time derivative of the scalar potential. As is well known, ϕ is due to the free charges appearing at resistivity interfaces and the sign of these charges varies with the current direction. The rate of change of charges results in a secondary current system

described by the anomalous vector potential A_{y_a}. These secondary currents produce in-phase magnetic fields and the time-variation of these magnetic fields induces the anomalous horizontal electric field $E_{y_{ja}}$. Origin of $E_{y_{ja}}$ is solely in the vector potential. (The total anomalous electric field component $E_{y_a} = E_{y_{ja}} + E_{y_{qa}}$, where $E_{y_{qa}}$ is the scalar potential-origin term.)

Finally, the lateral variation of $E_{y_{ja}}$, can be given by the following expression

$$\frac{\partial E_{y_{ja}}}{\partial y} = -\varepsilon\mu\omega^2\phi \ . \tag{17}$$

This relation means that the part of the anomalous electric field originating from the vector potential is only induced by the second time-derivative of the original boundary charges. In order to give a simple point-charge estimation for the relative importance of the anomalous vector and scalar potential terms, we write

$$\left| \frac{\frac{\partial E_{y_{ja}}}{\partial y}}{\frac{\partial E_{y_{qa}}}{\partial y}} \right| \sim \frac{\frac{1}{4\pi\varepsilon}\varepsilon\mu\omega^2\frac{q}{r}}{\frac{1}{4\pi\varepsilon}\frac{q}{r^3}} = \left(\frac{\omega r}{c}\right)^2 \tag{18}$$

where q is a point charge at a distance r from the observation site and c is the speed of light.

From Eq. (18) it can be seen that the lateral variation of the anomalous vector potential term is only comparable to the lateral variation due to the scalar potential at very long distances or at very high frequencies. Equation (18) means that in H-polarization practically the whole electric anomaly along the investigated profile originates from the direct Coulombian field of charges appearing at resistivity interfaces.

In Table III both vector and scalar potentials as well as the Lorentz condition are summarized.

E- AND H-POLARIZATION DISTORTIONS AT THE SURFACE

In Table IV and V all the vector potential components which occur, the scalar potential and the electric and magnetic field components are summarized for $z = 0$, as explicit functions of current density and implicit functions of the charge distribution (instead of charges we have given their associated potentials). The physical meaning of these quantitites is also given succintly. All of them, excepting H_z, have been discussed already. The vertical magnetic component in E-polarization can be interpreted as a lateral variation of the linear moment of the subsurface E-polarization currents $\left(j_E = j_x - \frac{\partial H_z}{\partial y}\right)$.

In the bottom part of Tables IV and V some meaningful combinations are also mentioned. Most of them of course, are not new applications. The aim of this summary is to point out their direct physical meaning.

1. The impedance, both for E- and H-polarizations, can be interpreted as a complex mean depth of subsurface currents (Szarka and Fischer 1989).
2. In E-polarization the expression $i\omega\mu H_z/E_x$ (suggested to use at first by Takács 1976) seems to be especially useful as it can physically be interpreted as a sort of normalized lateral variation of moment of subsurface currents.
3. Since H_z is perfectly free from H-polarization effects, and E_z is free from E-polarization distortions (it depends exlusively on charges), their systematic measurements could serve as basis to separate the vector potential- and scalar potential-origin effects.

SUMMARY

On the basis of earlier works (Szarka and Fischer 1989, 1990) relations between the electromagnetic fields and the subsurface current system were compared with the definitions of the potentials. The results of this paper can be summarized as follows:

Table I. Electric and magnetic field components in terms of currents at depth in the three basic magnetotelluric situations

1-D (E_x, H_y or E_y, H_x)	E-POL (E_x, H_y, H_z)	H-POL (H_x, E_y, E_z)
$E_{x_{1\text{-}D}}(z) = i\omega\mu \int_z^\infty (z'-z)\, j_{x_{1\text{-}D}}(z')\, dz'$	$E_{x_E}(y,z) = E_{x_{1\text{-}D}}(z) + i\omega\mu \int_z^\infty (z'-z) \left(j_{x_a}(y,z') - \frac{\partial H_z(y,z')}{\partial y} \right) dz'$	$E_{y_H}(y,z) = E_{y_{1\text{-}D}}(z) + i\omega\mu \int_z^\infty (z'-z)\, j_{y_a}(y,z')\, dz' - \frac{\partial}{\partial y} \int_z^\infty E_z(y,z')\, dz'$
$H_{y_{1\text{-}D}}(z) = \int_z^\infty j_{x_{1\text{-}D}}(z')\, dz'$		
$E_{y_{1\text{-}D}}(z) = i\omega\mu \int_z^\infty (z'-z)\, j_{y_{1\text{-}D}}(z')\, dz'$	$H_{y_E}(y,z) = H_{y_{1\text{-}D}}(z) + \int_z^\infty \left(j_{x_a}(y,z') - \frac{\partial H_z(y,z')}{\partial y} \right) dz'$	
$H_{x_{1\text{-}D}}(z) = - \int_z^\infty j_{y_{1\text{-}D}}(z')\, dz'$		$H_{x_H}(y,z) = H_{x_{1\text{-}D}}(z) - \int_z^\infty j_{y_a}(y,z')\, dz'$

Table II. Relations between the electromagnetic field components and the potentials in the three basic magnetotelluric situations

1-D E_x, H_y or E_y, H_x	E-POL (E_x, H_y, H_z)	H-POL (H_x, E_y, E_z)
$E_{x_{1\text{-}D}}(z) = -i\omega A_{x_{1\text{-}D}}(z)$	$E_{x_E}(y,z) = E_{x_{1\text{-}D}}(z) - i\omega A_{x_a}(y,z)$	
$H_{y_{1\text{-}D}}(z) = \frac{1}{\mu}\frac{\partial A_{x_{1\text{-}D}}(z)}{\partial z}$	$H_{y_E}(y,z) = H_{y_{1\text{-}D}}(z) + \frac{1}{\mu}\frac{\partial A_{x_a}(y,z)}{\partial z}$	
$E_{y_{1\text{-}D}}(z) = -i\omega A_{y_{1\text{-}D}}(z)$	$H_{z_E}(y,z) = -\frac{1}{\mu}\frac{\partial A_{x_a}(y,z)}{\partial z}$	$E_{y_H}(y,z) = E_{y_{1\text{-}D}}(z) - i\omega A_{y_a}(y,z) - \frac{\partial \phi(y,z)}{\partial y}$
$H_{x_{1\text{-}D}}(z) = -\frac{1}{\mu}\frac{\partial A_{y_{1\text{-}D}}(z)}{\partial z}$		$E_{z_a}(y,z) = -i\omega A_{z_H}(y,z) - \frac{\partial \phi(y,z)}{\partial z}$
		$H_{x_H}(y,z) = H_{x_{1\text{-}D}}(z) + \frac{1}{\mu}\frac{\partial A_{z_a}(y,z)}{\partial y} - \frac{1}{\mu}\frac{\partial A_{y_a}(y,z)}{\partial z}$

Table III. The vector potential components and the scalar potential in terms of subsurface current- and charge systems in the three basic magnetotelluric situations

1-D (E_x, H_y or E_y, H_x)	E-POL (E_x, H_y, H_z)	H-POL (H_x, E_y, E_z)
$A_{x_{1\text{-}D}}(z) = -\mu \int_z^\infty (z'-z)\, j_{x_{1\text{-}D}}(z')\, dz'$	$A_{x_E}(y,z) = A_{x_{1\text{-}D}}(z) -$	
$\left(\frac{\partial A_{x_{1\text{-}D}}(z)}{\partial z} = \mu \int_z^\infty j_{x_{1\text{-}D}}(z')\, dz' \right)$	$- \mu \int_z^\infty (z'-z) \left[j_{x_a}(y,z') - \frac{\partial H_z(y,z')}{\partial y} \right] dz'$	
$A_{y_{1\text{-}D}}(z) = -\mu \int_z^\infty (z'-z)\, j_{y_{1\text{-}D}}(z')\, dz'$	$\left(\frac{\partial A_{x_a}(y,z)}{\partial z} = \mu \int_z^\infty \left[j_{x_a}(y,z') - \frac{\partial H_z(y,z')}{\partial y} \right] dz' \right)$	$A_{y_H}(y,z) = A_{y_{1\text{-}D}}(z) - \mu \int_z^\infty (z'-z)\, j_{y_a}(y,z')\, dz'$
$\left(\frac{\partial A_{y_{1\text{-}D}}(z)}{\partial z} = \mu \int_z^\infty j_{y_{1\text{-}D}}(z')\, dz' \right)$		$\left(\frac{\partial A_{z_a}(y,z)}{\partial y} - \frac{\partial A_{y_a}(y,z)}{\partial z} = \mu \int_z^\infty j_{y_a}(y,z')\, dz' \right)$
		$\phi_H(y,z) = \int_z^\infty E_z(y,z')\, dz'$
		Lorentz-condition, $A_{z_a}(y,z) \equiv 0$, see Eq. (15) :
		$\frac{\partial A_{y_a}(y,z)}{\partial y} + i\omega\varepsilon\mu\, \phi_H(y,z) = 0$

Table IV. E-polarization relations at the surface

Name	Definition	Physical meaning
vector potential	$A_x(y) = -\mu \int_0^\infty z \left[j_x(y,z) - \frac{\partial H_z(y,z)}{\partial y} \right] dz$ $A_y(y) = A_z(y) = 0$	moment of subsurface E-polarization currents producted by $-\mu$
scalar potential	$\phi(y) = 0$	
electric field	$E_x(y) = i\omega\mu \int_0^\infty z \left[j_x(y,z) - \frac{\partial H_z(y,z)}{\partial y} \right] dz$ $E_y(y) = E_z(y) = 0$	moment of subsurface E-polarization currents producted by $i\omega\mu$
magnetic field	$H_x(y) = 0$	
	$H_y(y) = \int_0^\infty \left[j_x(y,z) - \frac{\partial H_z(y,z)}{\partial y} \right] dz$	the subsurface E-polarization currents
	$H_z(y) = \frac{\partial}{\partial y} \int_0^\infty z \left[j_x(y,z) - \frac{\partial H_z(y,z)}{\partial y} \right] dz$	lateral variation of moment of subsurface E-polarization currents
meaningful combinations	$E_x(y)/H_y(y)$	complex mean depth of subsurface E-polarization currents
	$i\omega\mu\, H_z(y)/E_x(y)$	the relative value of lateral variation of moment due to subsurface E-polarization currents
	$E_x(y)/i\omega\mu$ or $\int H_z(y)\, dy$	the vectorpotential (not influenced by H-polarization effects)

Table V. H-polarization relations at the surface

Name	Definition	Physical meaning
vector potential	$A_x(y) = A_z(y) = 0$ $A_y(y) = -\mu \int_0^\infty z\, j_y(y,z)\, dz$	moment of subsurface currents producted by $-\mu$
scalar potential	$\phi(y) = \int_0^\infty E_z(y,z)\, dz$	work needed to transport the unit charge from infinite depth to the surface
electric field	$E_x(y) = 0$ $E_y(y) = i\omega\mu \int_0^\infty z\, j_y(y,z)\, dz - \frac{\partial \phi(y)}{\partial y}$ $E_z(y) = 0 \left(E_z(y,z) = -\frac{\partial \phi}{\partial z} \right)$	moment of subsurface currents producted by $i\omega\mu$ minus the horizontal Coulombian field of boundary charges (the vertical Coulombian field of boundary charges)
magnetic field	$H_x(y) = -\int_0^\infty j_y(y,z)\, dz$ $H_y(y) = H_z(y) = 0$	the subsurface current
meaningful combinations	$-E_y(y)/H_x(y)$	complex mean depth of subsurface currents
	$E_z(y,z)$	the vertical Coulombian field of boundary charges (not influenced by E-polarization effects)

1. Meaningful integral formulas have been derived for the electromagnetic potentials which can probably not be obtained in any other way.
2. The E- and H-polarization distortions have been expressed in terms of corresponding components of the vector and scalar potentials. The sources of E- and H-polarization distortion terms at the surface are twofold:
   a) In E-polarization the anomalous part of the single-component vector potential results in an anomalous behaviour of both the electric and the magnetic field components.
   b) In H-polarization the distortions seem to be of two different origins: the scalar potential and the single-component vector potential, but these distortions are not independent of each other because the potentials are related by the Lorentz condition. In practice in the lateral variation of the electric anomaly the scalar potential term is always the dominating factor, and the anomalous vector potential term can be neglected.

Any method to get rid of three-dimensional distortions must begin with an exact recipe for the elimination of distortions in both H- and E-polarizations. Presentation of potential-oriented formulas and the discussion about the possible physical meaning of several known electromagnetic parameters could be one step in this attempt.

REFERENCES

Berdichevsky M N, Dmitriev V I 1976: In: Geoelectric and Geothermal Studies, KAPG Geophysical Monograph (Ádám A ed.). Akadémiai Kiadó, Budapest, 165-221.

Jiracek G R 1988: Near-surface and topographic distortions in electromagnetic induction. Review paper, presented at the IAGA EM workshop, October 24-31, Sochi

Jones A G 1983: Geophysical Survey, 6, 79-122.

Jones A G 1988: Geophysics, 53, 967-978.

Menvielle M 1988: Surveys in Geophysics, 9, 319-348.

Oldenburg D 1988: Inversion of electromagnetic data: an overview of new techniques. Review paper, presented at the IAGA EM workshop, October 24-31, Sochi

Spichak V V 1988: Electromagnetic field transformations and their use in interpretation. Review paper, presented at the IAGA EM workshop, October 24-31, Sochi

Szarka L, Fischer G 1989: Geophysical Transactions, 35, 157-172.

Szarka L, Fischer G 1990: Subsurface electromagnetic parameters in terms of currents at depth. Geophysical Transactions (in press)

Takács E 1976: In: Geoelectric and Geothermal Studies, KAPG Geophysical Monograph (Ádám A ed.). Akadémiai Kiadó, Budapest, 264-273.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 423-434 (1991)

EFFECT ON STABILITY OF THE INVERSE SOLUTION WITH 1 OHMM CHANGE IN EMBEDDED 'THICK' CONDUCTING LAYER'S THRESHOLD RESISTIVITY

Sri Niwas[1] and Pravin K Gupta[2]

[1]Department of Geophysics, Kurukshetra University, Kurukshetra-132 119, India
[2]Department of Earth Sciences, University of Roorkee, Roorkee-247 667, India

[Manuscript received March 11, 1991]

The problem of equivalence in resistivity data inversion culminates in "extremely small eigenvalues". These extremely small eigenvalues are the root cause of unstable inverse solutions. The question, how small the eigenvalues may be to qualify them as "extremely small ones" so that they still enable stable iterative methods to be used, is investigated here more critically by singular value analysis (SVA) of the matrix operator. It is found that even a thick layer can lead to unstable solutions if the resistivity is below a threshold value. This threshold will of course be changing when compared with the resistivity of the upper layer and bottom layers. The whole process is analysed by taking H-type curves.

Keywords: eigenvalues; equivalence; inverse solution; matrix operator; resistivity

1. INTRODUCTION

The exact quantitative interpretation of geometrical resistivity (e.g. Schlumberger) sounding data is, popularly, carried out using automatic iterative methods. The generalized linear inverse (GLI) technique, based on singular value decomposition, is the most versatile of all the available ones (Inman et al. 1973, Vozoff and Jupp 1975, Johansen 1977). The quality of an inverse solution is critically judged by the criteria of (1) existence (2) construction, (3) uniqueness and (4) stability. Because of the stability criterion, the quantitative interpretation of resistivity sounding data poses serious problems of reliability of the inverse solution from softrock terrain, comprising alternation of sand-clay/shale sequences, unlike the hard rock data. It is so because of the phenomena of "suppression" for ascending or descending type

curves and of 'equivalence' for K- or H-type curves (Maillet 1947, Rocroi 1975, Koefoed 1976, 1979). Bhattacharya and Patra (1968) have shown that the layer parameters, in case of a thin layer problem, happen to be linearly dependent when the thin layer is either highly conducting ($S=d/\varrho$) or highly resistive ($R=d\varrho$). The linear dependence of layer parameters results in a near singular system matrix. The presence of a 'thin' conducting or resistive layer, which is poorly represented in the data, destroys the orthogonality of the inverse problem. As a result the eigenvalues are either zero or 'extremely small'. Invoking the compatibility condition of Lanczos (1961), the zero eigenvalues and their associated eigenvectors can be excluded while solving the singular linear systems. However, the 'extremely small' eigenvalues, indicating a near-singular system, result in unstable solution in case of erroneous data, possessing enhanced variance which, in turn, affects the resolution. In order to enhance the resolution one can still follow, at the cost of accuracy, the Lanczos (1961) procedure and exclude the small eigenvalues and their associated eigenvectors also from the system.

However, for a better trade-off between resolution and error, the 'ridge-regression inversion' (Marquardt 1963, 1970, Horel and Kennard 1970a, 1970b, Inman 1975) or 'damped least square inversion' technique is more useful. It has been concluded by Marquardt (1963) that the GLI should be preferred for problems with mostly large eigenvalues and some zero eigenvalues, while the 'ridge regression' should be preferred for problems with some very small eigenvalues. However, it may be stressed here that 'ridge regression' is a remedy for a particular situation and should not be resorted to indiscriminately. Whether the 'ridge regression' is required or not, must be investigated through the 'condition number', which is the ratio of the largest eigenvalue (λ_i) with subsequent eigenvalues (λ_i).

Over the years the authors have been inverting resistivity sounding data procured from both hard rock terrain and alluvium zones of India. In their opinion also hard rock data are less

vulnerable to instability than the alluvium ones. The Indian alluvium comprises a top soil cover of approximately 100 ohmm resistivity, underlain by clays and sands. The resistivity of clays range from 2 ohmm to 10 ohmm while that of sand varies between 20 ohmm and 100 ohmm. In the present note the vulnerability of the resistivity inverse solution is studied, for the first time, for an otherwise 'thick layer'.

2. METHODOLOGY

The quantitative interpretation of resistivity sounding data can be performed in resistivity domain (r-domain) or in resistivity transform domain (λ-domain) with equal vigour and accuracy (Koefoed 1979, Sri Niwas and Israil 1987b). The computation of apparent resistivity from resistivity transform and that of resistivity transform from apparent resistivity can be easily performed either by the convolution method (Ghosh 1971a, 1971b) or using the matrix method (Sri Niwas and Israil, 1986, 1987a, 1987b). However, in the present analysis we would be performing all the operations in λ-domain. The theoretical computation of the resistivity transform function for the assumed model is obtained by a recurrence relation given by Koefoed (1979).

