Acta Geodaetica, Geophysica et Montanistica Hungarica

VOLUME 27, NUMBER 1, 1992

ELECTROMAGNETIC RESULTS IN ACTIVE OROGENIC ZONES

Akadémiai Kiadó, Budapest
AGGM 27 (1) 1–139 (1992) 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

AKADÉMIAI KIADÓ
H-1519 Budapest, P.O. Box 245

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

Special Issue
of the papers read at the
IAGA Symposium 1.8,

ELECTROMAGNETIC RESULTS IN ACTIVE OROGENIC ZONES,

held in Vienna during the
IUGG General Assembly in August, 1991

Edited by

A ÁDÁM

Preface

The IAGA symposium GAM 1.8 with the title **Electromagnetic Results in Active Orogenic Zones** convened by A Ádám/Sopron and G Duma/Vienna was held in Vienna during the IUGG General Assembly. Active orogenic zones are suitable targets for EM induction studies because of enhanced conductivity arising from anomalous temperature conditions, the presence of molten rock, rock fracturing, metamorphism and other tectonic conditions. During the symposium all aspects of the topic were covered by 7 posters and 11 oral presentations. These papers discussed EM results obtained in very different active regions, such as

— Izu-Oshima volcano

— Tucuman Basin

— Periadriatic lineament

— subduction area in Argentina

— Kyushu and Pannonian high heat flow areas

— frontal belt of Garhwal Himalaya

— an earthquake focal area in the Soviet Union

— Southern Central Andes

— Carpathians, *etc.*

together with their theoretical as well as numerical aspects. This special issue of Acta Geodaetica, Geophysica et Montanistica Hungarica presents 10 papers from the mentioned collection. In addition the editor included a paper by L Szarka and Z Nagy on physical modelling results held at another GAM symposium in Vienna.

The editor thanks to the authors for their co-operation.

Sopron, April 1992

A Ádám

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 5–17 (1992)

A NUMERICAL MODEL OF A THREE-DIMENSIONAL SUBDUCTING LITHOSPHERIC SLAB, ITS THERMAL REGIME AND ELECTROMAGNETIC RESPONSE

F W JONES[1], F PASCAL[1], M E ERTMAN[1]

A three-dimensional numerical model that simulates a subducting lithospheric slab is used to investigate the associated thermal field and electromagnetic response. Single slab sections as well as more complicated slab configurations can be considered. Slab movement is simulated by translation of slab temperatures within the numerical mesh, and the subsequent thermal regime is obtained using an alternating direction implicit technique.

The perturbations of time-varying electromagnetic fields caused by electrical conductivity variations related to the temperature variations are studied by means of a three-dimensional numerical model and the field components at the surface as well as induction arrows are computed.

Keywords: electrical conductivity variations; numerical model; 3D subducting slab

Introduction

Minear and Toksoz (1970) and Toksoz et al. (1971) used a quasi-dynamic computational scheme and a finite difference solution of the conservation of energy equation to investigate the thermal regime of a two-dimensional downgoing lithospheric slab and the effects of various parameters on associated geophysical quantities. Sydora et al. (1978) used the same basic method as Minear and Toksoz (1970) but with a modification to the method of temperature translation to investigate the effects of different rates of subduction and varying angles of subduction as well as the generation and effects of partial melt (Sydora et al. 1980). Subsequent to that work, Jones et al. (1981) employed a two-dimensional electromagnetic induction numerical model in which the electrical conductivity distribution was derived from the temperature distribution of a thermal model to investigate the electromagnetic response at the earth's surface of a downgoing slab.

Recently, de Jonge and Wortel (1990) presented temperature distribution maps of the Mediterranean area based on a combination of two-dimensional analyses of approximately three hundred cross-sections using the Minear and Toksoz (1970) technique. They constructed the three-dimensional temperature regime from the two-dimensional model results, and the intent of their work was to extend near-surface geophysical, geological and paleomagnetic observations to structures in the

[1]Department of Physics, University of Alberta,
Edmonton, Alberta, Canada T6G 2J1