Let us assume that the horizontally stratified earth is characterized by layer resistivities (ϱ_1, ϱ_2, ..., ϱ_N) and layer thicknesses (d_1, d_2, ..., d_{N-1}). Then the resistivity transform function can be expressed as a function of two parameter vectors,

$$T_i = T_i\,(P,U) \qquad (1)$$

where P is the vector of unknown parameters to be determined and its components being

$$P = (\varrho_1, \varrho_2, \ldots, \varrho_N, d_1, d_2 \ldots, d_{N-1}) \qquad (2)$$

and U is the vector of known parameters as

$$U = (u_1, u_2, \ldots, u_n) \tag{3}$$

where u = 1/separation constant (= half-current electrode spacing). Let P^o be an estimated solution of P and ΔP the desired deviater to render coincidence of calculated and observed resistivity transforms. Expanding (1) in Taylor's series and retaining only the first derivative term a matrix equation is obtained:

$$\underset{n\times 1}{\Delta T} = \underset{n\times m}{G}\ \underset{m\times 1}{\Delta P} \tag{4}$$

where $\Delta T = T_i(U_i P) - T_i(U_i, P^o)$, $n\times 1$ column vector of resistivity transform anomaly,

$\Delta P = P_j - P_j$, $m\times 1$ column vector of parameter corrections,

and $G = \dfrac{\partial T_i(U,P)}{\partial P_j}$ $n\times m$ system matrix of influence coefficients.

The problem as posed above has been linearized to exploit the facilities of the matrix formulation, although it means that the final solution would be obtained iteratively by updating the vector p^o after every iteration. However, for an 'ill posed' problem ($n \gtrsim m$), in a given iteration, the estimated $\hat{\Delta P}$ is obtained from the relation

$$\hat{\Delta P} = G_g^{-1}\ \Delta T \tag{5}$$

where G^{-1} is identified as g-inverse of matrix G.

Indeed, there are a number of g-inverses, but the one making use of the singular value decomposition (SVD) is particularly instructive as it quantifies many physical aspects of the inverse problem. It enables one to investigate the "goodness" of the inverse solution yielded by a given data set and analyse the relative influence of various data elements in illuminating the solution space. Accordingly Eqs (4) and (5) can be written as

$$\Delta T = U \Lambda V^t \Delta P \tag{6}$$

and

$$\Delta \hat{P} = V \Lambda^{-1} U^t \Delta T \tag{7}$$

where Λ is the diagonal matrix comprising the eigenvalues of G^tG or GG^t arranged in descending order of magnitude, V and U are, respectively, the orthogonal matrices comprising the eigenvectors of G^tG and GG^t.

The resolution of model parameters can be measured through the resolution matrix as

$$\Delta \hat{P} = R \Delta P \; ; \quad R = G_g^{-1} G = VV^t \; . \tag{8}$$

Ideally the resolution matrix R is desired to be an identity matrix.

3. DISCUSSION

We have considered a representative 3-layer earth model simulating an alluvium terrain with soil, clay and sand layers as top, intermediate and bottom layers, respectively. The model parameters of the top and bottom layers, ϱ_1 = 100 ohmm, ϱ_3 = 20 ohmm, d_1 = 10 m, and the thickness of the intermediate layers, d_2 = 30 m, are fixed. The models I, II and III have middle layer resistivities as 5, 4 and 10 ohmm, respectively. The kernel values, for inversion in λ-domain, were estimated, using the method of Sri Niwas and Israil (1986, 1987a), from the synthetic Schlumberger apparent resistivity curves for the three models. It has been pointed out by Sri Niwas and Israil (1987b) that the convergence of inverse solutions is more or less identical in both kernel domain and resistivity domain. The calculation of derivatives of the resistivity transform function with respect to model parameters is done using an expression developed by Sri Niwas et al. (1982). Figure 1 shows the convergence pattern for the three models. The initial model for three cases were taken the same, namely ϱ_1 = 100 ohmm, ϱ_2 = 7 ohmm, ϱ_3 = 20 ohmm, d_1 = 8 m, d_2 = 35 m. It is evident

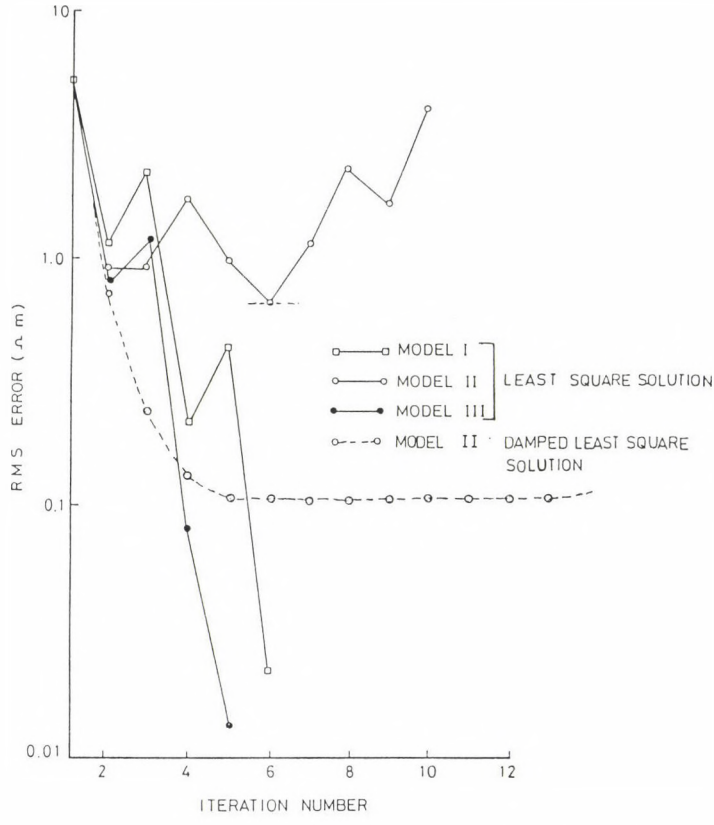


Fig. 1. Convergence pattern of rms errors in responses for different models

that models I and III possess good convergence characteristics while model II is divergent. The inverted results for model I and III are extremely accurate (0.001 percent error). In case of model II, the application of the 'ridge regression method' (damped least squares), with a starting value of the damping factor equal to 1 and its variation rate with each advancing iteration 1/3, although gets rid of divergence, yet the asymptotic rms error is still unacceptable. Since model III has nice convergence behaviour, subsequently, its results are not reproduced. It is instructive to note in Table I the inverted values of the layer parameters for model II (least squares and ridge regression). The inverted values for model II are

Table I. Inverted results of model II through Least Squares Inversion (LSI) and Ridge Regression Inversion (RRI)

Model parameter	True	Initial	L.S. Results of 6th iteration	R.R. Results of 6th iteration	R.R. Results of 12th iteration
ϱ_1 (ohmm)	100	100	96.826	97.936	97.873
ϱ_2 (ohmm)	4	7	4.731	3.999	3.564
ϱ_3 (ohmm)	20	20	20.314	19.890	19.860
d_1 (m)	10	8	10.458	10.070	10.070
d_2 (m)	30	35	35.477	35.055	30.050

extremely bad in case of d_2 and not so good in case of ϱ_2. Even using the ridge regression method there is no improvement up to the 5th iteration, however, at the 12th iteration there is a marked improvement in d_2 values but at the same time there is decline in ϱ_2 values. However, the conductance values remain more or less the same (8.75; 8.43).

The difference in models I and II is 1 ohmm in the resistivity of the middle layer. Figure 2 shows the variation of percentage error in layer parameters of models I and II with iteration number. Unlike model I, where errors in all parameters tend smoothly and quickly to zero, errors in the top and bottom layer parameters of model II show oscillatory but convergent behaviour while the intermediate layer parameters tend to diverge. This observation points towards the unstable nature of the data inversion for model II. Figure 3 gives the Singular Value Analysis (SVA) of the matrix G. It is apparent that the parameters associated with 'large' eigenvalues converge quickly and uniformly whereas the parameters associated with 'small' eigenvalues converge slowly. The parameter d_2 is associated with the smallest eigenvalues. Hence its convergence and resolution is weak and poor.

In order to ascertain the probable existence of inherent features discriminating the two models, the 'condition number' plot for the two models has been studied (Fig. 4). It reveals

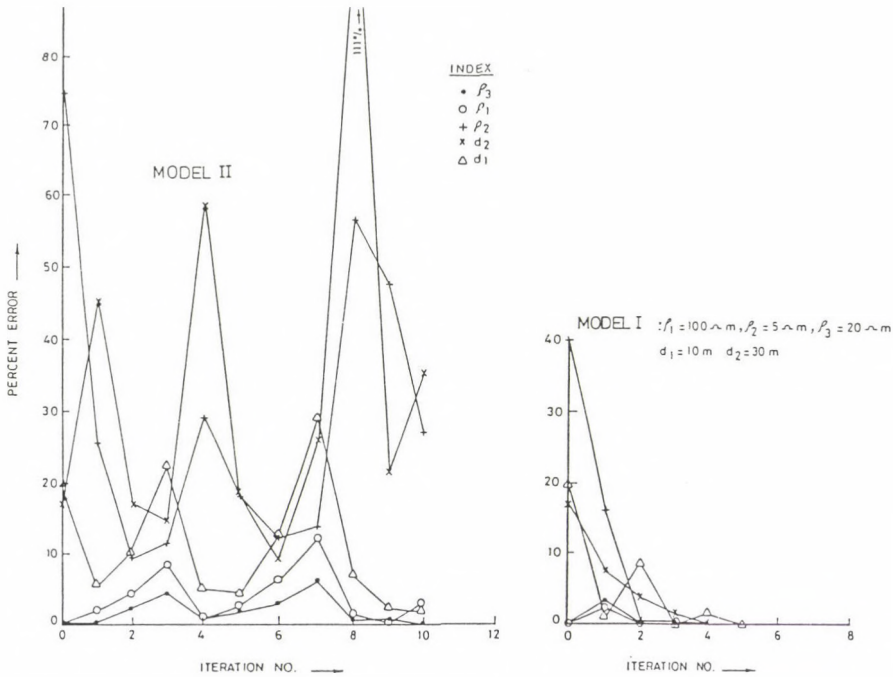


Fig. 2. Convergence pattern of rms errors in layer parameters for models I and II

no discriminatory difference between the two cases except for a minimum difference in the gradient of the last portions of the curves. The system matrices for inversion have similar structures in the two cases, yet the inversion for model I is stable while that for model II is unstable. Table II presents the structure of resolution matrices for models I and II. There is a drastic smearing of the elements of the matrix for model II in comparison to model I. It appears that in the present case an intermediate layer resistivity of 5 ohmm (ϱ_2/ϱ_3 = 5/20 and ϱ_2/ϱ_1 = 5/100), is the threshold value beyond which the equivalence effect becomes severe.

The final attempt to decipher why model II inversion is unstable, was made by using Pylaev's nomograms. The diagonals of rectangles of equivalence for models I, II and III are shown in Fig. 5. It is evident that the rectangles of equivalence for

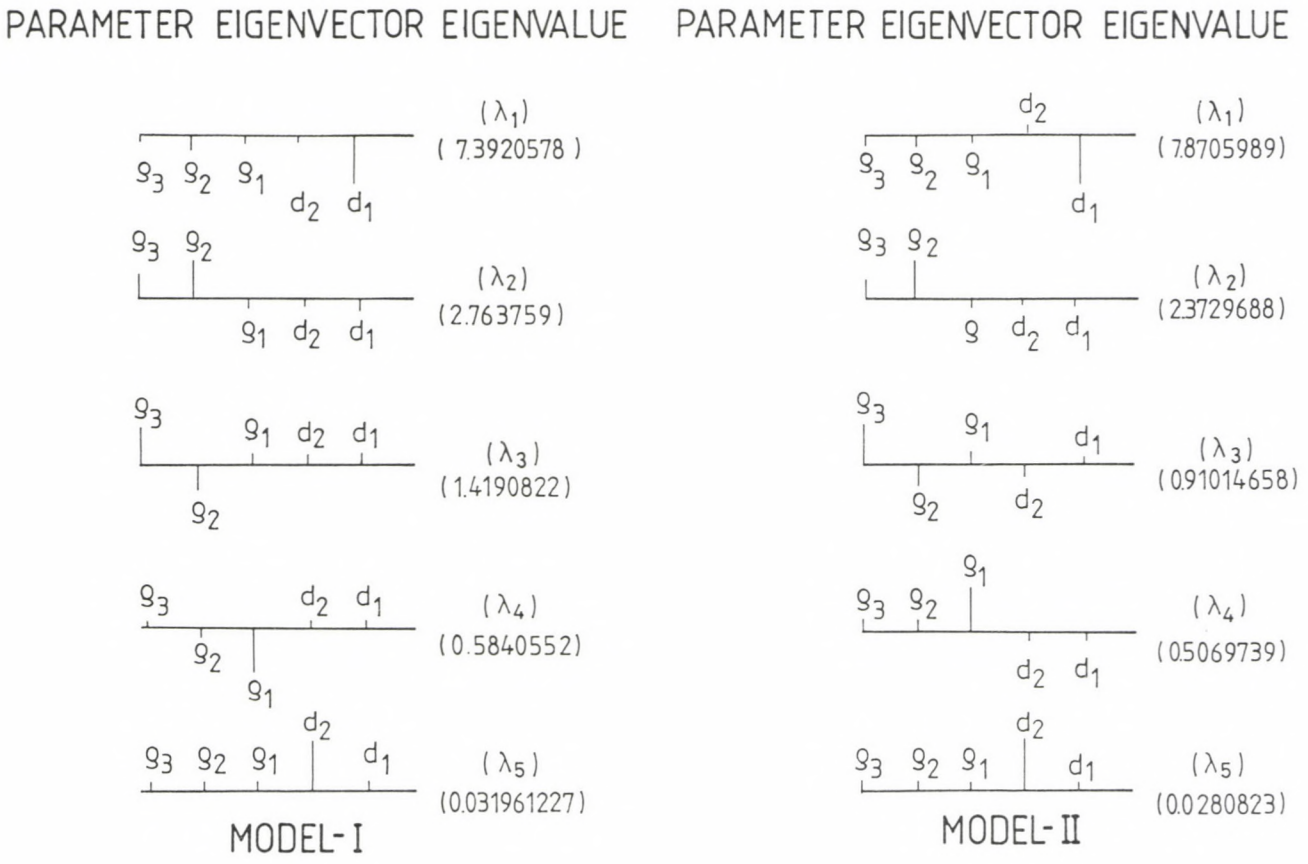


Fig. 3. Singular value analysis of matrix G for Models I and II

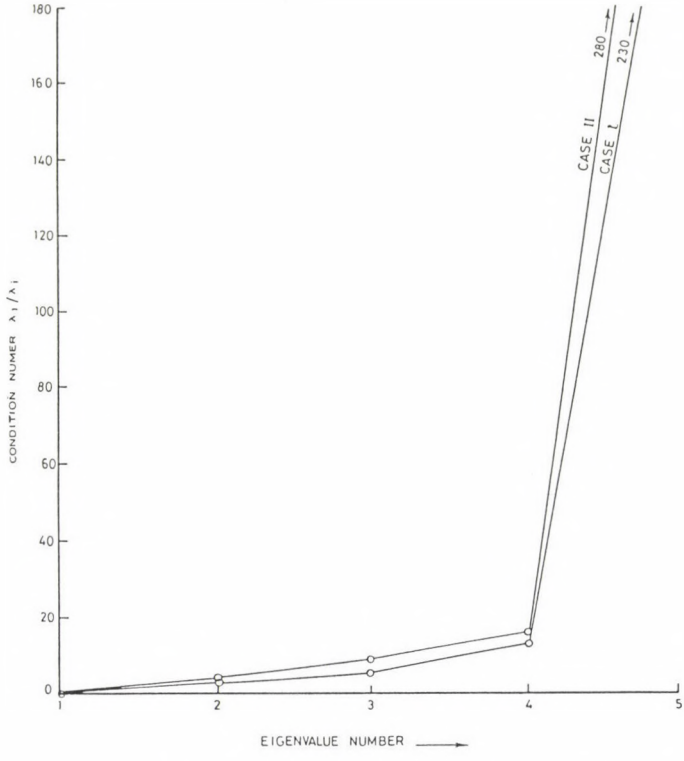


Fig. 4. Condition number plots for models I and II

Table II. Structure of Resolution Matrix R for model I and model II for LSI

Model I

$$\begin{bmatrix} \Delta\hat{\varrho}_1 \\ \Delta\hat{\varrho}_2 \\ \Delta\hat{\varrho}_3 \\ \Delta\hat{d}_1 \\ \Delta\hat{d}_2 \end{bmatrix} = \begin{bmatrix} 0.9999 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.9999 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.9999 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.9999 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.9999 \end{bmatrix} \begin{bmatrix} \Delta\varrho_1 \\ \Delta\varrho_2 \\ \Delta\varrho_3 \\ \Delta d_1 \\ \Delta d_2 \end{bmatrix}$$

Model II

$$\begin{bmatrix} \Delta\hat{\varrho}_1 \\ \Delta\hat{\varrho}_2 \\ \Delta\hat{\varrho}_3 \\ \Delta\hat{d}_1 \\ \Delta\hat{d}_2 \end{bmatrix} = \begin{bmatrix} 0.9786 & 0.0122 & 0.1805 & -0.0034 & -0.0979 \\ 0.0122 & 0.8109 & -0.3660 & 0.0468 & -0.4539 \\ 0.1805 & -0.3660 & -0.9037 & 0.0936 & 0.0893 \\ -0.0034 & 0.0408 & 0.0936 & 0.9884 & 0.1099 \\ -0.0978 & -0.4539 & -0.0893 & 0.1099 & -0.8742 \end{bmatrix} \begin{bmatrix} \Delta\varrho_1 \\ \Delta\varrho_2 \\ \Delta\varrho_3 \\ \Delta d_1 \\ \Delta d_2 \end{bmatrix}$$

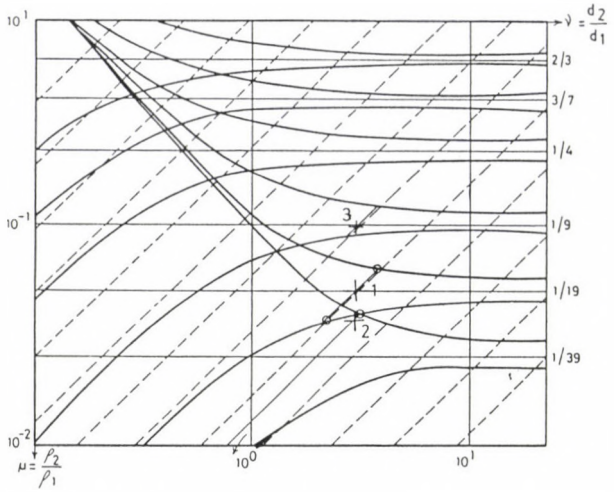


Fig. 5. Rectangles of equivalence on Pylaev's nomogram for different models
Case I: $\varrho_1 = 100\ \Omega$ m, $\varrho_2 = 5\ \Omega$ m, $\varrho_3 = 20\ \Omega$ m, $d_1 = 10$ m, $d_2 = 30$ m. Case II: $\varrho_1 = 100\ \Omega$ m, $\varrho_2 = 4\ \Omega$ m, $\varrho_3 = 20\ \Omega$ m, $d_1 = 10$ m, $d_2 = 30$ m. Case III: $\varrho_1 = 100\ \Omega$ m, $\varrho_2 = 10\ \Omega$ m, $\varrho_3 = 20\ \Omega$ m, $d_1 = 10$ m, $d_2 = 30$ m
H type Pylaev's nomogram for equivalence
----- lines of equal S_2/S_1
⟩⟩ limiting values of ϱ_2/ϱ_1 and d_2/d_1 for equivalence

models I and III are finite while that for model II, in effect, is infinite.