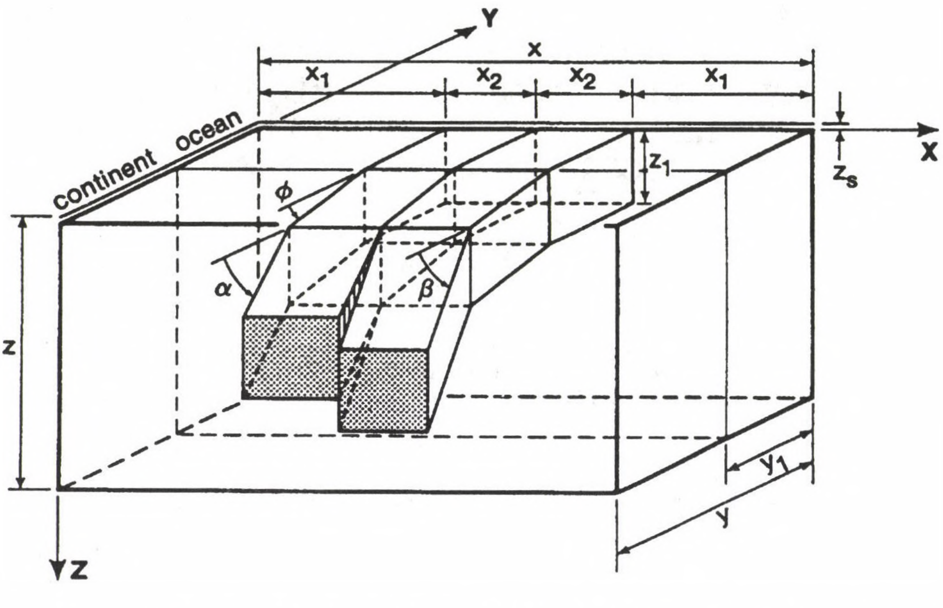


Fig. 1. The three-dimensional downgoing slab model. The slab moves in the $-y$ direction. In the models here, $x = 600$ km, $x_1 = 200$ km, $x_2 = 100$ km, $y = 591.6$ km, $y_1 = 217$ km, $z = 300$ km, $z_1 = 80$ km, $z_s = 5$ km. The model is superimposed on a mesh of $61 \times 61 \times 61 = 226,981$ points with $\Delta x = 10$ km, $\Delta y = 9.86$ km and $\Delta z = 5.0$ km. The time intervals for slab translation are $\Delta t = 0.35$ Ma and so the velocity of slab movement depends on the angles of subduction (α, β *and* ϕ) that can vary

upper mantle. They used their thermal model to do this by converting their derived temperature information to seismic *p*-velocities and thereby modelled the velocity structure.

De Jonge and Wortel (1990) recognized the limitation of constructing a three-dimensional model from many two-dimensional cross-sections in that any lateral heat flow related to along-trench variations of geometry is ignored.

The aim of the present work is to investigate the thermal field of a downgoing slab of limited along-trench extent and the associated lateral heat flow, as well as model the electromagnetic response to such a three-dimensional structure.

The thermal model

The physical model is shown in Fig. 1. The slab moves in the $-y$ direction and begins subduction at the ocean-continent interface. The slab can subduct at different angles during its downward motion and can consist of either a single slab or a slab composed of two parts that can dip downward at different angles as shown.

The quasi-dynamic model is a modification of that of Minear and Toksoz (1970) and Toksoz et al. (1971) in the manner of Sydora et al. (1978) and applied over a three-dimensional grid with grid spacings $\Delta x, \Delta y$, and Δz (see the title to Fig. 1). Computation of the thermal regime is accomplished using the conservation of energy equation

$$C_v \rho \frac{\partial T}{\partial t} = \nabla \cdot (K \nabla T) + H \tag{1}$$

where C_v is the specific heat, ρ the density, T is temperature, K is thermal conductivity, and H is heat generation per unit volume. An alternating-direction implicit finite-difference method (Peaceman and Rachford 1955) is used to solve this equation over the three-dimensional region of interest.

The value of C_v is taken equal to 1.3×10^3 J kg^{-1} K^{-1}, the density distribution follows the Bullen A density curve (see Sydora et al. 1978), and the thermal conductivity is taken as in MacDonald (1959) which includes a constant lattice conductivity together with a radiative transfer term that depends upon temperature.

Heat flux into the bottom of the region is taken constant and equal to 13.5 mWm^{-2}, the surface of the Earth is maintained at 0°C and there is no heat flux across the sides of the model. The initial vertical temperature distribution is taken from relationships given by Mercier and Carter (1975):

Continental:

$T = 7.99(p + 21.4) - 24914/(p + 21.4) + 996$

Oceanic:

$T = 4.34(p + 8.6) - 11840/(p + 8.6) + 1340$

where $p = 0.31z$ kbar km^{-1}. Both heating due to adiabatic compression and shear strain heating are included but no radioactive heat production is assumed. The shear-strain heating is assumed to occur in 10 km thick layers along the top, bottom, end and edges of the slab. Along the top of the slab it is taken as 1.6×10^{-5} Wm^{-3}, whereas along the bottom, end, and edges it is taken as 1.6×10^{-6} Wm^{-3}.

The slab is translated through the mesh for 20 time steps of 0.35 Ma each, for a total of 7 Ma.

Thermal results for particular models

A number of models were run, with different values of σ_0, E (the electrical conductivity and energy gap width for electronic conduction — see MacDonald 1959) and the angles ϕ, α, and β. Various aspects of the thermal field were investigated. However, since we are mainly interested in the possible lateral heat flow because of the finite extent of the three-dimensional model, only results related to this aspect of the model are presented here.

Figure 2 shows examples of the perturbations to the thermal field at several depths by a downgoing slab that is composed of two sections dipping together at 27° for 3.5 Ma and then separating with the two sections dipping at 45° and 56° for the remaining 3.5 Ma. It is seen that the temperature field is strongly affected

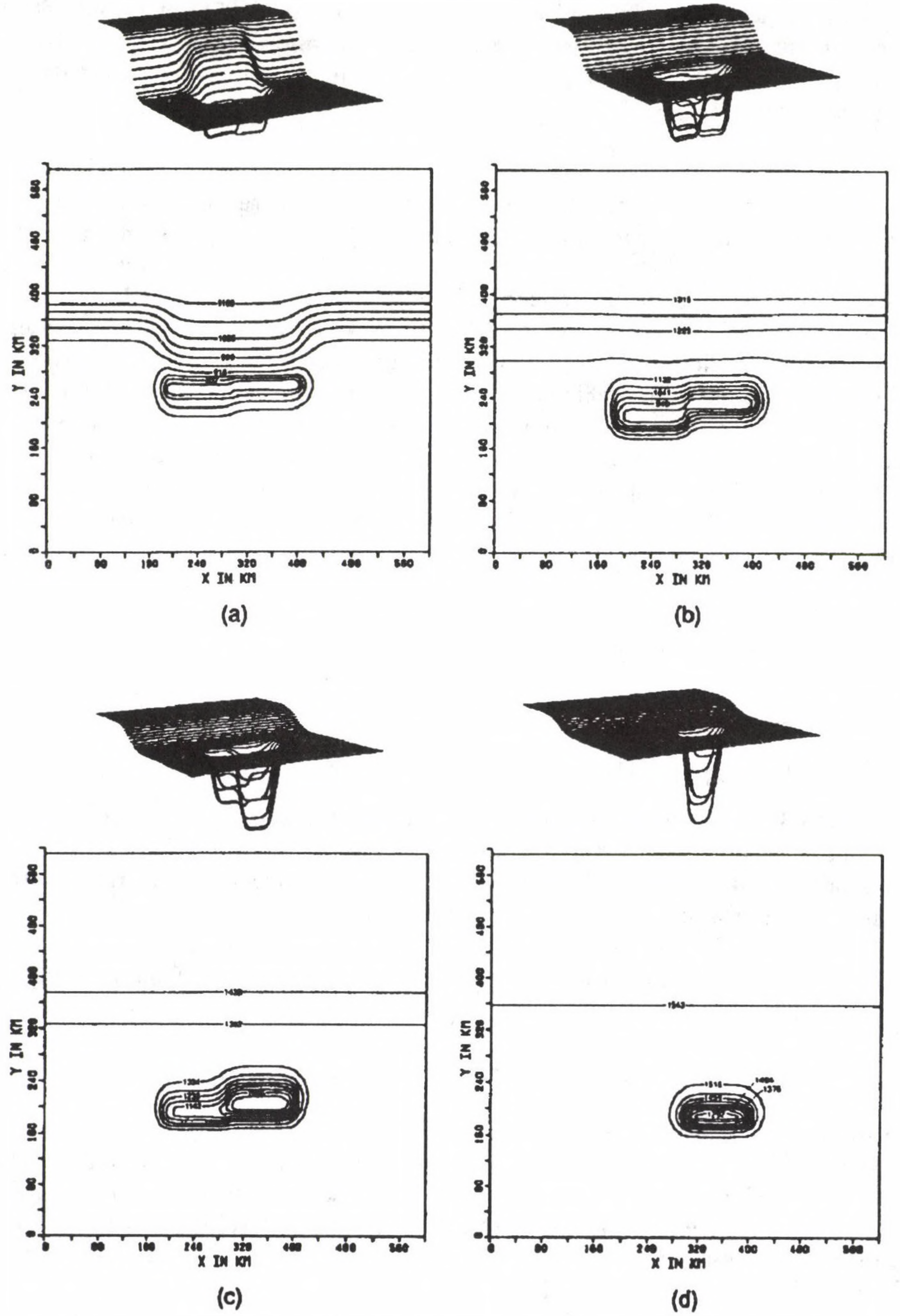


Fig. 2. Perturbation of the thermal field at four depths after 7 Ma by the downgoing slab: $\phi = 27°(0 < t \leq 3.5$ Ma); $\alpha = 45°$, $\beta = 56°$ (3.5 Ma $< t \leq 7$ Ma); $\sigma_0 = 10$ ohm^{-1} cm^{-1}; $E =$ 3.2 eV. The vertical velocities of the slab portions are: $v_z = 3.2$ cm/yr ($\phi = 27°$); $v_z = 4.0$ cm/yr ($\alpha = 45°$); $v_z = 5.1$ cm/yr ($\beta = 56°$). The temperature values are given in °C. a) depth 100 km, b) depth 150 km, c) depth 200 km, d) depth 250 km

by the downgoing slab, and the effect of the two sections is apparent, particularly at 150 and 200 km, while the temperature is affected mainly by the deeper slab section at 250 km depth. The shapes of the temperature contours are suggestive of those found by de Jonge and Wortel (1990) which similarly illustrate the changing perturbation of the temperature field with depth (see Fig. 4, de Jonge and Wortel). It is seen that a substantial horizontal gradient in the temperature field is present, and so lateral heat flow will occur.

The temperature field in vertical planes perpendicular to the direction of slab motion and the lateral heat flow for the same model are shown in Figs 3 and 4. From Fig. 3, it is seen that because of the different oceanic and continental geotherms, the temperatures of the upper parts of the down-going slabs are greater than their surroundings and heat will (initially) flow from the slab material to the surroundings (see Fig. 4). However, for deeper portions of the slabs, heat flows from the surroundings into the slabs. It should be noted that these figures are for time $t = 7$ Ma. At this time the downgoing slabs have just reached the position illustrated in Figs 3e and 4e. If the slab motion stops at this time, the slabs will gradually warm with accompanying decrease in the lateral temperature gradients and heat flows. This effect is seen in Figs 3d, 3c and 4d, 4c, positions where the slabs have been present for some time. Figure 4 shows that the lateral heat flow can be substantial (as much as 50 mWm^{-2} in this example). This depends on the various parameters chosen and some models have exhibited lateral heat flow values 100 mWm^{-2} or more. It is clear from the results obtained that much more must be learned about the rock properties under conditions that exist at depth in order to better assign the various parameters to the models.