These results shake the common belief that data inversion, based on a thick layer model, is always stable. The equivalence problem, which is generally believed to be associated with a thin layer model, can be operative even in a thick layer model. Pylaev's nomograms provide an explanation about the critical limit for an expected stable or unstable inverse solution.

ACKNOWLEDGEMENT

Authors are indebted to Prof. V K Gaur for his inspiration and fruitful discussions.

REFERENCES

Bhattacharya P K, Patra H P 1968: Direct current geoelectric

sounding-principle and interpretation. Elsevier

Ghosh D P 1971a: Geophysical Prospecting, 19, 192-217.

Ghosh D P 1971b: Geophysical Prospecting, 19, 769-775.

Horel A E, Kennard R W 1970a: Technometrics, 12, 55-67.

Horel A E, Kennard R W 1970b: Technometrics, 12, 69-82.

Inman J R 1975: Geophysics, 40, 798-817.

Inman J R, Ryu J, Ward S H 1973: Geophysics, 38, 1088-1108.

Johansen H K 1977: Geophysical Prospecting, 25, 667-691.

Jupp D L B, Vozoff K 1975: Geophys. Soc., 42, 957-976.

Koefoed O 1976: Geophysical Prospecting, 24, 233-240.

Koefoed O 1979: Geosounding principles I. Elsevier

Lanczos C 1961: Linear differential operators. D. Van Nostrand

Maillet R 1947: Geophysics, 12, 529-556.

Marquardt D W 1963: Jr. SIAM, 11, 431-441.

Marquardt D W 1970: Technometrics, 12, 591-612.

Rocroi J P 1975: Geophysical Prospecting, 23, 765-778.

Sri Niwas, Israil M 1986: Geophysics, 51, 1594-1602.

Sri Niwas, Israil M 1987a: Geophysical Prospecting, 35, 548-569.

Sri Niwas, Israil M 1987b: Geophysics, 52, 1412-1417.

Sri Niwas, Pawan Kumar, Wason H R 1982: Proc. Indian Acad. Sci. (Earth Planet. Sci.), 91, 29-41.

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 435-452 (1991)

RESOLVING POWER OF DIRECT CURRENT AND MAGNETOTELLURIC RESISTIVITY SOUNDINGS IN EXPLORING THICK SEDIMENTARY HORIZONS

Sri Niwas and Pramod Kumar

Department of Geophysics, Kurukshetra University, Kurukshetra-132 119, India

[Manuscript received March 11, 1991]

Thick sedimentary basins having marked contrast between their different litho units are ideal targets for geoelectromagnetic exploration. It has been brought out here that the D.C. resistivity and MT soundings are the best available tools to get information regarding conductivity distribution in the sub-surface layers. Data are interpreted using powerful ridge regression inversion. It has been shown by singular value analysis that the conducting layers are best resolved by MT data but resistive layers are poorly resolved. However, the resistive layers are best resolved using D.C. data but conducting layers are poorly resolved. Thus, D.C. and MT data are complimentary to each other and for complete subsurface information joint inversion may be tried. It is also established through extensive numerical experiments that D.C. data are much more prone to the principle of equivalence in comparison to MT data. However, D.C. data provide marked contrast between two formations, a need to start the inversion process.

Keywords: D.C. sounding; magnetotellurics; singular value analysis; resolving power

1. INTRODUCTION

Exploration geophysics is guided by the intellectual curiosity about the nature of the earth and also by the more secular need to know what resources and material are available to the mankind, where and in what quantities. Buried reserves of hydrocarbon resources are one such target which have become an important need for any developed/developing country. Thick sedimentary basins deposited under marine environments are most pronounced zones for the exploration for hydrocarbons. Reflection seismics and electromagnetics are considered as direct means of exploration of thick sedimentary basins and hence the hydrocarbon probabilities on the basis of thumb rule

that the thicker such a sedimentary column is the higher is the possibility of hydrocarbon resources. Maps of the parameter S (electrical unit conductance) are equivalent to seismic reflection time maps of t_0; the problem of the migration of these maps to depth-maps depends on the knowledge of the resistivity distribution or the velocity distribution. Generally the resolution of seismics is much higher than that of the geoelectrical responses, however, in some special cases like a low velocity zone lying below a high velocity zone, the resolution of seismics becomes suspect and that of the electromagnetic method becomes pronounced if the low velocity zone is highly conducting.

Porstendorfer (1975) attempted to review schematically the various methods to 'sound' the various sub-surface layers of thick sedimentary basins by using one form or another of electromagnetic probing. The sedimentary basins in consideration are characterized by a thick (2 km) well conducting top layer with a resistivity 1.8 ohmm followed by an insulating (resistivity 10^4 ohmm) salt layer of a thickness 1 km and 6.5 km sediments of medium resistivity (19 ohmm), lying over a resistive (resistivity = ∞) basement, thereby emerges a picture of a 4-layer section. It is well known that the low frequency electromagnetic waves capable to penetrate to depths are influenced by horizontal and vertical anomalies of the resistivity which produce distortions in the form of secondary fields used as convenient means to obtain knowledge about the sub-surface. However, the interaction of electromagnetic waves with the earth can be very complicated, even for a fairly simple geometry. Fortunately, the situation is much easier to handle at direct current excitation which is the limiting case of zero frequency and is an excellent starting point for such complicated physical phenomena. Porstendorfer (1975) has considered four possibilities for getting the geoelectrical profile of the sedimentary basin viz. (1) Deep electrical sounding (2) Frequency sounding (3) Transient electromagnetics and (4) Magnetotelluric sounding. However, there is a fundamental physical limitation of such

surface measurements in the form of an adverse masking influence of the overburden layers. When direct current is employed, the injected current tends to be diverted from the deeper regions by the upper layer and useful information from the target is completely screened (in case of an insulating subsequent layer). In the case of alternating current or transient excitation, there is an additional degradation caused by the attenuation of the propagated signal over the two-way path. In spite of such forbidding limitations, great efforts have been made over the years in overcoming the hurdles of the physics.

Let us study schematically the relative strength of all the four probing techniques in case of the above mentioned four-layer earth. In case of the direct current method the physics of the layered structure in question clearly indicates that far away from the source, the current lines flow in a horizontal plane parallel to the interface, thus vertical planes are equipotential surfaces and as such they give constant measured resistance irrespective of the subsurface condition and electrode separation. The apparent resistivity curve asymptotically approaches unit longitudinal conductance $S_1(=d_1/\varrho_1)$ of the layer over the salt layer. And it is virtually impossible to get any information about the sediments below the salt because of the screening effect. The methods using alternating field are helpful in such cases due to the law of the skin-effect. The 'frequency sounding' is capable of giving $S_3(=d_1/\varrho_1 + d_2/\varrho_2 + d_3/\varrho_3)$, however, high energy variations of discrete period ($T \approx 100$ s) are not practicable. In the 'transient method' (TEM) the computation of S_3 is reliable only if there is a great distance between source dipole and measuring dipole, but a limiting situation is soon reached depending upon the high generator power needed and the noise due to natural earth and industrial currents. However, considerable success has been recently claimed with long offset transient electromagnetic (LOTEM) depth sounding under such conditions. The apparent disadvantages of these methods are supposed to be minimized by using the always available passive

source field of natural electromagnetic variations by magnetotelluric sounding. Due to the simplified assumption of plane wave, the interpretation technique of the MT method is sound in comparison to 'frequency sounding', TEM and LOTEM, however, the exact nature of the source is still a problem.

Such a realistic situation, as discussed above, where conducting porous sediments are existing underneath a thick cover of basaltic lava flow (known as Deccan trap in India) exists in Saurashtra Peninsula, Western India. In seismic survey for the exploration of hydrocarbon, the sequence of sediments posed serious problems as the penetration of seismic energy in these rocks is difficult. Other problems such as the presence of high velocity noise and strong scattering of energy make the situation even more difficult. Sing et al. (1988) conducted a number of deep electric soundings employing Schlumberger and equatorial arrays and they obtained a different subsurface picture as compared to seismic results. Verma (1988) presented a theoretical study on the possibility of delineating sub-trappean sediments using a deep penetration time-domain EM method. Sri Niwas et al. (1988) presented a study on the inversion of field D.C. and synthetic MT resistivity data from the Saurashtra region. Recently Rao and Shastri (1991) reported LOTEM experiments carried out in Saurashtra Peninsula by taking 9 profiles using 13 transmitters and 361 receiver stations covering the entire area with the objective to delineate Mesozoic sediments between the top thick basaltic layer and the bottom mostly crystalline basement rock. The results are not yet available.

It is evident that the interpretation techniques for geoelectric exploration of thick sedimentary sections are only sound for direct current and magnetotelluric methods. The aim of the present paper is to analyse critically and quantitatively the resolving power of these two methods by singular value analysis (SVA) rather than qualitatively, in exploring the sedimentary sequences.

2. FORWARD AND INVERSE MODELLING

Any form of the quantitative interpretation requires in the first place a physical-mathematical model. The earth is supposed to be a linear system and modelling is presented by a convolution model (forward problem) and the quantitative interpretation (inversion) can be achieved by a deconvolution model (inverse problem). A general anomaly equation used in geophysics can be written as Fredholm type integral equation

$$f_i(r) = \int_0^1 G_i(r,\xi)\, m(\xi)\, d\xi \qquad (1)$$

where $G_i(r,\xi)$ is the Frechet kernel of the integral and the two continuous functions $f(r)$ and $m(r)$ represent data and model.

Modelling is presumed to be the process of predicting the results of measurements (predicting data) on the basis of some general principle or model and a set of specific conditions relevant to the problem at hand. As it is well known the direct current modelling is based on the solution of Laplace equation which has been obtained by Stefanescu (1930). The numerical computation would be performed by a technique given by Sri Niwas and Israil (1986, 1987). The magnetotelluric equation is obtained by solving the Helmholtz equation and we use the solution given by Kunetz (1972) and modified by Pek (1980).

The inversion process starts with the data and with a general principle or model and determines estimates of the model parameters. The solution of a type (1) problem is obtained through continuous inverse theory. However, it limits practically the rigour with which continuous functions can be studied. However, simpler discrete models easily permit considerable insight by discrete inverse theory since they rely mainly on the theory of vectors and matrices rather than on the somewhat more complicated theory of continuous functions and operators. Equation (1) is as it has been described for a linear system, however, for a quasi-linear one f_i stands for

the observed data minus that calculated for the initial model and m stands for the profile of a certain quantity in the desired earth model minus that profile for the initial model. Their kernel $G_i(r)$ is calculated for the initial model. In most inverse problems, f_i-s are simply a table of numerical values, they can be represented by the elements of a vector $\underset{\sim}{f}$ of length N (if i=1, N is the number of data points). Similarly, the model parameters can be represented as the elements of a vector m of length M (if the number of model parameters is M). Then the relationship between data and model parameters may be written in discrete form, by Eq. (1) as:

$$f_i = \sum_{j=1}^{M} G_{ij} m_j \quad (i=1,\ N;\ j=1,\ M) \tag{2}$$

which is in matrix form

$$\underset{(N\times 1)}{\underset{\sim}{f}} = \underset{(N\times M)}{\underset{\approx}{G}}\ \underset{(M\times 1)}{\underset{\sim}{m}} \tag{3}$$

where matrix G is given by $\partial G_i/\partial m_j$ and is called the data kernel in analogy to the theory of integral equations.

The freedom of choosing the coordinate system is of greatest importance in both physics and mathematics. Thus, if the coordinate system of (3) is transformed to the principal axes system in which the eigenvectors (u_i) of $\underset{\approx}{G}\underset{\approx}{G}^t$ i.e. the column of the orthogonal matrix $\underset{\approx}{U}$ and the eigenvectors (V_i) of $\underset{\approx}{G}^t\underset{\approx}{G}$ i.e. the column of the orthogonal matrix $\underset{\approx}{V}$ define the principal axes. This gives considerable insight of the physics of the problem. This means that both data and model can be represented by the generalized vectors $\underset{\sim}{f}^1$ and $\underset{\sim}{m}^1$ and the matrix $\underset{\approx}{G}$ by a diagonal matrix Λ i.e.

$$\underset{\sim}{f} = \underset{\approx}{U}\underset{\sim}{f}^1 \ ; \quad \underset{\sim}{f}^1 = \underset{\approx}{U}^t\underset{\sim}{f} \tag{4}$$

$$\underset{\sim}{m} = \underset{\approx}{V}\underset{\sim}{m}^1 \ ; \quad \underset{\sim}{m}^1 = \underset{\approx}{V}^t\underset{\sim}{m} \tag{5}$$

$$\underset{\sim}{f}^{1} = \underset{\approx}{\Lambda}\,\underset{\sim}{m}^{1} \quad ; \quad \underset{\sim}{m}^{1} = \underset{\approx}{\Lambda}^{-1}\underset{\sim}{f} \tag{6}$$

where Λ is a diagonal matrix having elements λ_i, eigenvalues of $\underset{\approx}{G}^{t}\underset{\approx}{G}$ or $\underset{\approx}{G}\underset{\approx}{G}^{t}$, placed on the diagonal in decreasing order of magnitude. If all the parameters are linearly independent then M non-vanishing eigenvalues are expected and if it is an overdetermined system ($N \gtrsim M$) then (N-M) eigenvalues would be equal to zero. Equation (4) shows that the data eigenvectors (u_i) are a combination of the original data and Eq. (5) shows that the parameter eigenvectors (v_i) represent a combination of the original parameters that may be uniquely determined.

Using the singular value decomposition (SVD) theorem (Lanczos 1961) we can write Eq. (3) and its inverse solutions through (4), (5) and (6) as

$$\underset{\sim}{f} = \underset{\approx}{U}\,\underset{\approx}{\Lambda}\,\underset{\approx}{V}_{t}\ m \tag{7}$$

and

$$\hat{\underset{\sim}{m}} = \underset{\approx}{V}\,\underset{\approx}{\Lambda}^{-1}\underset{\approx}{U}^{t}\ \underset{\sim}{f} \tag{8}$$

$\hat{\underset{\sim}{m}}$ is some estimate of $\underset{\sim}{m}$ as in case of an ill posed problem ($N \lesssim M$) it is only possible to get some estimate of the possible solution and not the solution itself. The answer to the question how close $\hat{\underset{\sim}{m}}$ is to $\hat{\underset{\sim}{m}}$ is obtained through

$$\hat{\underset{\sim}{m}} = \underset{\approx}{V}\underset{\approx}{V}^{t}\ m. \tag{9}$$

Thus any particular solution is a weighted average of the true solution with weights given by row vectors of matrix $\underset{\approx}{V}\underset{\approx}{V}^{t} = R$, known as resolution matrix.

However, as due to linear dependence of some model parameters, the orthogonality of the system is destroyed and the matrix G is nearly singular, one is bound to get 'vanishing eigenvalues'. If the data are erroneous, then the relative values of the eigenvalues become important in visualising data error propagation in the solution. Equation (8) clearly demonstrates that if the data points, along which the data eigenvectors are principally directed, contain error α_i and if

the associated eigenvalue is small, the error in parameter β_i will be greatly magnified in the direction of the associated parameter eigenvector which results in an unstable solution and could result in a very large $\|\beta\|^2$. In such a situation the reliability of the solution becomes suspect. One measure of reliability of the solution is its co-variance matrix $<\beta\beta^t>$ which becomes large when λ_i is extremely small. The measure to distinguish between 'vanishing eigenvalues' and 'non-vanishing eigenvalues' is the condition number (η_i) given by

$$\frac{|\alpha|}{|f|} \leq \|\Lambda\| \cdot \|\Lambda^{-1}\| \frac{|\beta|}{|m|} \tag{10}$$

and is numerically equal to

$$\eta_i = \lambda_1/\lambda_i \ . \tag{11}$$

If the condition number is large the solution becomes unstable in the sense that the relative error in the solution may considerably exceed that of the parameter itself. The condition number is also specified by $\|\Lambda\| \cdot \|\Lambda^{-1}\|$ and if it is close to 1 then the problem is 'well conditioned' and if much larger than one, the problem is 'ill-conditioned'.

The ridge regression inversion (Marquardt 1963) if some compromise value is used for the Marquardt constant (k) obtains a stable solution which approximately minimizes the prediction error and also the solution length. Any 'vanishing eigenvalues' of $\underset{\approx}{G}^t\underset{\approx}{G}$ are increased in the ridge regression by a factor k^2. Increasing the size of all the eigenvalues results in a significant decrease of the mean of the squared solution length $(\underset{\sim}{\hat{m}}-\underset{\sim}{m})$ and variance of the estimated solution. The ridge--regression estimate of $\underset{\sim}{m}$ is

$$\underset{\sim}{\hat{m}}* = \underset{\approx}{V} \ (\Lambda^2+k^2\underset{\approx}{I})^{-1} \ \Lambda \ \underset{\approx}{U}^t \ \underset{\sim}{f} \ . \tag{12}$$

Here the choice of the value of k^2 is extremely important because of its physical significance: the contributions of eigenvectors with eigenvalues smaller than k^2 are automatically

supressed thereby the solution gets degraded in both model and data spaces.

3. NUMERICAL EXPERIMENTS OF FIELD DATA (D.C.) AND SYNTHETIC DATA (MT)

For studying the relative resolving power of the D.C. and MT data we have considered the most typical model as given by Porstendorfer (1975) and mentioned in Section 1.