It should be noted that these models are for conduction only. Although adiabatic compression and shear strain heating effects have been included, no account is taken of the possible generation of partial melt, latent heat, or the effect of water.

The electromagnetic model

The electromagnetic model is that originally developed by Jones and Pascoe (1972) and used subsequently by Lines and Jones (1973a, b). The numerical method is described by Jones and Vozoff (1978). The model considered is that of a semi-infinite conductor occupying the region $z > 0$ with a plane boundary and which has regions of different conductivity. The electrical conductivities are derived from the temperature field through the relationship

$$\sigma = \sigma_0 e^{-E/kT} \tag{2}$$

where σ_0 is the limiting conductivity as T approaches infinity, E is the width of the energy gap for electronic conduction, k is Boltzman's constant (1.38×10^{-23} JK^{-1}) and T is the absolute temperature. We have found that small changes in E result in large changes in the electrical conductivities. Published values of E differ greatly, depending on the mode of conduction assumed and the material considered (see for example MacDonald (1959) and Rikitake (1966)). The values chosen here for

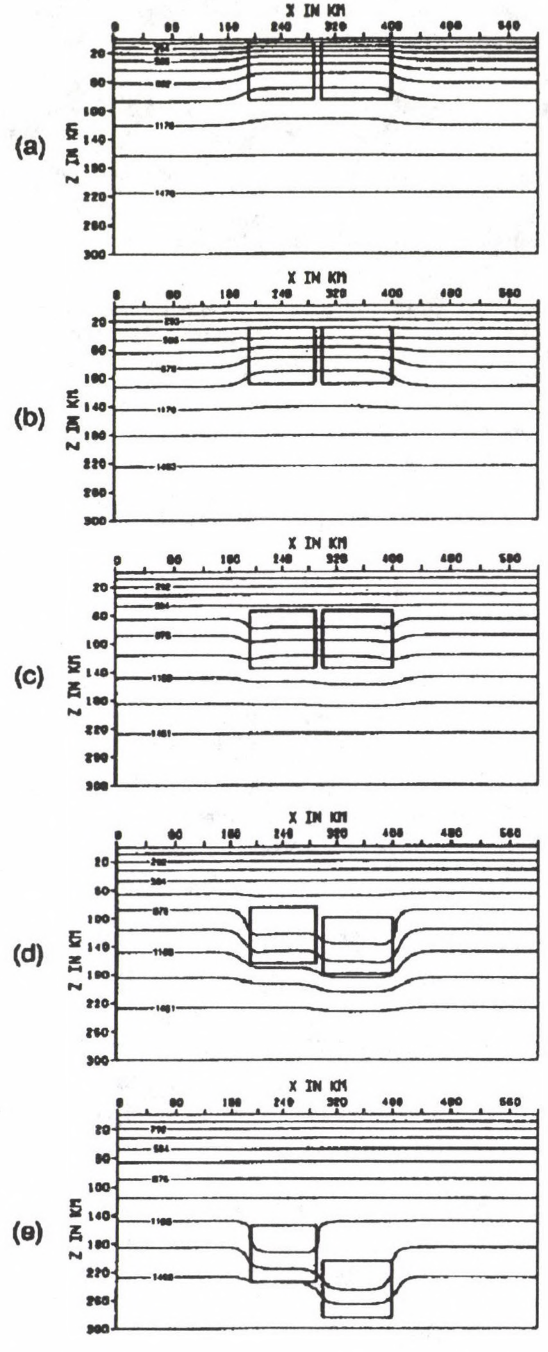


Fig. 3. Same model as in Fig. 2. Temperatures in the $x-z$ plane at five positions in the y-direction at 7 Ma: a) 374.7 km (at ocean-continent interface), b) 325.4 km (between first and second bends), c) 276.1 km (at second bend), d) 246.5 km (between second bend and ends), e) 177.5 km (ends of slabs)