We would call it Model 1. Computed apparent resistivities for the model using both D.C. and MT source fields are given in Fig. 1. It is interesting to note that when the conducting third layer is supposed to be removed the nature of the anomaly is quite distinguishable in the MT case whereas there is hardly any change in D.C. source anomaly. The further interpretation of the D.C. apparent resistivity is not useful as it cannot give any information beyond the second resistive layer. However, the quantitative interpretation of the MT sounding

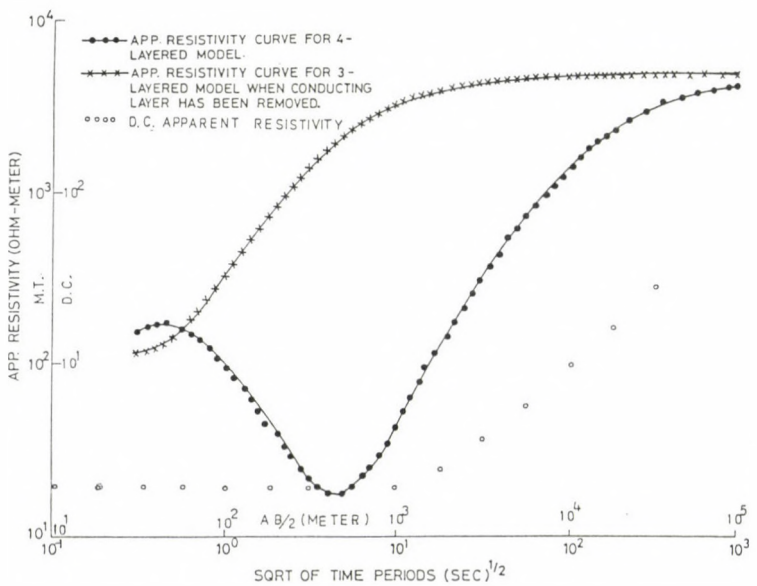


Fig. 1. Forward MT and D.C. apparent resistivity response over Model 1

curve is performed using two powerful inversion techniques, i.e. Ridge-regression Inversion (Inman 1975) and Backus-Gilbert Inversion (Oldenberg 1979). It is apparent to note that the Backus-Gilbert results (Fig. 2) are superior to the Ridge regression solution. Figure 3 presents the condition number plot. The condition number curves are quickly rising with a high gradient, showing the 'ill conditioned' nature of the problem. To quantify the resolving power of the D.C. and MT resistivity data one has to refer to the Singular Value Analysis (SVA) of matrix G which is given in Fig. 4 for Model 1, for both D.C. and MT cases. The physical parameter (i.e. resistivity) governs the physics of the problem more than the geometrical parameters (thicknesses) of the model which are clearly brought in quantitative terms by SVA. The first two largest eigenvalues (λ_1 = 426.3, λ_2 = 7.426) in the MT case have associated eigenvector points almost entirely in the ϱ_3 and ϱ_1 directions, respectively, indicating that these two parameters are very strongly defined in the MT data. And in the

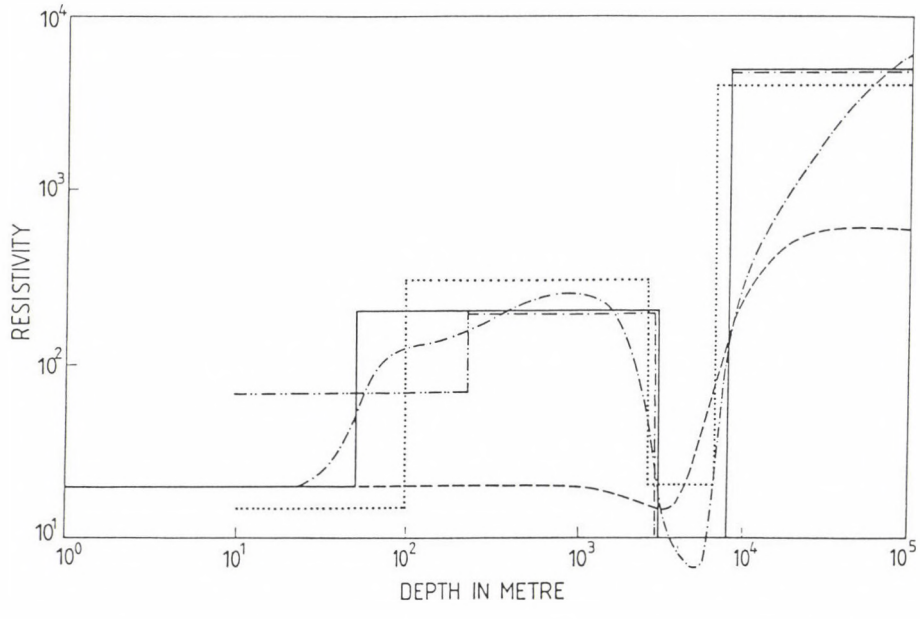


Fig. 2. Interpreted models for synthetic MT response for Model 1

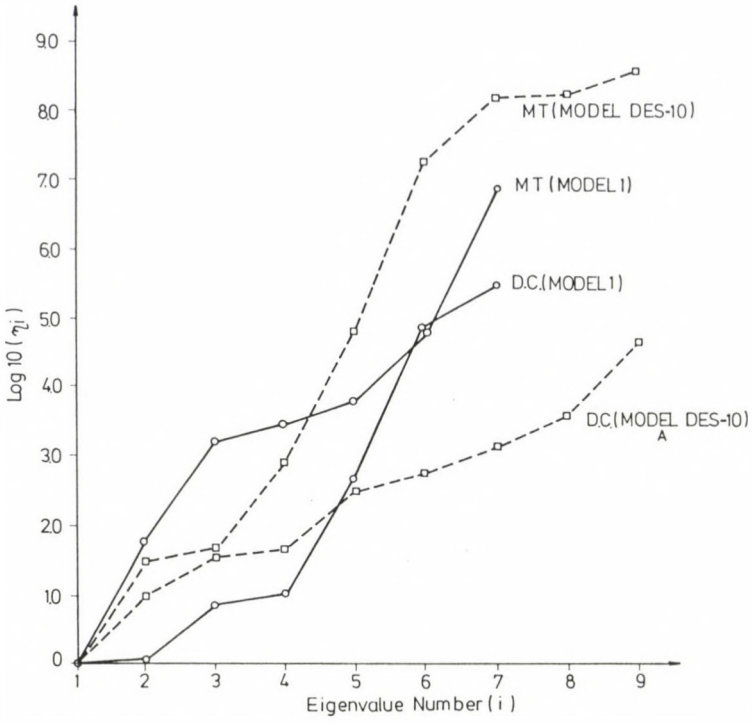


Fig. 3. Condition number plot for Model 1 (D.C. and MT) and for model DES-10A (D.C. and MT)

D.C. case the first two largest eigenvalues (λ_1 = 6.921 and λ_2 = 6.320) have their associated eigenvector points entirely in the directions of ϱ_2 and ϱ_4, respectively. Thus, these two parameters are strongly defined by the D.C. data. Considering the fact that the parameters associated with the largest eigenvalues are best resolved and less prone to errors in the data, it can be seen in quantitative term that the MT method provides better resolution for conducting layers and the D.C. method provides better resolution for resistive layers. However, in the present case, the MT response is computed for periods greater than 0.1 s, as such a top surface layer cannot be resolved at this period range and for this purpose still lower periods are needed. At the same time ϱ_4 in D.C. case may not be obtained in field observation, because of the screening effect of the second layer.

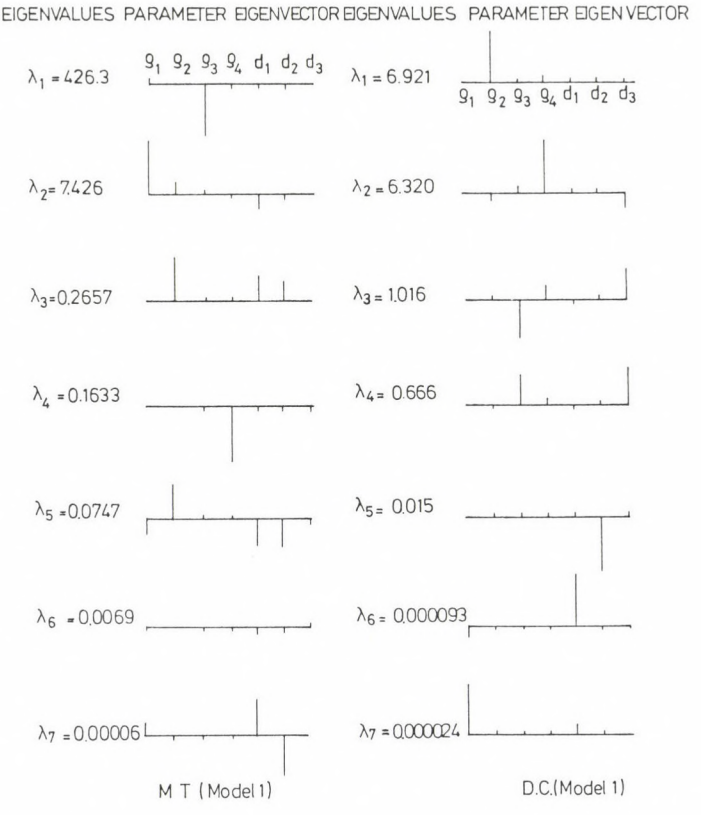


Fig. 4. Singular Value Analysis of matrix G for Model 1 (D.C. and MT)

Let us move on to the problem of geophysical exploration in Saurashtra Peninsula, Western India, where the expected geological sequences are weathered deccan-trap, compact deccan-trap, Mesozoic sediments followed by basement. Because of thick volcanic rocks (deccan-traps), the conventional seismic survey was not successful. Singh et al. (1988) conducted a number of deep electrical soundings in an attempt to provide a representative subsurface lithological configuration. Since the resistivity of the traps is not very high (200-500 ohmm), D.C. methods may succeed to provide meaningful results. We have taken a few sounding curves from Singh et al. (1988) for the purpose of the present study and present them without mentioning their specific location. We have interpreted these

soundings by Ridge-Regression Inversion having a starting value of the Marquardt constant equal to 1 and with a variation rate of 1/3 in subsequent iterations. The interpreted results are given by Model A. Singh et al. (1988) interpreted these soundings by a trial and error method given by Das and Verma (1977) and the result is reproduced by Model B. We have arrived to yet another interpretation of the data by Ridge-Regression Inversion, taking Model B as initial guess. This interpretation is shown as Model C. It is significant to note that although Models A, B and C are different, their forward response matches well with the observed apparent resistivity data within reasonable limits (rms error $< .1$ ohmm). This shows that the D.C. resistivity data are more prone to the principle of equivalence. One such sounding and its interpretation is presented in Fig. 5 for DES-10. These realistic models (Models A, B and C) based on Deep Electrical Sounding data are used for computing MT forward responses in apparent resistivity and phase variation for periods from 0.1 to 20,000 secs and are

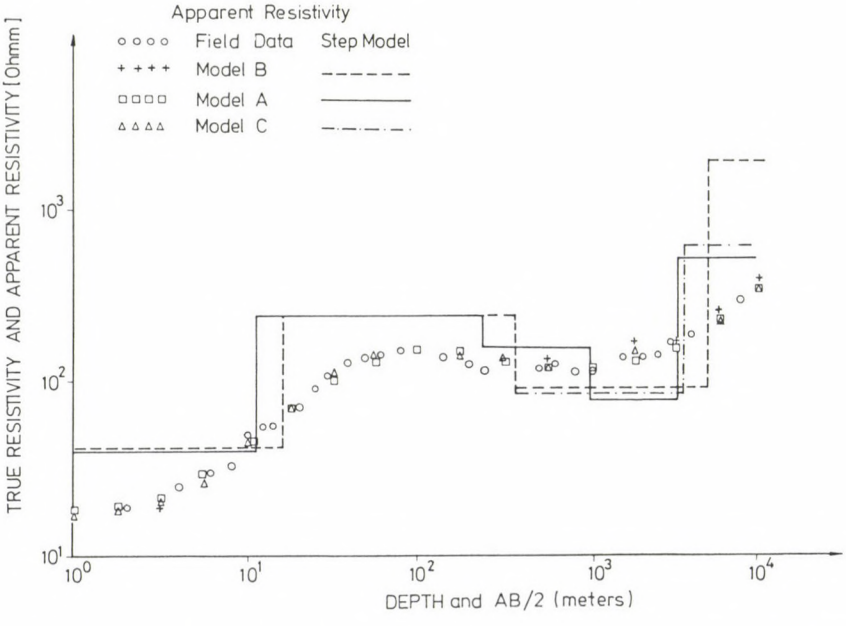


Fig. 5. D.C. apparent resistivity curves and various interpreted model for field curve DES-10

given in Fig. 6. The MT responses for all the three models are distinct proving that the equivalence is less severe in MT in comparison to the D.C. method. Singular value decomposition of the matrix G for Model DES-10A has been carried our for D.C. and MT cases and the condition number plot is included in Fig. 4. The Singular Value Analysis is given in Fig. 7. In Model A the initial guess is for a five layer sequence and fourth layer (ϱ_4, d_4) represents the conducting Mesozoic sediments. The largest eigenvalue (λ_1=128.7) in the MT case shows that the associated eigenvector strongly points in ϱ_4-direction (resistivity of the conducting layer) whereas the largest eigenvalue in the D.C. case (λ_1=44.56) has an associated eigenvector which is directed chiefly in d_4 direction. But the next largest eigenvalue (λ_2=4.29) has an eigenvector uniquely directed in ϱ_2 (resistivity of resistive deccan-trap) direction. This also supports the earlier conclusion that an MT survey in necessary for the resolution of conducting layers and

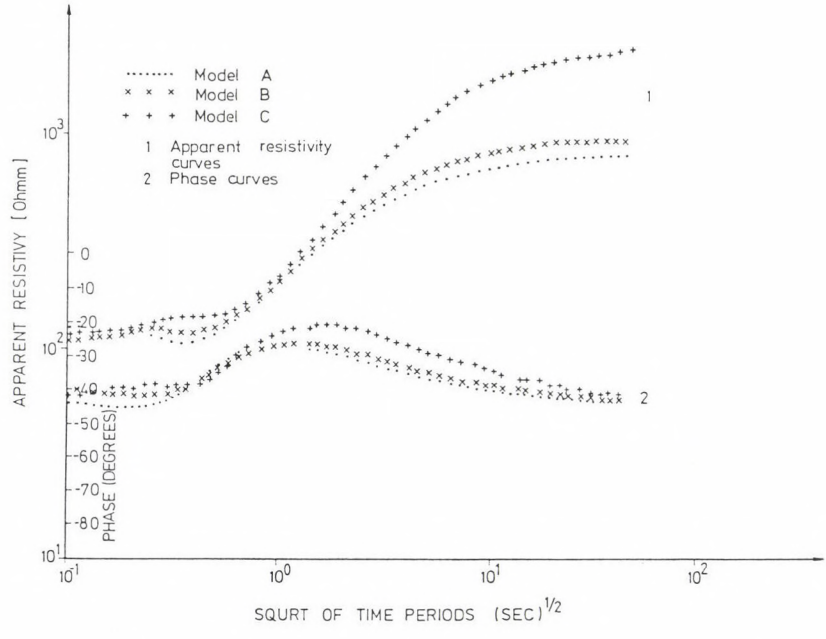


Fig. 6. Forward MT (apparent resistivity and phase) response for Model A, B and C of Fig. 3

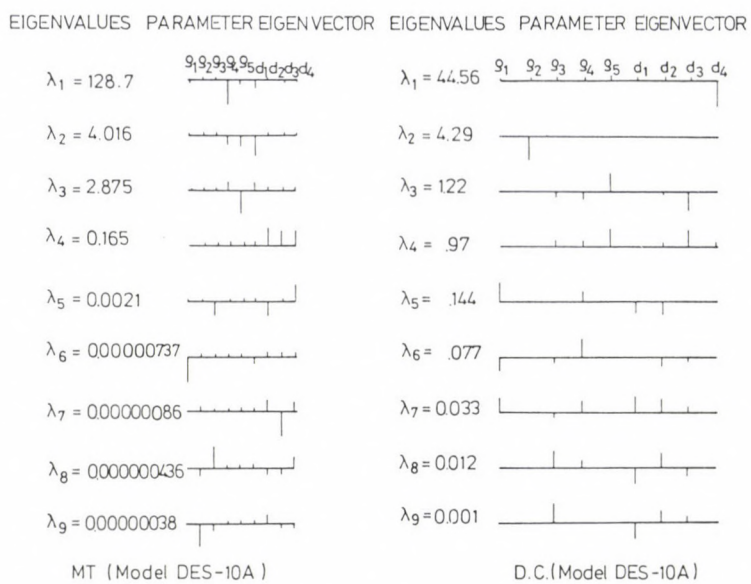


Fig. 7. Singular Value Analysis of matrix G for model DES-10A (D.C. and MT)

D.C. survey in essential for the resolution of resistive layers. It is important to note that other eigenvectors associated with eigenvalues smaller than the largest ones (λ_1) in case of MT and λ_1 and λ_2 in case of D.C.) are more complicated and do not point in one direction.

However, the linear combinations of parameters associated with extreme eigenvalues (higher condition number) are more prone to the error propagation and their determination would be inaccurate. Figure 8 presents the D.C. field sounding curve DES-9 and the three interpreted Models A, B and C. The computed forward response for all the three models resembles well to the field data, however, the MT forward response (apparent resistivity and phase) differs significantly from each other in case of Models A, B and C (Fig. 9).

4. CONCLUSION

From the above discussion it can be concluded that D.C. and MT resistivity soundings are effective tools with well

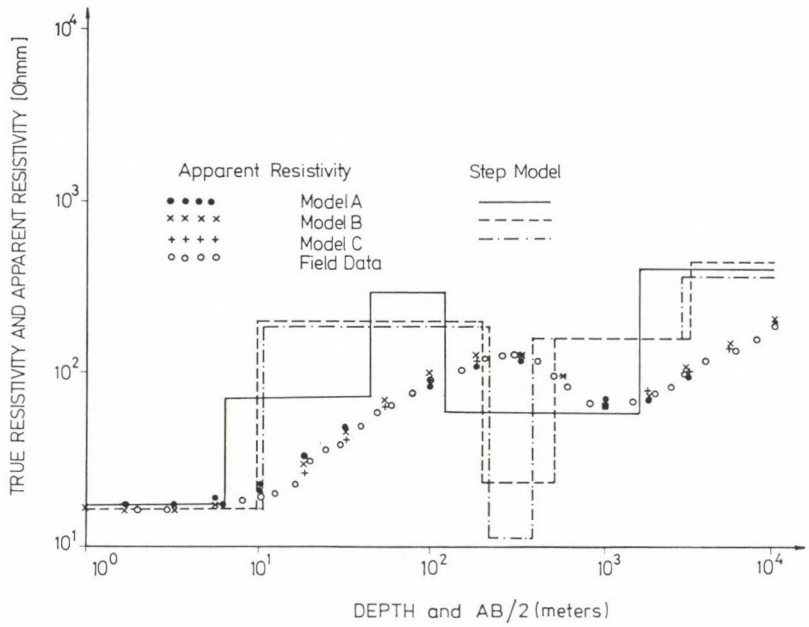


Fig. 8. D.C. apparent resistivity curves and various interpreted models for field curve DES-9

developed forward and inverse modelling technique for the exploration of thick sedimentary basins. MT sounding theory is developed for H-polarization and E-polarization. In H-polarization the longitudinal conductance of the sediments is correctly determined from the ascending branch of the sounding curve and with the E-polarization the depth of the well conducting layer is correctly determined from the descending branch of the sounding curve. However, with D.C. resistivity, the longitudinal conductance of the sediments is correctly determined from the ascending branch of the sounding curve but no such facility is available to determine correctly the depth of the conducting layer from the descending branch of the sounding curve. At the same time if there are conducting layers below a virtually insulating layer like salt (in Model 1) the D.C. resistivity method is incapable of giving information due to screening effect of the salt layer. But the MT method is not affected by such situations. Through extensive numerical

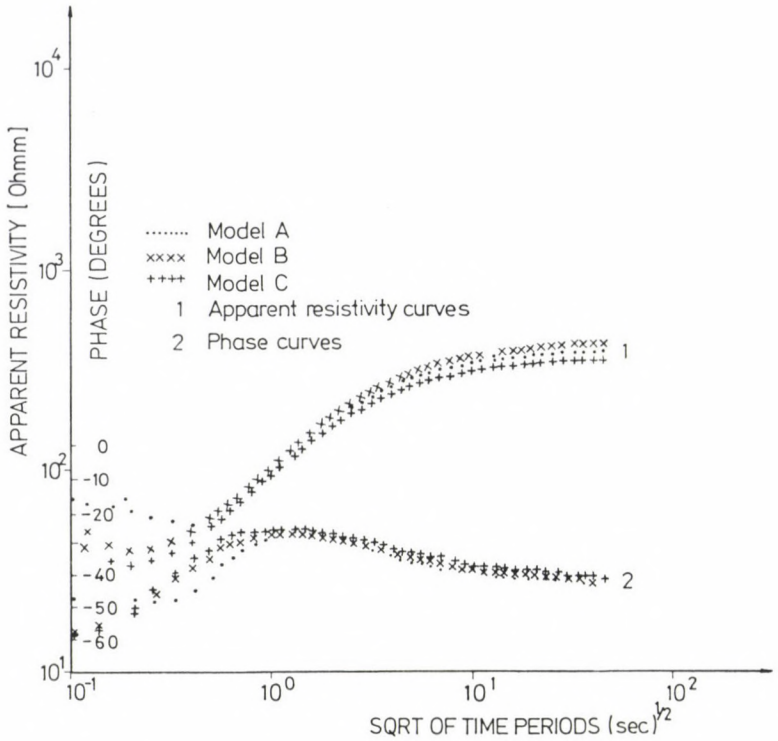


Fig. 9. Forward MT (apparent resistivity and phase) response for Models A, B and C of Fig. 8

experiments and by considering realistic models (from Saurashtra) it has been established that the equivalence principle in D.C. resistivity is strongly valid whereas in MT method it is partially valid. On the basis of the singular value analysis of the matrix operator G, it has been quantitatively shown that MT data are necessarily required to resolve conducting horizons whereas D.C. data are essentially required to resolve resistive horizons. One advantage of the D.C. data is that if the screening layer (like salt) is replaced by an other resistive layer like volcanic rocks (as deccan-trap in India) then it is the only method to provide a starting model for inverting MT data. If a full interpretation of all the sections of thick sedimentary horizons is required there is no scope of ignoring D.C. resistivity sounding even if

the cost of survey is higher as complementary data to the inductive methods like MT sounding or if the theory is perfected, to the LOTEM depth soundings.

5. ACKNOWLEDGEMENT

The authors are thankful to Prof. Vinod Gaur for providing encouragement to investigate these problems which are aimed at an understanding of the fundamental nature of the problem.

REFERENCES

Das U C, Verma S K 1977: Unpublished NGRI Report

Inman J R 1975: Geophysics, 40, 798-817.

Kunetz G 1972: Geophysics, 37, 1005-1021.

Lanczos C 1962: Linear Differential Operators. D. Van Nostrand

Marquardt D W 1963: Jr. SIAM, 11, 431-441.

Oldenberg D W 1979: Geophysics, 44, 1228-1244.

Pek J 1980: Studia Geoph. et Geol., 24, 39-50.

Porstendorfer G 1975: Principles of magnetotelluric prospecting. Borntraeger, Berlin-West - Stuttgart

Rao E P. Shastri N L 1991: Proc. International Seminar on Deep Electromagnetic Exploration. Kharagpur, India

Singh S B, Sunder A, Gupta M L, Pandey B P, Das D 1988: Proc. International Workshop on Deep Electromagnetic Exploration. Hyderabad, India

Sri Niwas, Israil M 1986: Geophysics, 51, 1594-1602.

Sri Niwas, Israil M 1987: Geophysics, 52, 1412-1417.

Sri Niwas, Verma S K, Kumar P, Manglik A 1988: Proc. International Workshop on Deep Electromagnetic Exploration. Hyderabad, India

Stefanescu S S 1930: Le Journal de Physique et le Radium, 7, Series 1

Verma S K 1988: Proc. International Workshop on Deep Electromagnetic Exploration. Hyderabad, India

Acta Geod. Geoph. Mont. Hung., Vol. 26 (1-4), pp. 453-468 (1991)

DISTORTION OF ERROR CHARACTERISTICS BY OUTLIERS

F Steiner

Geophysical Department, Miskolc University, H-3515 Miskolc,
Egyetemváros, Hungary

[Manuscript received April 30, 1991]

It is a justified demand in all branches of geosciences that the uncertainty of the data uninfluenced by foreign (outlier-generating) effects, i.e. the errors of reliable data should be expressed numerically. The author analyses in this paper seven kinds of error characteristics using a two-parameter, simple model for the outliers; the relative distortions due to the outliers reached often several orders of magnitude in these parameters. As the standard deviation (called in short "scatter") is extremely sensitive against outliers, it is advisable to use other kinds of error parameters. If for any special reason just the scatter (or the variance) is to be determined with the greatest possible accuracy (e.g. if a variogram is to be constructed), even in such cases some other kind of error, e.g. the empirical value of the probable error q (i.e., the sample semi-interquartile range) should be determined in the first step, and this value is then to be multiplied with the value σ/q which quotient is well known for any kind of probability distribution.

Keywords: asymptotic scatter; dihesion; error characteristics; model of outliers; most frequent value; outliers; resistance; scatter; semi-interquartile range

1. DEFINITION OF SOME ERROR CHARACTERISTICS

Several kinds of error definitions can be used for the characterization of the uncertainty of data, resulting from the essence of the phenomenon studied and from the normal conditions of the measurements. In order to make the formulas simpler, in the following the probability density function $f(x)$ of the deviations x from the true value shall be supposed to have symmetrical distribution with respect to the origin (general formulas are given e.g. by Steiner 1988 and 1990). The distribution function corresponding to $f(x)$ is $F(x)$, its inverse will be denoted by F^{-1}.

Primary error characteristics being independent of statistical algorithms are the so-called semi-interquantile ranges. Among them the semi-interquartile range

$$q = \frac{1}{2}\,[F^{-1}(3/4)-F^{-1}(1/4)] \qquad (1)$$

and the semi-intersextile range

$$Q = \frac{1}{2}\,[F^{-1}(5/6)-F^{-1}(1/6)] \qquad (2)$$

are the most often used ones; q is also called probable error (Bessel proposed its use at the beginning of the last century).

The most usual application of the scatter (standard deviation) defined as

$$\sigma = \sqrt{\int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx} \qquad (3)$$

is due to the fact that statistical algorithms are mostly based on the least squares principle, i.e., the L_2-norm of the deviations is made a minimum; σ is the minimum norm of the x-s (e.g. Tarantola 1987).

A large table following p. 334 in Steiner (1990) gives the definition of the so-called P-norm for symmetrical unimodal distributions. The location of the minimum for this norm is found by the algorithms called "most frequent value procedures". Other names are: MFV-method (Dresen et al. 1989), M-fitting (Silva and Cutrim 1989); in German texts, HWM does occur, too. The value ε (which plays a central role in the expression of the P-norm) fulfils the equation

$$\varepsilon^2 = 3\,\frac{\displaystyle\int_{-\infty}^{\infty} \frac{x^2}{[\varepsilon^2+x^2]^2}\,f(x)dx}{\displaystyle\int_{-\infty}^{\infty} \frac{1}{[\varepsilon^2+x^2]^2}\,f(x)dx} \qquad (4)$$

in an iterative sense; this value ε belonging to the minimum place of the P-norm is called dihesion and it is a measure of

the compactness of the data with the density function f(x), - thus it gives a possibility for the characterization of the errors in the data. The uncertainty of the most frequent value M (the location of the minimum of the P-norm) is characterized by the asymptotic scatter (denoted here by A) similarly to σ which latter is the asymptotic scatter of the expected value E. The formula for A is somewhat complicated: if the function s(x) is defined by the dihesion ε as

$$s(x) = \frac{4\varepsilon}{4\varepsilon^2+x^2} \tag{5}$$

then

$$A = 2\varepsilon \frac{\sqrt{n_1-n_2}}{2n_2-n_1}, \tag{6}$$

where

$$n_1 = \int_{-\infty}^{\infty} s(x) \cdot f(x) \, dx \tag{6a}$$

and

$$n_2 = \int_{-\infty}^{\infty} s^2(x) \cdot f(x) \, dx . \tag{6b}$$

A is in many cases well approximated by the following much simpler formula (which has an interesting similarity to Eq. 3):

$$A = \sqrt{\frac{9}{5} \int_{-\infty}^{\infty} s(x) \cdot x^2 \cdot f(x) \, dx} . \tag{7}$$

Finally it should be mentioned that the uncertainty of the data system itself can be measured with the scale parameter appearing in the algorithm called by Andrews et al. (1972) "CML-estimation" which fulfils the equation:

$$S_{CML} = \frac{\int_{-\infty}^{\infty} \frac{x^2}{S_{CML}^2+x^2} f(x) \, dx}{\int_{-\infty}^{\infty} \frac{1}{S_{CML}^2+x^2} f(x) \, dx} ; \tag{8}$$

is also to be understood in an iterative sense.

It is to be remarked here that different error characteristics have different numerical values (as they have different meanings), nevertheless, for a given type of f(x), the transformation factors among them can be determined, consequently if any kind of (finite) error characteristics is determined, any other can be computed from it. If one has a so-called geostatistical distribution, which is expected to occur according to Eqs (6-14) in Steiner (1990) with the highest probability being characterized within the supermodel $f_a(x)$ by the type parameter a=5, then the scatter σ can be calculated as 1.91 q, 1.29 Q or 1.47 ε (see formulas 1-13 and 1-14 (Steiner 1990) for the definition of $f_a(x)$, the numerical values of the error characteristics are to be found in Tables 1.1 and 3.4). It is therefore necessary to be informed about the distorting effect of outliers resulting from effects due to foreign causes which surpass significantly the statistical variability of the data due to inherent changes in the studied process.

2. THE MODEL OF OUTLIERS USED

The appearance of outliers may be very different (even rhapsodic). If statistical algorithms are to be tested about their sensitivity or insensitivity (resistance) against outliers, then the distribution of the outliers is to be modelled, too. Andrews et al. (1972) (see pp. 102-103) supposed that from the sample with n elements n-j are standard Gauss-distributed random numbers, j are outliers with the values 100; 200; ...; j x 100.

Such a "one-sided" distribution of the outliers (only positive deviations) is justified in this case by the fact that the breakdown bounds of algorithms for the determination of the location parameter are studied. The distortion of error characteristics may be caused by outliers of symmetrical distribution, too, therefore (for illustrative purposes, and for simplicity, resp.) the outliers are supposed to have a

symmetrical distribution in the present model in the range $(-m \leq x < -1)$ and $(1 < x \leq m)$ but the distribution is else uniform as at Andrews et al. The range being undistorted by outliers is $(-1 \leq x \leq 1)$; several points of view speak in favour of a uniform distribution of the "good" data, too, whitin this interval. (The most important such point is the lack of flanks, thus values with $|x| > 1$ should be really considered as outliers, if the undisturbed data have this advantageous distribution.)

Naturally all statistical algorithms - disregarding some theoretically unfounded ad hoc procedures - handle the outliers (being present in a ratio r) together with the undisturbed data, and therefore it is logical to characterize the present model of outliers by a single density function f(x):

$$f(x) = \begin{cases} \frac{1-r}{2} & \text{if } |x| \leq 1 \\ c = \frac{r}{2(m-1)} & \text{if } 1 < |x| \leq m \\ 0 & \text{else} \end{cases} \qquad (9)$$

(see Fig. 1). The integrals in the different error formulas can be easier studied if f(x) from Eq. (9) is written as

$$f(x) = f_1(x) + f_2(x) \qquad (10)$$

where

$$f_1(x) = \begin{cases} d = \frac{1}{2}\left(1 - r - \frac{r}{m-1}\right), & \text{if } -1 \leq x \leq 1 \\ 0 & \text{else} \end{cases} \qquad (10a)$$

further

$$f_2(x) = \begin{cases} c = \frac{r}{2(m-1)}, & \text{if } -m \leq x \leq m \\ 0 & \text{else} \end{cases} \qquad (10b)$$

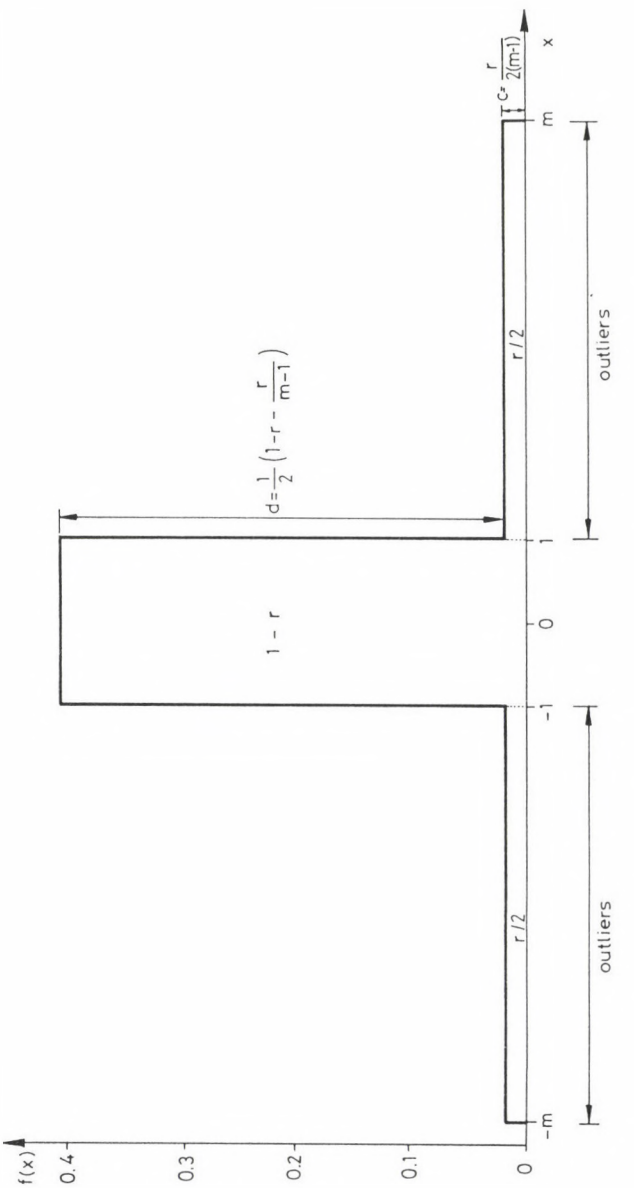

f(x)
0.4
0.3
0.2
0.1
0
-m
-1
0
1
m
x
r/2
r/2
1 - r
$d=\frac{1}{2}\left(1-r-\frac{r}{m-1}\right)$
$c=\frac{r}{2(m-1)}$
outliers
outliers

Fig. 1.

(see again Fig. 1).

According to these, any integral appearing in Eqs (3, 4, 6, 7) and (8) can be given in an explicit form. The next section gives the formulas for the different error characteristics in form of a list.

3. FORMULAS FOR THE ERROR CHARACTERISTICS FOR THE PRESENT MODEL OF OUTLIERS

The introduction of certain simplifying notations helps a better arrangement of the program. Before listing them, the definition of notations is repeated which have been already used in the previous sections, too:

r: rate of the outliers

m: maximum absolute value of the outliers

ε: dihesion (see the general formula 4, further Eq. (14) concerning the present model outliers).

$$c = \frac{r}{2(m-1)}$$

$$d = \frac{1}{2}\left(1 - r - \frac{r}{m-1}\right)$$

$$q1 = \frac{1}{1+\varepsilon^2} \qquad q2 = \frac{1}{1+(2\varepsilon)^2}$$

$$qm1 = \frac{m}{m^2+\varepsilon^2} \qquad qm2 = \frac{m}{m^2+(2\varepsilon)^2}$$

$$t1 = \frac{1}{\varepsilon}\ \text{arc tan}\ \frac{1}{\varepsilon} \qquad t2 = \frac{1}{2\varepsilon}\ \text{arc tan}\ \frac{1}{2\varepsilon}$$

$$tm1 = \frac{1}{\varepsilon}\ \text{arc tan}\ \frac{m}{\varepsilon} \qquad tm2 = \frac{1}{2\varepsilon}\ \text{arc tan}\ \frac{m}{2\varepsilon}$$

$$tS = \text{arc tan}\ \frac{1}{S_{CML}}$$

$$tmS = \text{arc tan}\ \frac{m}{S_{CML}}\ .$$

The formulas for the error characteristics are with these notations the following:

$$q = \begin{cases} \dfrac{1}{2(1-r)} , & \text{if } r \leq \dfrac{1}{2} \\ m - \dfrac{m-1}{2r} , & \text{if } r > \dfrac{1}{2} \end{cases} ; \qquad (11)$$

$$Q = \begin{cases} \dfrac{2}{3(1-r)} , & \text{if } r \leq \dfrac{1}{3} \\ m - \dfrac{m-1}{3r} , & \text{if } r > \dfrac{1}{3} \end{cases} ; \qquad (12)$$

$$б = \frac{1}{\sqrt{3}} \sqrt{1 + r\left(\frac{m^3-1}{m-1} - \right)1} \quad ; \qquad (13)$$

$$\varepsilon^2 = 3\varepsilon^2 \cdot \frac{d \cdot [t1-q1]+c \cdot [tm1-qm1]}{d \cdot [t1+q1]+c \cdot [tm1+qm1]} \quad \text{in an iterative sense ;} \qquad (14)$$

$$A = 2\varepsilon \cdot \frac{\sqrt{n1-n2}}{2 \cdot n2-n1} , \qquad (15)$$

where

$$n1 = 8\varepsilon^2 [d \cdot t2+c \cdot tm2] \qquad (15a)$$

and

$$n2 = 4\varepsilon^2 [d \cdot (q2+t2)+c \cdot (qm2+tm2)] ; \qquad (15b)$$

$$\bar{A} = 14.4 \cdot [d \cdot (1-4\varepsilon^2 t2)+c \cdot (m-4\varepsilon^2 \cdot tm2)] ; \qquad (16)$$

$$S^2_{CML} = S_{CML} \cdot \frac{d \cdot [1-S_{CML} \cdot tS]+c \quad [m-S_{CML} \cdot tmS]}{d \cdot tS+c \cdot tmS} . \qquad (17)$$

Equation (17) is also to be understood in an iterative sense similarly to Eq. (14). (Equations 14-17 can be easily derived on ground of the Eqs 4-8, i.e., of the general definitions, applying known mathematical relations given e.g. by Gröbner and Hofreiter (1961), see Eqs 15a,b on page 15 and Eq. 3c on p. 18.)

4. NUMERICAL VALUES OF THE DISTORTIONS OF THE ERROR CHARACTERISTICS CAUSED BY OUTLIERS

In a case without outliers (r=0) the numerical values of the error characteristics are the following for a uniform distribution in the range (-1, +1):

$q = 0.50000$
$Q = 0.66667$
$\sigma = 0.57735$
$\varepsilon = 0.71852$
$A = 0.67433$
$S_{CML} = 0.42898$.

(As it has been already mentioned, if any of the error characteristics is known, all others can be computed for an arbitrary type of distribution, e.g. 1.155-times the q-value determined from a sample yields the value of σ for a uniform distribution.)

In cases $r > 0$ (with any value of m) Eqs (11-17) give the error characteristics for a distribution distorted by outliers. If the surplus due to the outliers is divided by the correct value referring to the sterile case, a surplus in percents is obtained, and these values enable a comparison of the distorting effect of outliers on different error characteristics. (The distorted values of A are referred immediately to the correct value of A, as A was ab ovo defined as an approximation to A.)