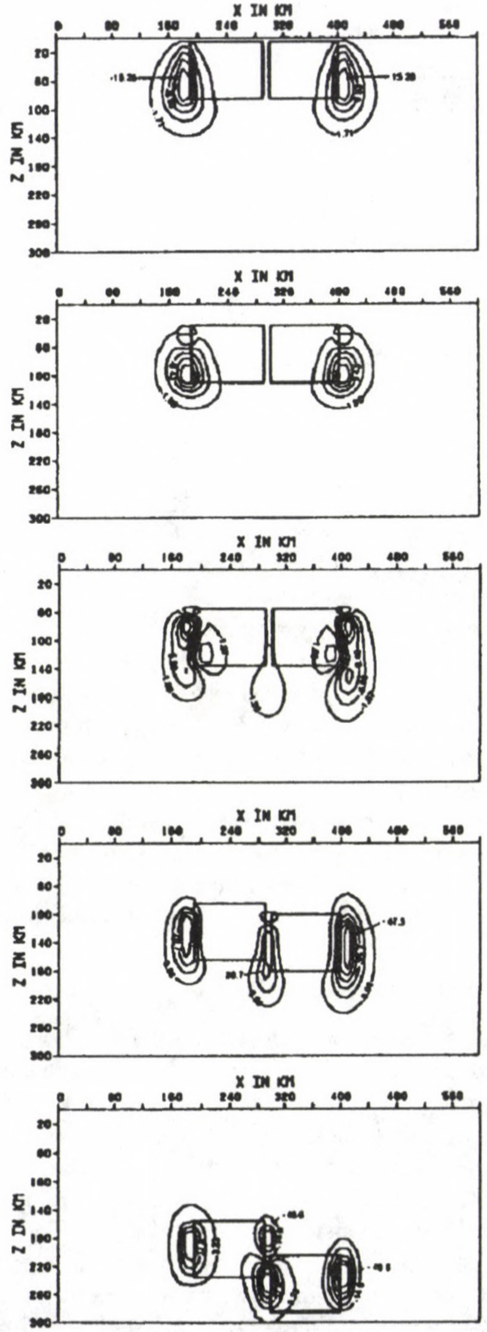


Fig. 4. Same model as in Fig. 2. Heat flow values in the $x - z$ plane at five positions in the y-direction as in Fig. 3 at 7 Ma

σ_0 and E are reasonable for the model and give electrical conductivities consistent with those given by Duba and Lilley (1972).

The source field is taken to be uniform and oscillating with period $2\pi/\omega$ which is sufficiently long that displacement currents can be ignored. The magnetic permeability is taken as that of free space. Under these conditions, Maxwell's equations are:

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} \tag{3}$$

and

$$\nabla \times \mathbf{E} = -i\omega\mu_0 \mathbf{H} \tag{4}$$

where the time factor $\exp(i\omega t)$ is understood in all field quantities and σ is the conductivity appropriate to each region.

If we take the curl of equation (4) and substitute $\nabla \times \mathbf{H}$ from Eq. (3), we obtain an equation in $\mathbf{E}$:

$$\nabla^2 \mathbf{E} - \nabla(\nabla \cdot \mathbf{E}) = i\eta^2 \mathbf{E} \tag{5}$$

where $\eta^2 = \omega\mu_0\sigma$.

The vector Eq. (5) may be re-written as three scalar equations in Cartesian coordinates, converted to finite difference form and solved simultaneously for the components of $\mathbf{E}$ in the three directions at each point of a mesh of grid points surrounding the region of interest by a numerical method. After the electric fields are approximated by the numerical method, the magnetic field components can be calculated by using Eq. (4). Apparent resistivities, induction arrows, and other quantities of interest can be calculated from the field quantities.

The electromagnetic model consists of a surface conducting layer 10 km thick with continental conductivity 10^{-3} ohm^{-1} m^{-1} and ocean conductivity 1 ohm^{-1} m^{-1} overlying the electrical conductivity structure derived from the thermal model. A grid of $38 \times 38 \times 38 (= 54,872)$ mesh points is superimposed over the whole region. The uniform source field is taken to oscillate in the x-direction, and the model is solved iteratively as described by Lines and Jones (1973a, b).

Electromagnetic results for particular models

Examples of results for a two-section slab model are shown in Figs 5–7. Figure 5 shows the electric and magnetic field amplitudes over the surface of the Earth for the model. Both the ocean-continent interface and the presence of the downgoing slab perturb the fields. It is seen that the electric field component parallel to the coastline (E_x) at the ocean surface is substantially less than that over the continent and is perturbed by the presence of the slab. The E_y and E_z components (Figs 5b, c) respond to the three-dimensional nature of the model and illustrate how the induced currents are deflected by regions with changes in electrical conductivity. It should be noted that the changes in electrical conductivity are not abrupt, since they depend on the temperature differences generated by the thermal model which are gradual from one region to another. Effects due to the separation of the two parts of the slab and subsequent difference in subduction angle are apparent in these figures, but are small for this model.