Equations (3, 4, 6, 7) and (8) show that the greatest distortion due to outliers occurs at the scatter, σ , as the greatest multiplier, x^2 is obtained here on the integrand for the great values of $|x|$ in the function $f(x)$ being greater than 0 due to the outliers. Figure 2 shows in a double logarithmic scale the relative values of the distortions in σ due to outliers for different rates of the outliers (r) vs. the maximum outlier value. (The computations have been carried out both in r and in m with equal logarithmic steps, thus the ratio

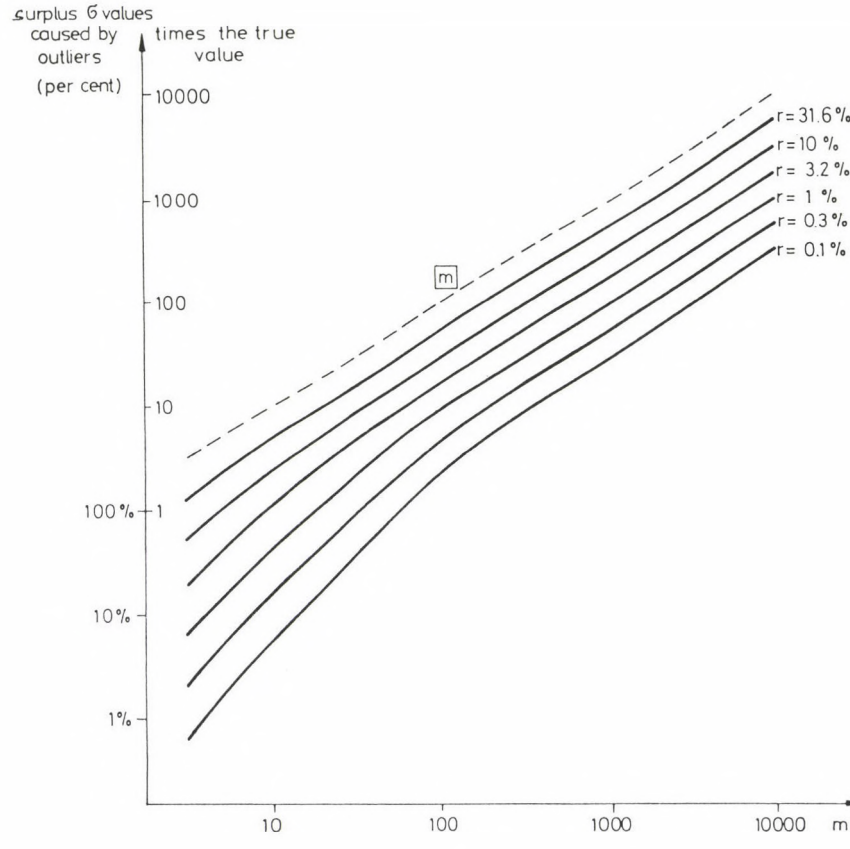


Fig. 2.

of two values following each other is always $\sqrt{10}$.)

It is characteristic that in the case of m=100 the distortion is more than twice of the real value even if the rate of the outliers (r) is as small as one per mille. That means that the distorted $\tilde{\sigma}$ -value is more than three times greater than that of the sample without outliers. The choice m=100 is similar to the procedure used by Andrews et al.: if the sample has an extent of n=1000, the outlier j=1 has just the value 100; Andrews et al's original method leads generally to even greater m values.

The unexpectedly high values shown in Fig. 2 are connected

(as the formula of the scatter, Eq. (3) shows) with the fact that the greatest weights are obtained for the farthest part of the flanks of f(x), thus the values increase strongly with increasing m. (If the minimum value of the L_∞-norm would be used for the characterization of the errors - the definition is given e.g. by Tarantola (1987) -, then it would be exactly equal with m. These values are represented in Fig. 2 by the dotted line and denoted by m, and this gives a curve fitted to the curve set σ, above the curve with the parameter r=31.6 percent.) The distorted values of σ are mostly (with the exception of the smallest values of r-s and m-s) no adequate error characteristics of the "good" data.

One could suppose that other error characteristics get similarly useless at most pairs of the parameters (r,m) represented in Fig. 2, even if less than valid for the scatter σ. Figure 3 shows, nevertheless, that the greatest distortion of the dihesion in the range of the parameters (r,m) studied in Fig. 2 is only slightly more than 30 percent, and these distortions decrease strongly with the increase of the maximum value of the outliers, m. This decrease is the consequence of the fact that in the integrals of Eq. (4) the multiplier of f(x) decreases strongly with increasing |x|, therefore the values of the integrals are only influenced by outliers lying relatively near the greatest data densities (i.e., which are not too far from the sterile data), and according to Eq. (9), in the present model of the outliers the probability density f(x) decreases in these parts of the flanks (similarly to everywhere else) as r/2(m-1).

The multiplier of f(x) in the integrals of Eq. (8), defining iteratively S_{CML}, does not decrease as strongly, as necessary to give decreasing distortions for increasing m-s (with constant r). In Fig. 4 the relative distortions of S_{CML} are plotted against m for r=31.6 percent. These values increase with increasing m, but this increase is very slight, moreover it reaches quickly saturation. Concerning the level of distortion: the saturation value is nearly 100 percent, in contrast to the values for the dihesion for the same value of

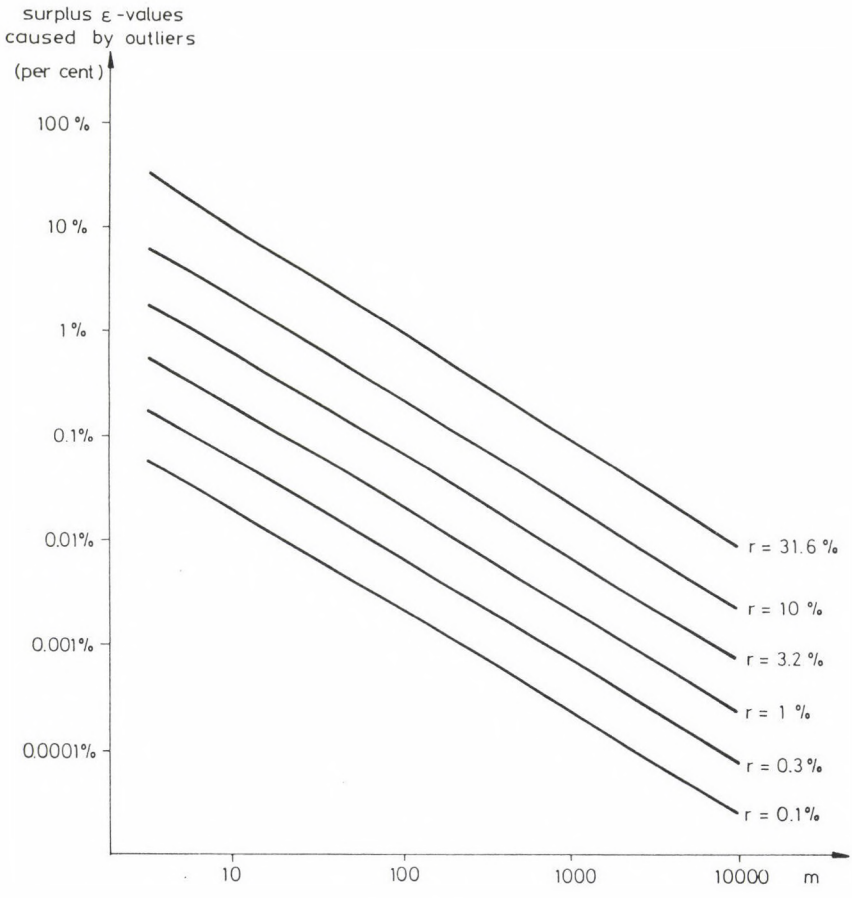


Fig. 3.

r, which are smaller by orders of magnitude. (The sensitivity of S_{CML} against outliers and the resistance of the dihesion ε, respectively, have been shown already in a concrete case of a data system by Csernyák and Steiner 1985.) A comparison of the curve in Fig. 3 for ε is enabled by replotting it in Fig. 4; the corresponding ε-curve is also shown here from Fig. 2 to see the effect of an increasing m in all the three possible cases.

The approximation of a constant value of the distortion with icreasing m (for a given value of r) is characteristic not

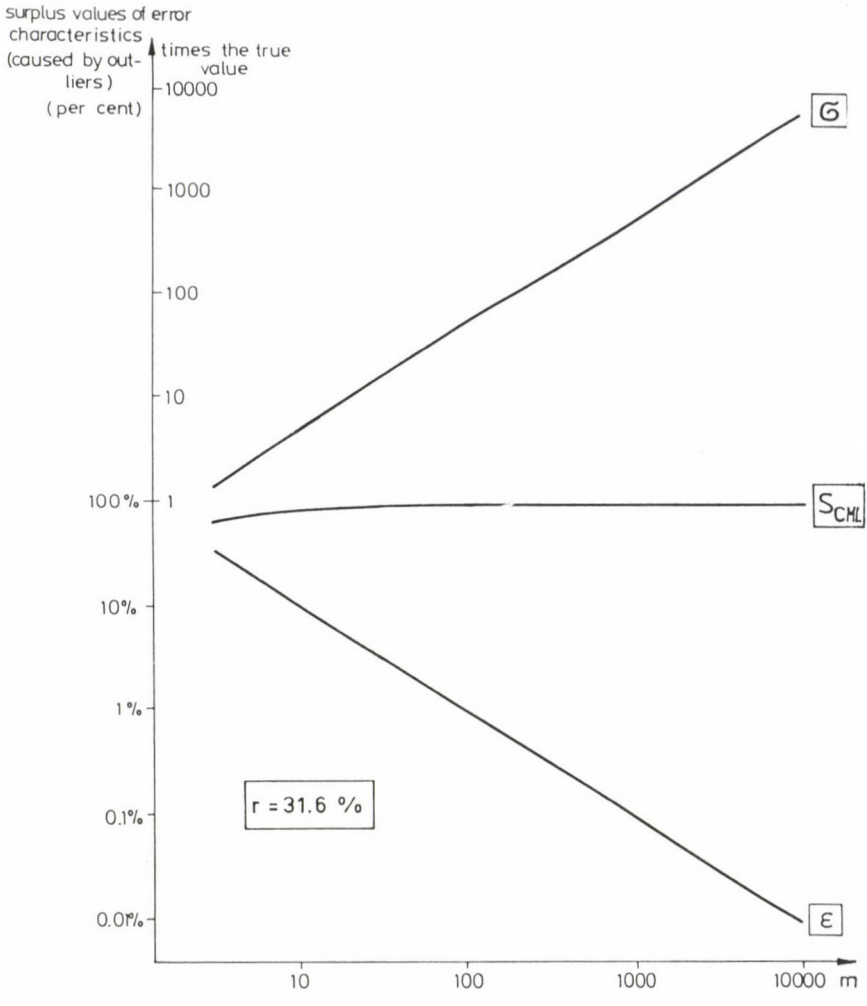


Fig. 4.

only for S_{CML}, but also for the asymptotic scatter of the most frequent value, $\bar{A}$ (and of A approximating it). As the similarity concerns also the fact that the values at m=10 approximate already rather closely the values at $m \to \infty$, these latter constant values are given in Fig. 5 for S_{CML}, A and $\bar{A}$ vs. r. (The distortion of the asymptotic scatter A is even at a 10 percent rate of the outliers only about 5 percent,

independently of the value of m, i.e. even in the case of large values of m.)

Finally we consider shortly the error characteristics q and Q, the distortions of which are independent of m, if $r \leqslant 1/3$ (see Eqs. 11 and 12). Elementary computations show that the relative distortions are to be computed in this range in both cases $r/(1-r)$, and these values are plotted with thick line in Fig. 6 (notation "q;Q"). For some values of m (thin

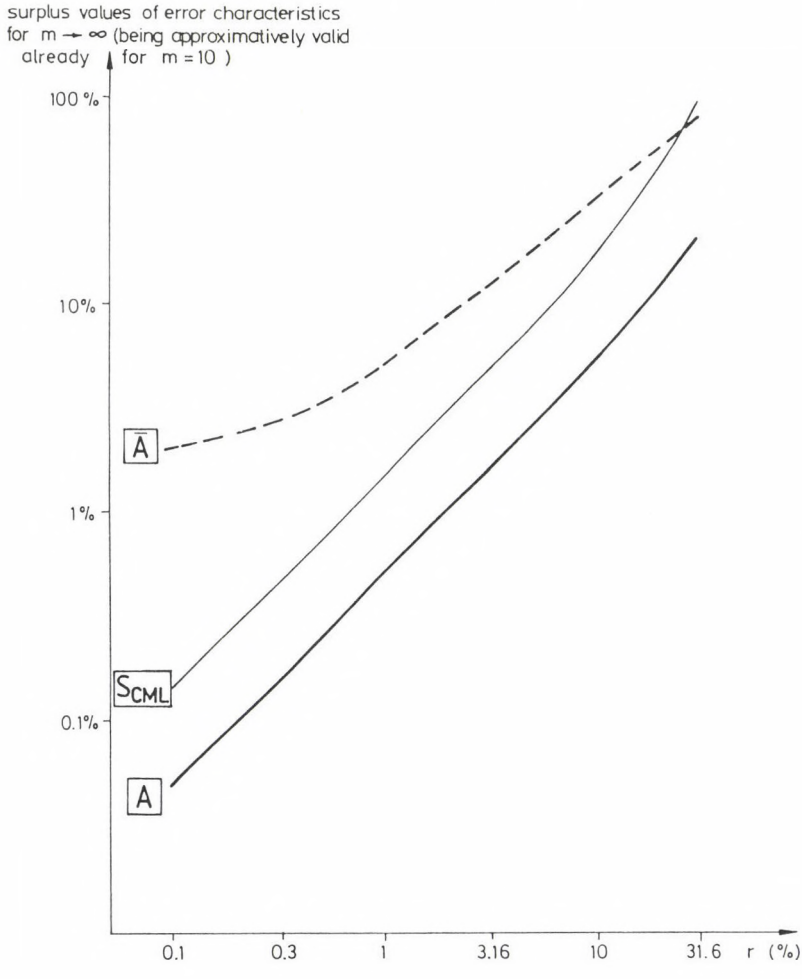


Fig. 5.

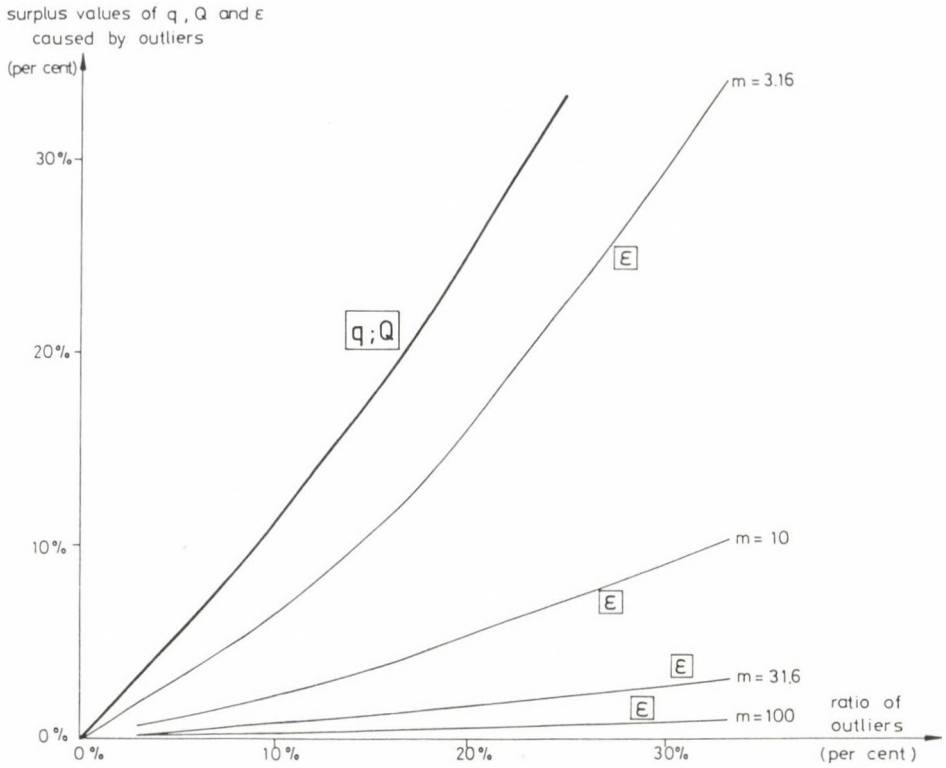


Fig. 6.

lines) the relative distortions of the ε -s are also shown in this figure. The case m=3.16 is just at the boundary to model realistic outliers using the present model of the outliers, and the distortion of the dihesion is even in these cases less than the distortion of q and Q. (And it has been shown previously that with icreasing m the distortions of ε decrease quickly, see Fig. 3.) If the model of the outliers is modified so that all the outliers appear in the range (1,m) (with a rate of r), say, with uniform distribution, the range of the m-independence and together with it, the range of the small values for q and Q will be halved (e.g. for Q, $0 \leqslant r \leqslant 1/6$, instead of $r_{max}=1/3$).

In contrast the asymmetry of the outliers causes no significant distortion in ε . This can be seen from Eq. (4) defining the dihesion where x appears only squared. This is, however, insufficient, as Eq. (4) is given here for simplicity at a zero value of the location parameter T=M, and therefore, in the general case, in the denominator of the fractions is $(x-M)^2$ instead of x^2 what means that if the asymmetric outliers would significantly modify the most frequent value M, then this would cause a secondary, but not neglectable distortion of ε . Such a distortion of the location parameter is, however, impossible even at the greatest m-values studied in the present paper, if the standard option of the computation of the most frequent value is used (see the Figs in Steiner 1988, pp. 67 and 68, especially the ordinates for k=2, i.e. roughly for M'). Thus the statements in the present paper about the outlier-insensitivity of the dihesion (i.e. about the resistance) are not connected with the symmetry of the present model of the outliers.

REFERENCES

Andrews D F, Bickel P J, Hampel F R, Huber P J, Rogers W, Tukey J W 1972: Robust Estimates of Location. Princeton University Press, Princeton, N.J.

Csernyák L, Steiner F 1985: Publ. of the Univ of Miskolc, Series A, Mining, 40, 183-224.

Dresen L, Csókás J, Dobróka M, Gyulai Á, Ormos T 1989: On the influence of model and measurement errors in the inversion of the in-mine measured geoelectric and seismic data. Proceedings of the 34th International Geophysical Symposium. MGE, Budapest

Gröbner W, Hofreiter N 1961: Integraltafel I. Unbestimmte Integrale. Springer-Verlag, Wien

Kerékfy P 1978: Alkalmazott Matematikai Lapok, 4, 327-357.

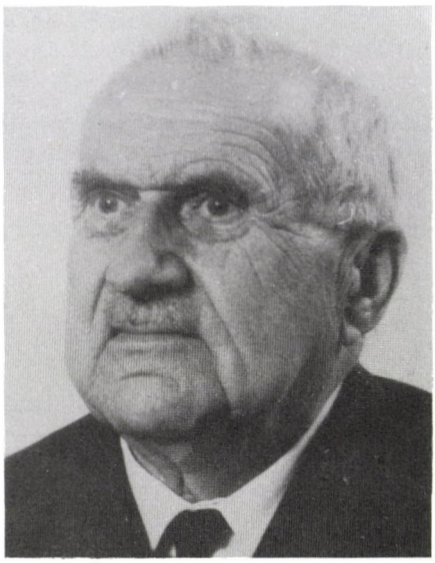
Silva J B C, Cutrim A O 1989: Geoexploration, 26, 1-31.