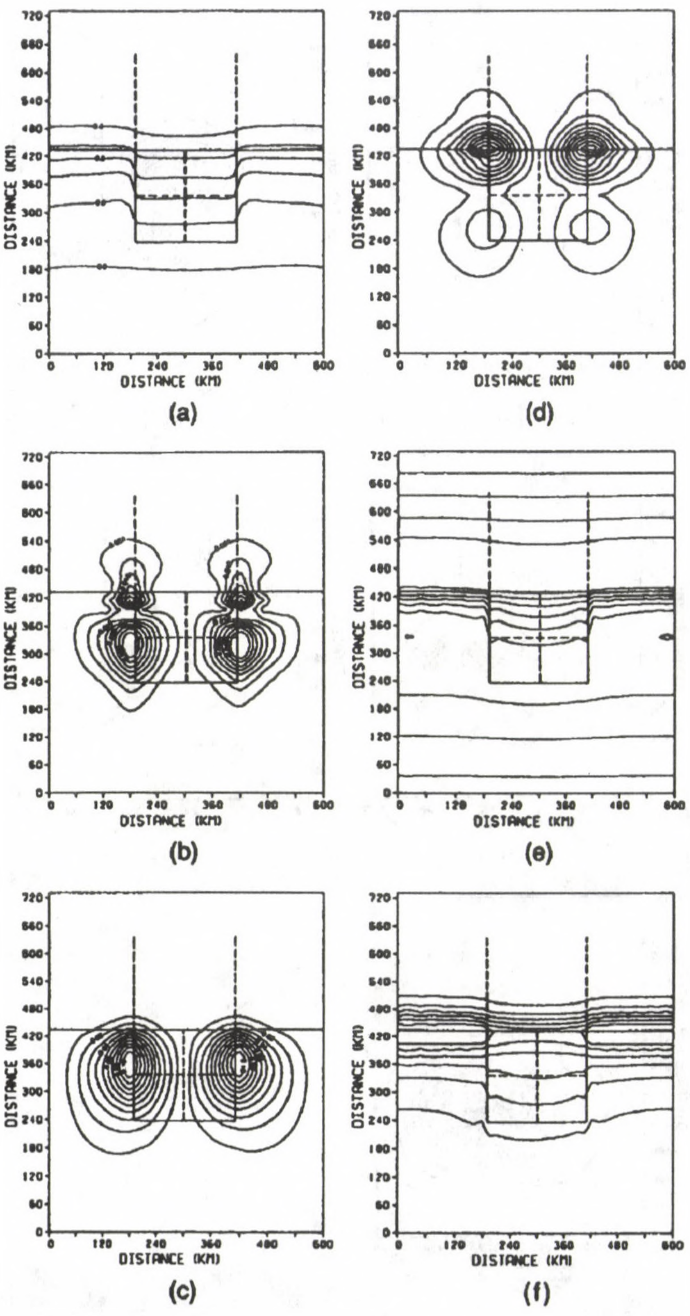


Fig. 5. Amplitudes of the electric and magnetic fields at the Earth's surface for the model: $\phi = 27°(0 < t \leq 3.5$ Ma), $\alpha = 27°$, $\beta = 45°$ (3.5 Ma $< t \leq 7$ Ma), $\sigma_0 = 5 \times 10^2$ ohm^{-1} cm^{-1} for oceanic lithosphere and subducting slab, $\sigma_0 = 10$ ohm^{-1} cm^{-1} for continental lithosphere, $E =$ 0.7 eV everywhere, frequency = 0.0003 sec^{-1} (period = 55.5 min). All quantities are normalized with respect to $E_x = 1.0$ at (0,0). a) E_x, b) E_y, c) E_z, d) H_x, e) H_y, f) H_z

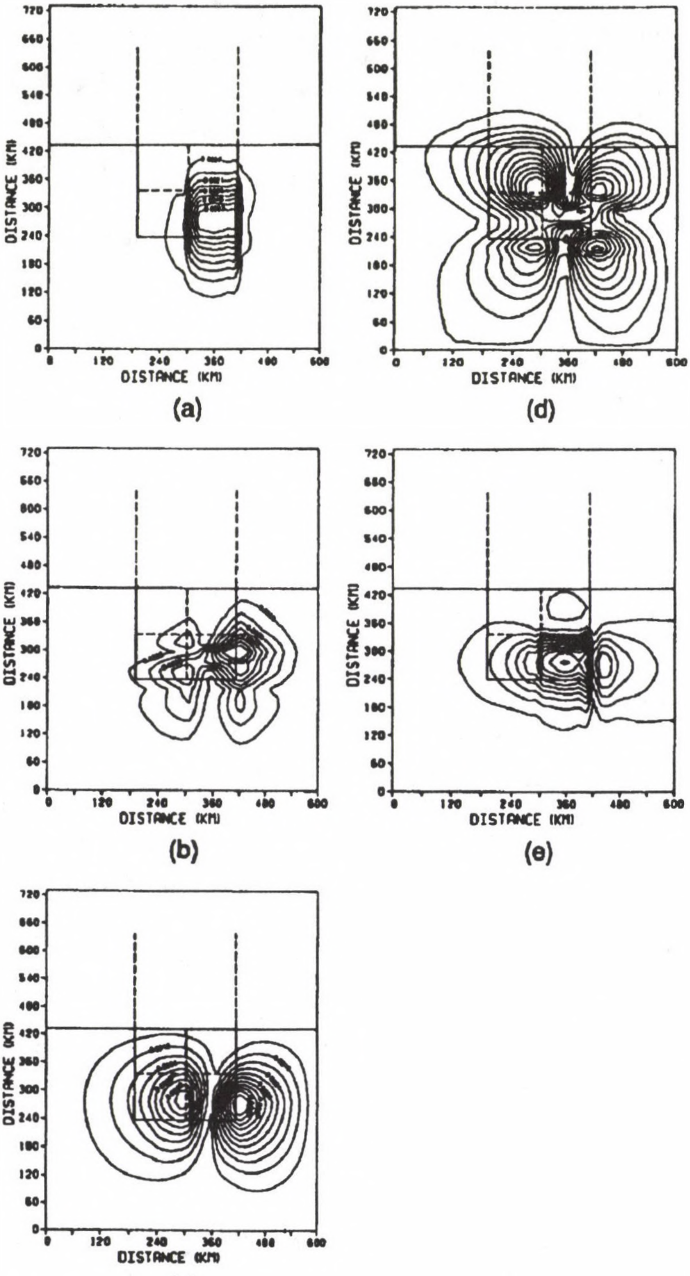


Fig. 6. Differences in the electric and magnetic field amplitudes at the Earth's surface $\Delta F = F_A - F_B$, where Model A = The model of Fig. 5. Model B = Similar to model A but with both α and $\beta = 27°$. a) ΔE_x, b) ΔE_y, c) ΔE_z, d) ΔH_x, e) ΔH_y (ΔH_z was not contoured because its values were small and irregular)