Steiner F 1988: Geophysical Transactions, 34, No. 2-3, 139-260.

Steiner F 1990: Foundation of Geostatistics (in Hungarian). Tankönyvkiadó, Budapest

Tarantola A 1987: Inverse Problem Theory. Elsevier, Amsterdam

Professor Hazay nonagenarian



A crowd of his former students, former and present co-operators, colleagues and all the Hungarian geoscientists congratulate Professor Hazay at his 90th birthday.

His extremely effective carreer is closely connected partly with the Hungarian National Land Survey, partly with the Technical University of Budapest.

After graduating in civil engineering at the Technical University Budapest in 1922, he entered state service as a land surveyor at the National Land Survey 1923. During his successful activity for about 30 years he gained a lot of experience in all phases of land surveying works: triangulation, precise levelling, geodetic astronomy etc.

As a result of his outstanding abilities he received soon managing tasks: from 1932 in the works of the geodetic control net and survey of Budapest, from 1938 in the national land survey, and from 1942 on, he became the head of the Hungarian National Land Survey.

His abilities manifested themselves already in his youth both in the scientific research work and the professional education.

His first scientific papers have appeared in 1930. Up to now he has published about 100 papers and 6 scientific books, among them "Adjusting Calculations in Surveying" also in English. He concentrated his research activity on the current theoretical and practical problems of geodesy, surveying and mapping. Several of his ideas have became topics of international interest.

The following steps of his scientific carreer have been: he was awarded the doctor degree by the Technical University of Budapest in 1934; at the same university he was awarded the Professor title in 1941; the Hungarian Academy of Sciences elected him its corresponding member in 1967, and full member in 1976. As president of the Geodetic Commission of the

Hungarian Academy of Sciences, he played a leading role in the Hungarian geodetic science for a long time.

Professor Hazay's connection with the professional education started in 1929, when he began to teach Adjustment and Geodetic Astronomy at special courses. Later when the special education of civil engineers in geodesy, surveying and mapping was started in this country, he has put all his force into the service of professional education.

From 1951 he educated many generations of geodesists at the Technical University in Sopron, from 1959 at the Technical University of Budapest as the Professor of Mathematical Cartography, Adjustment, National Survey etc. and Head of Department of Geodesy.

His excellent lecture notes and textbooks were published in several editions. He was Dean of the Faculty for Civil Engineering in the period 1960-1964, and subsequently Rector of the University until 1967. Meanwhile several of his students and assistants have became professors, outstanding engineers or scientists.

Professor Hazay is holder of the State Kossuth Prize, of three governmental and two ministerial medals, and of the Honorary Medals of Hungarian geodesists. The Technical University Budapest awarded him the title "doctor honoris causa" and the Honorary Medal of the University.

His deep knowledge, undinimished energy and outstanding human character are examples to be followed for all of us. We express our best wishes at his 90th birthday in the hope that he will be further active for a long time in a good health.

The Editorial Board

Publications of Professor István Hazay in the period 1981-1990

Addenda to the list of his publications related to the period 1930-1980, published in the Acta Geod. Geoph. Mont. Hung., 16 (1981), 7-10.

97. Quick zone-to-zone transformation in the Gauss-Krüger projection. Acta Geod. Geoph. Mont. Hung., 18 (1983), 71-82.
98. The scale of the map (A térkép méretaránya). Geodézia és Kartográfia, 1986.
99. The importance of projections in mapping (A vetületek szerepe a térképészetben). Geodézia és Kartográfia, 1988.
100. Remembrance of Gyula Papp (Emlékezés Papp Gyulára). Geodézia és Kartográfia, 1988.

BOOK REVIEWS

L FIALOVSZKY ed.-in-Chief: Survey Instruments and their Operational Principles. Akadémiai Kiadó, Budapest, 1991. 750 pp, 646 figs, 26 tables

The book is the revised and enlarged version of the handbook in Hungarian language "Geodéziai műszerek, edited also by L Fialovszky (Műszaki Könyvkiadó, Budapest, 1979).

This new edition gives a detailed theoretical background to understand the basic principles of construction and operation of the instruments necessary for the most general tasks of surveying, and presents many types of these instruments, together with their descriptions. As compared to the earlier edition the scope is greatly expanded, especially in respect of the new types of modern instruments, such as electronic distance meters and tacheometers, automatic equipments for mapping etc.

The book contains three parts and 14 chapters in them, with the following topics:

Part I. Opto-Mechanical Surveying Instruments

1. Fundamentals and elements of optics; levels
2. Instruments for measuring angles
3. Instruments for measuring elevation
4. Precision instruments for horizontal and vertical setting out
5. Instruments and equipment for distance measurement
6. Instruments for tacheometry

Part II. Electronic Surveying Instruments

7. Elements of electronic instruments
8. Distance measuring instruments using the principle of interference
9. Electronic distance measuring instruments
10. Electronic theodolites and tacheometers

Part III. Office Instruments and Equipment for Processing

11. Grid square and coordinate plotting devices
12. Equipment for automatic mapping and plotting
13. Automatic equipment for map production
14. Instruments to measure areas.

In the Appendix one can find a list of some important firms producing surveying instruments, a selected bibliography and a detailed name and subject index.

The book was published (as Number 62) in the series "Developments in Geotechnical Engineering", as a joint edition with Elsevier, Amsterdam-Oxford-New York-Tokyo.

This well compiled, clearly written uptodate work will surely be helpful for undergraduate and graduate students and teachers of surveying, for practising builders, architects and technicians and for everybody who is interested in any branch of surveying.

At the first pages of the book there is a short biography of the Editor-in-Chief with the comment under his photo: "Professor Fialovszky died two weeks after the final version of this manuscript was received by the publishers".

Thus, when taking this book - his very last work - in hand, it reminds us of him, the untiring scientist, too.

L Miskolczi

K R KOCH: Bayesian Inference with Geodetic Applications. Springer Verlag, Berlin, Heidelberg, New York, London, 1990. 198 pp, 4 figs, 6 tables, Appendix

The Bayesian approach provides a unified approach to Analytic Statistics. There is no need to distinguish the different hypotheses, tests and estimations. Bayesian analysis uses the same basis. In the Bayesian phylosophy the elements in a statistical model can be assigned a probabilistic structure. Parameters appearing in the model can be assigned probability distributions. The Bayesian statistics is a very useful tool to analyze different kinds of data.

This book gives a very nice presentation of the Bayesian Inference with geodetic applications.

After a short introduction, Chapter 2 deals with the basic concepts of Bayesian inference. It gives the derivation of the Bayes theorem summarizes the prior information for the unknown parameters. It treats the point estimation, confidence regions, the hypothesis testing and predictive analysis. Finally the numerical techniques for the statistical inference are presented.

Chapter 3 is focussed on problems of models and special applications. Different models are introduced. Thus the linear models with noninformative and informative priors for the unknown parameters, nonlinear models, mixed models and linear models with unknown variance and covariance components. It deals with the problem of classification solved by applying the Bayes rule. The so-called M or robust estimates are also derived by Bayesian inference. Finally, the reconstruction of digital images is discussed.

The appendix summarizes the different kinds of distributions.

A basic knowledge of statistical methods is needed to read the book. The book is very clearly written and helps the scientists who are interested in evaluating and analysing of data.

J Somogyi

ISPRS Commission II/VII International Workshop Proceedings. Advances in Spatial Information Extraction and Analysis for Remote Sensing. American Society for Photogrammetry and Remote Sensing, 1990. 112 pp, 42 figs, 5 tables

The book contains the papers of ISPRS Commission II/VII International Workshop held at the University of Maine Orono (13-17 January 1990).

The five invited papers and the ten presented papers are grouped into the following subjects: "Data Fusion in Remote Sensing and GIS. Knowledge-Based Systems. Processing on Small Computers". The topics of the papers include the following: Integration of GIS with remote sensing; Standards for data transfer and exchange; Man-machine interaction; Database design and data management systems; Computer vision and image understanding techniques in remote sensing; Large shared data bases; Algorithms for information extraction from image data of high spatial and spectral resolution.

J Somogyi

Protecting Natural Resources with Remote Sensing. Proceedings of the Third Forest Service Remote Sensing Applications Conference - April 9-13, 1990. American Society for Photogrammetry and Remote Sensing. 488 pp, 75 figs, 43 tables

This issue contains papers from the Third Forest Service Remote Sensing Application Conference held at the University of Arizona and the Doubletree Inn Tucson, Arizona.

Most of the papers deal with the use and application of remote sensing technology in some way or another related to the monitoring and protection of natural resources.

The proceedings give a very informative summary about the topic.

J Somogyi

Global Natural Resource Monitoring and Assessments: Preparing for the 21st Century Proceedings of the International Conference and Workshop - September 24-30, 1989. Three volumes. Published by American Society for Photogrammetry and Remote Sensing. 1990. 1495 pp

The three volumes collect the papers presented to the International Conference and Workshop of Global Natural Resource Monitoring and Assessments: Preparing for the 21st Century. The objectives of the conference were: To follow up on the recommendations from past meetings, and improve the integration and coordination of national and international inventory and monitoring systems and their associated data bases to support environmental assessments in the 1990s and beyond.

Volume I included 65 papers of seven sessions with the following subjects: Global monitoring, the planning perspective, the environmental perspective and the social/economic perspective. Importance of global data in monitoring forests and desertification, range resources, soil and water resources. Review of current global assessment efforts. Linking monitoring and assessment to planning and action for sustainable development. Social aspects of global monitoring. International range inventory and monitoring. Water resources. Monitoring environmental health.

Volume II includes 67 papers of seven sessions. The main topics are: Resource inventories. Technical forecasts for the future. Global coordination efforts. Resource management planning. Policy aspects of monitoring systems. Ecological approaches to resource inventory and monitoring.

Volume III includes 69 papers of eight session with the following themes: Future satellite sensing capabilities. Resource measurements and modeling. Data bases and geographic information systems. Remote sensing opportunities. Remote sensing case studies. Achieving the future. Blueprint of the future, recommendations and proposals. Responses and closing.

The lectures which are collected in these three volumes deal with the questions of the moment and can be useful for differently skilled experts.

J Somogyi

W J RIPPLE ed.: Fundamentals of Geographic Information Systems: A Compendium. Published by American Society for Photogrammetry and Remote Sensing, 1989. 248 pp, 84 figs, 9 tables

One meets with an increasing number of resource management and

environmental problems. These include global climate change, acid deposition stemming from the burning of fossil fuels, the disposal of toxic wastes, increasing soil erosion and decreasing water quality, population growth etc. Geographic information systems have the potential to help for solutions to such problems.

The book is divided into eight sections, the first section deals with the definition of GIS, the history of the technology, and requirements and principles for GIS implementation. The second section focuses on entry of spatial digital data into a GIS. Section three gives a review on the principles of logic as they apply to geographic data analysis, fundamental GIS operations, and digital terrain models. The fourth section deals with GIS data quality issues and error assessment. Part five treats the recent developments and future trends in GIS. Examples of GIS applications can be found in sectin six. One can find in section seven information on how to select and evaluate a GIS. Section eight collects the GIS literature.

J Somogyi

Twelfth Biennial Workshop on Color Aerial Photography and Videography in the Plant Sciences and Related Fields. American Society for Photogrammetry and Remote Sensing, 1990. 303 pp, 84 figs, 52 tables

The book contains the papers from the Workshop on Color Aerial Photography and Videography in the Plant Sciences and Related Fields held in Sparks, Nevada, 23-26, May 1989. Most of the papers deal with the color and color infrared photography for vegetation and eleven of the papers discuss the use of videography for the plant sciences.

J Somogyi

Image Processing '89, Sparks, Nevada, 23 May 1989, American Society for Photogrammetry and Remote Sensing, 258 pp, 121 figs, 46 tables

The book contains the papers of the Image Processing 89' meeting which was held in conjunction with the twelfth Biennial Workshop on Color Aerial Photography and Videography in the Plant Sciences and Related Fields.

The papers are divided into the following topics:

I. Imaging Spectroscopy with six papers
II. Geology with five papers
III. Forestry with seven papers
IV. Image Processing Techniques with six papers
V. Land and Water Resources with five papers.

The Proceedings give a very informative summary about the remote sensing problems in USA.

J Somogyi

J R WILLIAMSON and M H BRILL: Dimensional Analysis Through Perspective. A Reference Manual American Society for Photogrammetry and Remote Sensing, Kendal/Hunt Publishing Company, Iowa, 1990. 242 pp, 52 figs, 1 table

This book gives a summary of perspective which is one of the first geometric properties of photography. According to the authors it is the

purpose of this manual to provide the photogrammetric analyst with step-by-step procedures without dealing how the procedures have been derived. They have used specific aspects of single-photograph perspective to present a fresh review of techniques commonly used in photogrammetric analysis of one-, two- and three-point perspective photography. The use of single-image perspective techniques are not widely treated in photogrammetric literature. This work gives a help to the photogrammetric analyst using single images.

This manual is intended for the professional photogrammetrist.

J Somogyi

Earth Observation Systems, Legal Considerations for the '90s. Published by American Society for Photogrammetry and Remote Sensing and American Bar Association, 1990. 253 pp

This book contains the product of two in-depth workshops. The first part of this volume "Satellites and Remote Sensing" is a collection of papers presented at the 1987 Workshops and gives a nice presentation on the technology of earth remote sensing, the current and potential applications of this technology, and the current state of the industry and government activity in this area.

The second part "Current Developments in Remote Sensing Law and Practice" consists of transcripts of the actual proceedings of the 1989 Workshop. This Workshop was held as a roundtable discussion among the participants.

The proceedings give a very useful help to experts who a working in this field.

J Somogyi

P VYSKOCIL, C REIGBER, P A CROSS eds: Global and Regional Geodynamics. International Association of Geodesy Symposia, Symposium No. 101. Series editor: I I Mueller. Springer-Verlag, New York, 1989, 349 pp., 193 figs

This volume contains the proceedings of a symposium held within the General Meeting of the International Association of Geodesy (IAG) to commemorate its 125th Anniversary (Edinburgh, Scotland, August 3-5, 1989).

This book is a direct result of the improved cooperation between the IAG Commission VII on Recent Crustal Movements (CRMC) and IAG Commission VIII of Space Techniques for Geodesy and Geodynamics (CSTG) with the aim to extend information on recent crustal movements on global, regional and local scales as well as to improve the ideas, techniques and theoretical studies for a deeper understanding of the geodynamic phenomena.

The CRMC has reached a great progress with the terrestrial instrumentation on local and partly regional scales incorporating advanced estimation and filtering methods to monitor the displacement and strain accumulation fields of special interest.

The CSTG is developing and coordinating the links between the various groups engaged in the field of space techniques as well as elaborating and proposing projects of international cooperations. It has proved the capability of the relative expensive techniques of Satellite Laser Ranging (SLR) and Very Long Baseline Interferometry (VLBI) to monitor the recent global plate motions.

The appearance of coordinate differences measured by a relative

unexpensive NAVSTAR Global Positioning System (GPS) among the classical observables was the first signal of the cooperation. Moreover, the GPS and the mobile SLR/VLBI techniques are able to fill the regional gap between the local and global approaches.

The most pregnant proof of the facts above is a large number of high quality papers included in this volume.

In the first five papers SLR/VLBI data collected in the last years for global plate motion investigations are analysed. The space geodetic results are compared to the widely accepted geophysical models.

The instrumentation and modelling of SLR/VLBI observations are presented in four additional papers.

The regional application of SLR/VLBI technique to the Japanese island arc contraction, to the Pacific Basin and to the meeting of the Eurasian, African and Arabic Plates (WEGENER MEDLAS project) are presented in four papers.

The modelling in non-Cartesian frame, the effects of precipitation and groundwater on ground deformation as well as the lithospheric deformation and asthenospheric pressure are discussed in the next three papers, respectively.

Six papers deal with GPS applications from regional to local scales, specified as the Transatlantic Net (TANGO), the Australian Pilot Project for orbit inprovement and local fields of special interest in China, Iceland, Venezuela and Canada.

The horizontal investigations by terrestrial methods are covered by ten papers. Beside local projects in Venezuela, Romania, Canada and Egypt the related design, random distribution and environmental problems are presented in different papers. Some papers investigating the repeated distance measurements and strain observations with holographic interferometry in Japan are also included.

The vertical investigations by terrestrial methods in Japan and Italy, the Kalman Filter technique and the water-tube tiltmeter application in Japan are discussed in four papers, respectively.

The volume is closed by the investigation of the secular change in gravity on the Fennoscandian region.

Summarizing, this book is a very characteristic cross section of the geodynamic investigations by geodetic methods giving an insight not only for geodesists interested in the "strain", but to other geoscientists engaged mainly in the "stress" part of the same phenomena.

L Bányai

H SÜNKEL, T BAKER eds: Sea Surface Topography and the Geoid. International Association of Geodesy Symposia, Symposium No. 104. Series editor: I I Mueller. Springer-Verlag, New York, 1989. 187 pp, 66 figs

This volume contains papers presented August 3-5, 1989 at the Symposium 104 of the General Meeting of the IAG in Edinburgh. It is divided into four parts according to the sessions of the Symposium.

In Part 1 on the Geoid, the published papers reflect the recent practical and theoretical efforts to improve the accuracy of global, regional and local geoid determinations in national and international relation. The available high precision of GPS positioning measurements which are now in cm order in relative sense require similar accuracy of the resolution of gravity field models and geoid computation, so the presented geopotential models and the computational methods aimed at this precision.

It seems me from the papers that the larger and larger number of satellite tracking data, altimeter data and surface gravity data with topographical information and the sophisticated computational methods like collocation and Fast Fourier Transform make this accuracy possible. There are also very impressive and initiative efforts to make geoid determination and levelling automatic by the combination of GPS and Inertial Navigation System.

In Part 2 on Sea Surface Topography, most of the papers deal with the problems of the combination of satellite altimetry data with sea or land gravimetry. The knowledge on sea surface topography and the shape of the geoid on the sea is very important for oceanographers as well as for geodesists since 2/3 of the Earth surface is covered by water. So any geodetic information about the sea surface improves the global picture about the Earth's surface and its gravity field. The computational results show that the combination of altimeter data with gravity data improves the prediction accuracy both of altimeter heights (that is the sea topography) and of gravity field (that is the geoid). By this combination the so called crossover differences have a standard deviation of some decimeters after the adjustment of data.

In Part 3 on The Vertical Datum, several papers treated the problem of the connection of local vertical datum between continents. Since the mean see levels detected by local Tide Gauge stations are not on the same equipotential surface of the Earth's gravity field so the separation should be known for the exact connection. The authors described three main approaches to solve this problem which are based on (1) global gravity field models, local gravity surveys and levelings, (2) the solution of a geodetic boundary value problem, and (3) satellite altimetry. It is clear from the papers that in solving such problems - which serve the geodetic background to solve other problems of the natural sciences - the contribution of space techniques like SLR, GPS, VLBI will increase significantly in the near future.

In the joint-view part 4, a simultaneous determination of a general ocean circulation model and an Earth gravity field model, a sea surface topography model, a geopotential model derived from altimeter, Doppler and laser tracking data is presented in terms of spherical harmonics. Some questions of precise leveling are also discussed here with special attention to the corrections for atmospheric refraction.