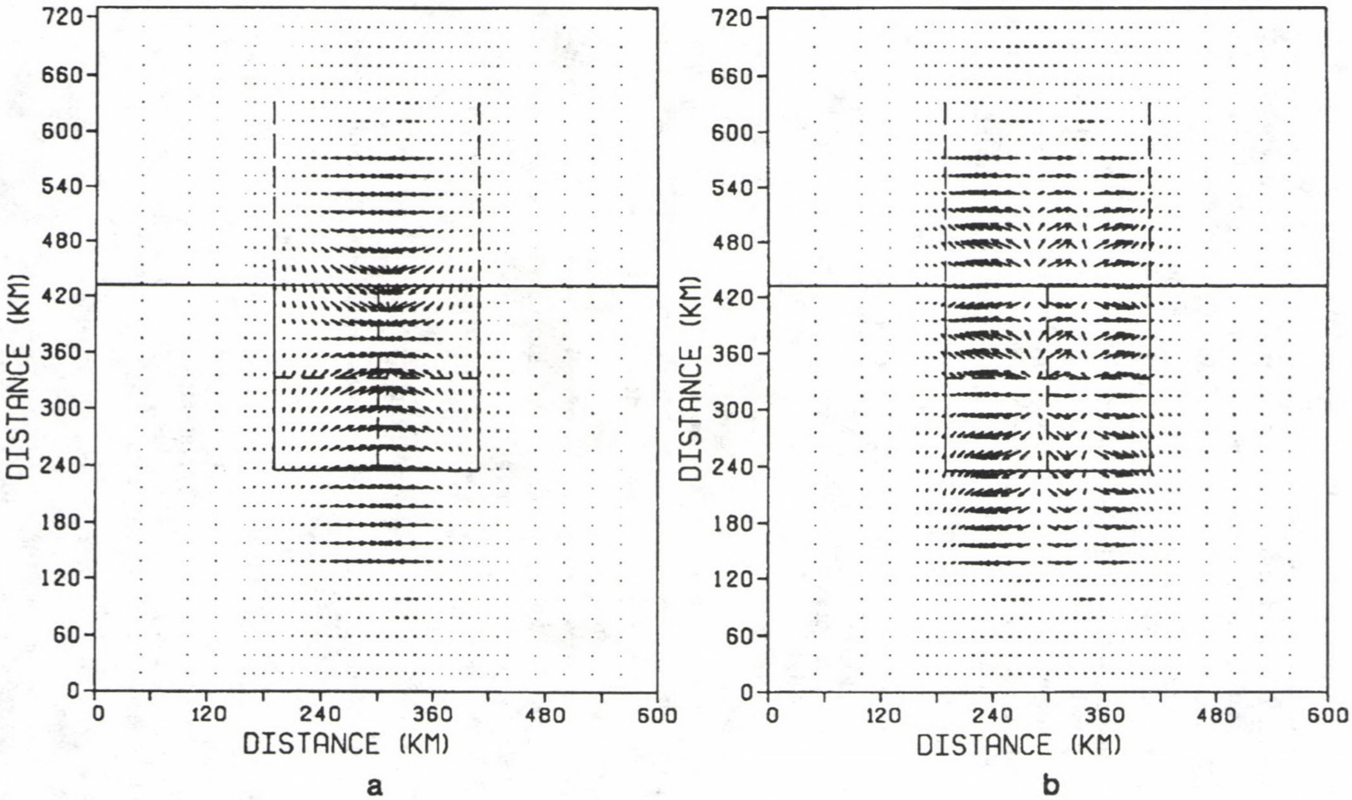


Fig. 7. In-phase (a) and quadrature (b) induction arrows for model A of Fig. 6

Although the effect of slab separation on the field quantities can be seen in Fig. 5, it appears small. However, when two models are compared, the differences become much more apparent. Figure 6 illustrates the differences that exist in the amplitudes of the field quantities betwen a single conducting slab and one that separates into two parts that subduct at different angles. These difference plots clearly show the slab section that has separated from the main slab and subducted at a different angle beginning at 3.5 Ma. Indeed, for E_z (Fig. 6c), we even see the boundaries of that portion outlined, between that portion and the surrounding continental material as well as between that portion and the main slab.

The in-phase and quadrature induction arrows for a case in which two slab sections separate (as in Fig. 5) are shown in Fig. 7. For this case, with the source field in the x-direction, we would expect the current flow to concentrate in the downgoing slabs and to be deflected into the upper part of the slab that dips more abruptly (the right hand side of the slab in Fig. 7). Such current concentration is indicated by the in-phase induction arrows of Fig. 7a.