In conclusion this proceedings give an impressive picture about the recent state and accuracy of the determination of geoid related quantities and show clear connection points to other Earth's sciences in which the geodetic information is necessary to study global processes. It is also clear that for such tasks the development of the methods of space geodesy is indispensable.

G Papp

Gy BÁRDOSSY and G J J ALEVA: Lateritic bauxites. Akadémia Kiadó, Budapest, 1990. 624 pp, 225 figs, 53 tables, 16 coloured plates

The book of one of the authors (Gy Bárdossy: Karst bauxites, 1982, Akadémiai Kiadó and Elsevier, Budapest) has been highly appreciated by experts dealing with bauxite exploration. Simultaneously, the need has come up that as a continuation of this work the knowledge on lateritic bauxites ought to be summarized, as well. It is a laudable aim that authors prepared this book wthin a relatively short time. The book "Lateritic bauxites"

together witht he "Karst bauxites" can be regarded as a two-volume work and it has to be found in the library of the scientist, teacher and practical expert. The value of the book is reflected also by the fact that the work published jointly by the Akadémiai Kiadó and Elsevier Science Publishers B.V. is the book No. 27 of the well-known series "Developments in Economic Geology".

The nine chapters include all the things that are worth knowing about lateritic bauxites. The large number of references (about 500 items), the Glossary at the end of the book and the exact definitions make this work not only an important source but also a fundamental collection of information for all the experts dealing with bauxite exploration. The work gives indispensable aid for experts to speak a uniform professional language all over the world.

The review of lateritization and of the genetic processes of lateritic bauxites, the classification of bauxite deposits as well as the outline of the world distribution are important introductory aspects to discuss the lateritic bauxites in detail. Discussing the main characteristics of lateritic bauxites in harmony with the external and internal factors the following aspects are dealt with: climate, tectonics and source rocks as external, and geometry of deposits, lithology (structures and textures, colours etc.) geochemistry and mineral compositions of the deposits as the most important internal factors. The review of information concerning the genetics of lateritic bauxites is a summary of these chapters.

Having summarized the main characteristics of lateritic bauxite formation, the lateritic bauxite deposits of the world are described and listed over about 300 pages in more or less details according to the significance of them. In the course of this discussion the lateritic bauxite deposits are assigned to eight provinces: South America, Southeast Brazil, West Africa, Southeast Africa, India, Southeast Asia, West and North Australia, Southeast Australia. The isolated bauxitic areas are grouped at the end of this chapter.

Information on each deposits within the provinces derive from published or partly unpublished data and have been provided by bauxite companies for the specific purpose of inclusion in this book. Authors studied most of the beds in the field.

The authors followed the same sequence in the concise description of the deposits. The main aspects of this descriptions are: name, location, size, bauxite profile, chemical and mineral composition, geomorphology, age, genesis, ownership, reserves, exploration and mining.

The last chapter deals with bauxite as an aluminium ore and it becomes obvious to the reader here how the scientific knowledge serves the economic purposes. In this chapter useful information can be obtained on the methodology of bauxite exploration starting from the field studies through laboratory measurements to the reserve estimation, on the quality control, on the mining, concentration and processing of the ore. The comprehension of data of the world bauxite reserves and production fairly well reflect the situation and dimensions of this branch of industry.

As for myself who have been in contact with bauxite exploration in many aspects, first of all in the fields of mineralogical and geochemical researches, I can safely declare that the geological literature of the world has gained a very important, informative book written with scientific authority.

Gy Pantó

J FÜLÖP: Bevezetés Magyarország geológiájába (Introduction to Hungary's geology). Akadémiai Kiadó, Budapest, 1989. 246 pp, 109 figures and three in colour

This book is the second part of a voluminous work on Hungary's geology. The first part included the history of the exploitation of mineral resources in Hungary, this second part gives the outlines of the geological structure of Hungary. The author extends the discussion in two directions: first, he gives a historical account, i.e. he starts with early ideas and shows how the geologic structure became more and more deeply known both by exploration and by theoretical hypotheses. The second extension is in space: he does not restrict to Hungary, or to the Carpathian Basin within which Hungary is the central part, but he discusses surrounding mountains of the Carpathians towards the Alps and towards the Balkan peninsula.

Geologic and geophysical exploration of the basin and in the surrounding mountain chains yielded ample information about near-surface and deep structures, age and tectonics of the area and enabled plate tectonical reconstructions which are sometimes contradictory, nevertheless they point toward a final solution for the reconstruction of the geologic past of the basin. The historic viewpoint leads to the reconstruction of not only the geologic history, but also to the history of the ideas. Especially the basin itself was the object of different theories (intermediate massif, orogenic origin, heterogeneous composition, mantle diapir and finally the plate tectonic model). As the political history of the basin was nearly as variable as the geologic one, lists of geographic names in different languages (and different times) help the orientation together with an extensive bibliography. It is to be considered if the book would be worth being translated into English as the interest in the Carpathian Basin has increased recently, and a Hungaro-American geological--geophysical cooperation continues to complete our ideas on the development of the basin.

J Verő

W SCHRÖDER ed.: Advances in Geosciences Interdivisional Commission on History of the IAGA. Bremen-Roennebeck, 1990. 358 pp, many figures

This book contains papers presented at the symposia "The history of geomagnetism and aeronomy" and "Problems of uncertainties in geophysical time series" of the Interdivisional Commission on History of the IAGA during the IAGA General Assembly, held in Exeter, August 1989.

The papers presented in this book are concentrated around a few topics being presently of specific interest. The first of these topics is the eleven-years solar cycle, its history, with special emphasis on the Maunder minimum. Legrande and LaGoffe look for connection between the Maunder minimum and climatic changes (no unambiguous correspondence is found), Attolini, Cecchini, Galli and Nanni use tree ring radiocarbon data together with auroral occurrences for the establishment of the solar cycle in historic times, Krakovetsky, Loisha and Popov conclude from Sibirian auroral data that no Maunder minimum can be found there, Schröder brings evidence for the eleven-years cycle before the Maunder minimum, Kopecky relates the auroral occurrence to solar activity, Bumba and Hejna propose a new index of high geomagnetic activity, and somewhat differing from these, Banzon et al. report on the effect of the volcanic activity on Italian floods. Methods for similar studies are presented by Sneyers and by

Gregori. In spite of the deviating data, there is little doubt about the existence of the Maunder minimum, its climatological effect, however, is much less well established. Anyway, one has to be extremely careful with the analysis of historic data, and basic data are as far as possible to be cross-checked from historic, social, economic and thematic points of view.

A second group of papers deals with the early history of sun-earth relations, including geomagnetism (Skinner with he early history of the electrojet, McNoe with the Apia observatory, Steleanu with Edmund Halley, Wiederkehr with Neumayer's plan to establish an observatory in Australia, and Humboldt's role in this project, Legrand with de Mairan's surprisingly modern ideas on the aurora). Hall added to this group a scientometric analysis of nineteenth century literature on geomagnetism. The most interesting points of these accounts is that the parameter time played a different role in that time: the time needed to collect funds for a project lasted nearly as long, as e.g. an expedition itself, or even much longer, and the realisation of the plan could be a tiny part of the whole project. Heider's account on the one and a half centuries of geophysics in Canada, and Schmidlin's report on the Thyssen-gravimeter refer only partly to geomagnetism, nevertheless, they are interesting to read.

It is a pity that two papers by Debarbat and by Brown about the role of women in earth sciences are very short summaries. The last group deals with persons in sun-earth physics of the present century (Baer and Liebowitz on Landsberg, Dobson on Schove, Dieminger on Lange-Hesse and Tiemann on Foerster). They show a wide variety of personalities, beginning from enthusiastic organizers to self-devoting teachers.

The volume is recommended to all who are interested in the history of the sun-earth physics, be it from a historical point of view, be it through the use of historical data.

J Verő

W WITTKE: Rock Mechanics. Theory and Applications with Case Histories. Springer-Verlag, Berlin, Heidelberg, 1990. 1075 pp, 861 figs

This book is the English translation of the German edition extended by reports on particularly difficult rock engineering projects.

The book is based on more than 20 years experience in research, teaching and practicising of the author in the relatively new engineering field of rock mechanics. He focussed on the constitutive laws governing the mechanical and hydraulic behaviour of jointed rock, on their numerical implementation and on the application of this knowledge to rock construction.

Special attention is paid to the computer-oriented methods of analysis corresponding to the current "state of the art" in rock mechanics, namely to the finite element method applied to stability analyses of various types of construction during the design and different construction stages.

This work consists of 28 continuously numbered chapters devided into 5 large logically built up parts.

In part "A" the conceptual models are presented comprising the grain and rock mass fabric, the discontinuities, the stress-strain behaviour of different rock masses, the basis for the consideration of deformation due to swelling and the model of seepage flow through a rock mass.

In part "B" the procedures of the finite element analysis are developed describing the stresses and strains in a rock mass, the stresses and deformations due to swelling, the rock mass wedge stability and the

three-dimensional seepage flow using the homogeneous model. The seepage flow analysis using the discrete model is also presented.

In part "C" the analysis procedures are applied to the different rock engineering constructions as traffic tunnels and adits, caverns, pressure tunnels, foundations of gravity and arch dams as well as slopes.

In part "D" the different investigations, explorations and tests are presented which are required to build up the mechanical characteristics, the loadings, the boundary conditions and geometry of the finite elements system during the design as well as necessary to be performed during the construction stages to check the previously made assumptions. The chapters describe the site observations, the laboratory tests, the shear tests on discontinuities, the tests to determine the deformability and in-situ stresses as well as the groundwater conditions and permeability investigations. Examples of rock mechanic investigations are also included.

In part "E" the case histories of the Hasenberg tunnel in Stuttgart Rapid-Transit Railway, the Wehr and the Estangento-Sallente powerhouse caverns are presented. These studies show the methodology of the complex rock mechanical investigations from the project level to the monitoring stages.

A large number of excellent figures makes through the chapters the understanding of the text easy.

The preliminary demand is entirely fulfilled to be a text-book not only for students, practicising engineers and geologists but for the surveying engineers or geodesists, working together in the displacement measurement part of the same rock engineering projects.

The bases of conceptual models for rock mass behaviour and discontinuities or faults as well as the finite element approach may be useful for the geoscientists engaged in stress-strain investigations of regional scales caused by geodynamic processes, too.

L Bányai

CONTENTS

Announcement
Cultural Heritage Collected in Libraries of Geoscience, Mining and Metallurgy – Past, Present and Strategy for the Next Millennium

The libraries of geoscience, mining and metallurgy in Europe (as well as on other continents) contain a rich cultural heritage consisting of "old" books, manuscripts, sketches, maps and unpublished works, sometimes also coins, medallions, pewter figures, carvings, etc. The importance of these collections is well known to librarians, historians, restorers, antiquarians, academies, scientific societies and associations, curators of monuments, and many more...
In September, 1993, an international symposium of several days will be held on the above subject in Freiberg, Saxony. This symposium is jointly organized by the Department of reserve précieuse of the Library of Bergakademie Freiberg and the University Library of Montanuniversität Leoben. The first announcement ist to be distributed in mid-1992.
In order to efficiently prepare the symposium, we request all those interested to send us their wishes, comments, suggestions for papers, etc. now.

PRINTED IN HUNGARY

Akadémiai Kiadó és Nyomda Vállalat, Budapest

— treble underlining: bold-face italics
— red underlining: Greek letters
— green underlining: script letters.

Rules for mathematical-physical notations:

— trigonometric, logarithmic, analytic symbols, symbols for units and functions are in roman type (not underlined)
— letter symbols in mathematical and physical formulas, scalars, and subscripts of algebraic and physical quantities are in italics (underlined)
— vectors, matrices, operators in probability theory are in bold-face roman type (double underlining)
— tensors, operators and some special functions are in script letters (green underlining). These cannot be bold.
— Greek letters (red underlining) cannot be bold or extra bold type (thus they cannot be used for vectors or tensors)
— void upper lines e.g. for vectors
— avoid possible confusion between o (letter) and 0 (zero), l (letter) and 1 (one), ν (Greek nu) and v, u (letters) etc.
— explain ambiguous or uncommon symbols by making marginal notes in pencil
— be careful about superscripts and subscripts
— formulae must be numbered consecutively with the number in parentheses to the right of the formula. References in text to the equations may then usually be made by the number in parenthesis. When the word equation is used with a number, it is to be abbreviated, Eq. or Eqs in the plural
— the International System of Units (SI) should be used.

Authors are liable for the cost of alteration in the *proofs*. It is, therefore, the responsibility of the author to check the text for errors of facts before submitting the paper for publication.

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, the title of the paper, title of the journal abbreviated, 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 references.

Examples:

Bomford G 1971: Geodesy. Clarendon Press, Oxford

Brosche P, Sündermann J 1976: Effects of oceanic tides on the rotation of the earth. Manuscript. Univ. of Bonn

Buntebarth G 1976: Temperature calculations on the Hungarian seismic profile-section NP-2. In: *Geoelectric and Geothermal Studies (East-Central Europe, Soviet Asia), KAPG Geophysical Monograph.* Akadémiai Kiadó, Budapest, 561–566.

Gibbs N E, Poole W G, Stockmeyer P K 1976a: An algorithm for reducing the bandwidth and profile of a sparse matrix. *SIAM J. Number. Anal.,* 13, 236–250.

Gibbs N E, Poole W G, Stockmeyer P K 1976b: A comparison of several bandwidth and profile reduction algorithms. *ACM Trans. on Math. Software,* 2, 322–330.

Szarka L 1980: Potenciáltérképezés analóg modellezéssel (Analogue modeling of potential mapping). *Magyar Geofizika,* 21, 193–200.

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

5. *Legends* should be short and clear. The place of the 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.

All the illustrations should contain the figure number and author's name in pencil on the reverse.

Figures will be redrawn. Therefore the most important point is clearness of the figures, even pencil-drawings are accepted (with a duplicate).

Photographs and *half-tone* illustrations should be sharp and well contrasted.

If a specific reduction or enlargement is required, please indicate this in blue pencil on the figure.

The editors will send information to the first author about the *arrival* and acceptance of the papers. A galley proof is also sent to the first author for *correction.* Hundred *offprints* are supplied free of charge.

Periodicals of the Hungarian Academy of Sciences are obtainable at the following addresses:

AUSTRALIA
C.B.D. LIBRARY AND SUBSCRIPTION SERVICE
39 East Splanade
P.O. Box 1001, Manly N.S.W. 2095

AUSTRIA
GLOBUS, Höchstädtplatz 3, *1206 Wien XX*

BELGIUM
OFFICE INTERNATIONAL DES PERIODIQUES
Avenue Louise, 485, *1050 Bruxelles*
E. STORY-SCIENTIA P.V.B.A.
P. van Duyseplein 8, *9000 Gent*

BULGARIA
HEMUS, Bulvar Ruszki 6, *Sofia*

CANADA
PANNONIA BOOKS, P.O. Box 1017
Postal Station "B", *Toronto, Ont. M5T 2T8*

CHINA
CNPICOR, Periodical Department, P.O. Box 50
Peking

CZECH AND SLOVAK FEDERAL REPUBLIC
MAD'ARSKA KULTURA, Národni třida 22
115 66 Praha
PNS DOVOZ TISKU, Vinohradská 46, *Praha 2*
PNS DOVOZ TLAČE, *Bratislava 2*

DENMARK
EJNAR MUNKSGAARD, 35, Nørre Søgade
1370 Copenhagen K

FEDERAL REPUBLIC OF GERMANY
KUNST UND WISSEN ERICH BIEBER
Postfach 10 28 44
7000 Stuttgart 10

FINLAND
AKATEEMINEN KIRJAKAUPPA, P.O. Box 128
00101 Helsinki 10

FRANCE
DAWSON-FRANCE S.A., B.P. 40, *91121 Palaiseau*
OFFICE INTERNATIONAL DE DOCUMENTATION ET LIBRAIRIE, 48 rue Gay-Lussac
75240 Paris, Cedex 05

GREAT BRITAIN
BLACKWELL'S PERIODICALS DIVISION
Hythe Bridge Street, *Oxford OX1 2ET*
BUMPUS, HALDANE AND MAXWELL LTD.
Cowper Works, *Olney, Bucks MK46 4BN*
COLLET'S HOLDINGS LTD., Denington Estate,
Wellingborough, Northants NN8 2QT
WM DAWSON AND SONS LTD., Cannon House
Folkstone, Kent CT19 5EE

GREECE
KOSTARAKIS BROTHERS INTERNATIONAL BOOKSELLERS, 2 Hippokratous Street, *Athens-143*

HOLLAND
FAXON EUROPE, P.O. Box 167
1000 AD Amsterdam
MARTINUS NIJHOFF B. V.
Lange Voorhout 9–11, *Den Haag*
SWETS SUBSCRIPTION SERVICE
P.O. Box 830, 2160 Sz Lisse

INDIA
ALLIED PUBLISHING PVT. LTD.
750 Mount Road, *Madras 600002*
CENTRAL NEWS AGENCY PVT. LTD.
Connaught Circus, *New Delhi 110001*
INTERNATIONAL BOOK HOUSE PVT. LTD.
Madame Cama Road, *Bombay 400039*

ITALY
D. E. A., Via Lima 28, *00198 Roma*
INTERSCIENTIA, Via Mazzé 28, *10149 Torino*
LIBRERIA COMMISSIONARIA SANSONI
Via Lamarmora 45, *50121 Firenze*

JAPAN
KINOKUNIYA COMPANY LTD.
Journal Department, P.O. Box 55
Chitose, *Tokyo 156*
MARUZEN COMPANY LTD., Book Department
P.O. Box 5050 Tokyo International, *Tokyo 100-31*
NAUKA LTD., Import Department
2-30-19 Minami Ikebukuro, Toshima-ku, *Tokyo 171*

KOREA
CHULPANMUL, *Phenjan*

NORWAY
S.A. Narvesens Litteraturjeneste
Box 6125, Etterstad
1000 Oslo

POLAND
WĘGIERSKI INSTYTUT KULTURY
Marszalkowska 80, *00-517 Warszawa*
CKP I W, ul. Towarowa 28, *00-958 Warszawa*

ROUMANIA
D. E. P., *Bucuresti*
ILEXIM, Calea Grivitei 64–66, *Bucuresti*

SOVIET UNION
SOYUZPECHAT — IMPORT, *Moscow*
and the post offices in each town
MEZHDUNARODNAYA KNIGA, *Moscow G-200*

SPAIN
DIAZ DE SANTOS Lagasca 95, *Madrid 6*

SWEDEN
ESSELTE TIDSKRIFTSCENTRALEN
Box 62, *101 20 Stockholm*

SWITZERLAND
KARGER LIBRI AG, Petersgraben 31, *4011 Basel*

USA
EBSCO SUBSCRIPTION SERVICES
P.O. Box 1943, *Birmingham, Alabama 35201*
F. W. FAXON COMPANY, INC.
15 Southwest Park, *Westwood Mass. 02090*
MAJOR SCIENTIFIC SUBSCRIPTIONS
1851 Diplomat, P.O. Box 819074,
Pallas, Tx. 75381-9074
REDMORE PUBLICATIONS, Inc.
22 Cortlandt Street, New York, N.Y. 1007

YUGOSLAVIA
JUGOSLOVENSKA KNJIGA, Terazije 27, *Beograd*
FORUM, Vojvode Mišića 1. *21000 Novi Sad*

Index: 26.031