Conclusions

A three-dimensional numerical model of a downgoing lithospheric slab has been developed. Both the thermal regime and electromagnetic response of the slab can be investigated. It is found that lateral heat flow can be significant, equal to or greater than the vertical heat flow, though this depends on the particular model and the time after subduction has occurred. The electromagnetic response of the slab has been studied. The field quantities are perturbed and differences exist between single slab and multiple slab models. These differences depend on the model parameters, but in general appear to be small.

The results depend to a great extent on the parameters chosen, and it is clear that much more must be known about the relationship between temperature and electrical conductivity as well as such quantitites as the width of the electronic energy gap and their dependence on the conditions at depth within the Earth.

Acknowledgements

This work has been supported by the Natural Sciences and Engineering Research Council of Canada.

References

de Jonge M R, Wortel M J R 1990: *Terra Nova*, 2, 609–616.
Duba A, Lilley F E M 1972: *J. Geophys. Rev.*, 77, 7100–7105.
Jones F W, Pascoe L J, Ramaswamy V, Sydora L J 1981: *J. Geoph. Res.*, 86, 10870–10874.
Jones F W, Vozoff K 1978: *Geophysics*, 43, 1167–1175.
Jones F W, Pascoe L J 1972: *Geophys. J. R. astr. Soc.*, 27, 479–485.
Lines L R, Jones F W 1973a: *Geophys. J. R. astr. Soc.*, 32, 133–154.
Lines L R, Jones F W 1973b: *Can. J. Earth Sci.*, 10, 510–518.
MacDonald G J F 1959: *J. Geoph. Res.*, 64, 1967–2000.
Mercier J C, Carter N L 1975: *J. Geoph. Res.*, 80, 3349–3362.
Minear J W, Toksoz M N 1970: *J. Geoph. Res.*, 75, 1397–1419.

Peaceman D W, Rachford Jr. H H 1955: *J. Soc. Incl. Appl. Math.*, 3, 28–41.
Rikitake T 1966: Electromagnetism and the Earth's Interior. Elsevier, Amsterdam
Sydora L J, Jones F W, Lambert R St J 1978: *Can. J. Earh Sci.*, 154, 626–641.
Sydora L J, Jones F W, Lambert R St J 1980: *Tectonophysics*, 62, 233–249.
Toksoz M N, Minear J W, Julian B R 1971: *J. Geoph. Res.*, 76, 1113–1138.

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 19–34 (1992)

GEOMAGNETIC INDUCTION ANOMALIES IN THE FRONTAL BELT OF GARHWAL HIMALAYA

C D REDDY[1] AND B R ARORA[1]

The analysis of transient geomagnetic variational events recorded by a simultaneously operated chain of magnetometers across the Main Central Thrust of the Garhwal Himalaya indicates that induction effects are determined by a small wavelength feature superimposed on an anomaly of regional character. Regional anomaly is shown to result from the concentration of induced currents in the Indo-Gangetic block, south of the Main Frontal Thrust. Estimate of geoelectric parameters from frequency dependence and also from simple 2-D numerical modelling suggests that thick conducting sediments in Indo-Gangetic foredeep contribute little to the overall conductance required to produce the spatial characteristics of observed induction anomalies. Much of the conductance is provided by some deep seated structure.

A narrow conductivity anomaly, embedded in a high seismicity zone, mapped a little south of the Main Central Thrust, is believed to be caused by free water trapped in some obducted or a back thrusted belt. A suggestion is made that the recent Uttarkashi earthquake of October 20, 1991 may be a manifestation of the release of strain energy accumulated in the fluids of this conducting belt.

Keywords: conductivity anomaly; Garhwal Himalaya; geomagnetic variations; high seismicity; Indo-Gangetic block

1. Introduction

The collision of the Indian and Asian plates along the Indus-Tsangpo Suture Zone (ITSZ) and subsequent successive underthrusting of the Indian plate at the Main Central Thrust (MCT) and Main Boundary Thrust (MBT) have played key role in the evolution of the Himalayan orogenic belt (for review see Windley 1988). The ITSZ together with these regional intra-crustal thrusts (MCT and MBT) subdivide the Himalayan tectogene from north to south into the Tethyan belt, the main central crystalline belt of the Higher Himalaya, the lesser Himalaya and Foothill Siwalik belt. The yet vaguely recognised Main Frontal Thrust (MFT), also referred to as Himalayan Frontal Fault (HFF), separates the Siwalik molasse of the foothills from the thick alluvial deposits in the Indo-Gangetic plains. These major thrust planes form foci of earthquakes of various magnitude which signify continued state of tectonic restlessness (Valdiya 1984).

Over the last decade, a number of magnetovariational studies have been carried out to investigate the sub-surface structures associated with various thrusts belts of the Himalayan tectogene, in terms of electrical conductivity. The high sensitivity of this physical parameter to fluids, conducting mineral as well as to high

[1]Indian Institute of Geomagnetism, Colaba, Bombay 400 005, India

temperatures prevailing at greater depths makes such an investigation a useful tool to map high electrical conductivity zones resulting from the trapping of free water released during metamorphic dehydration or thermally induced silicate melting and also those associated with enriched minerals like graphite or carbon. Mapping fluids

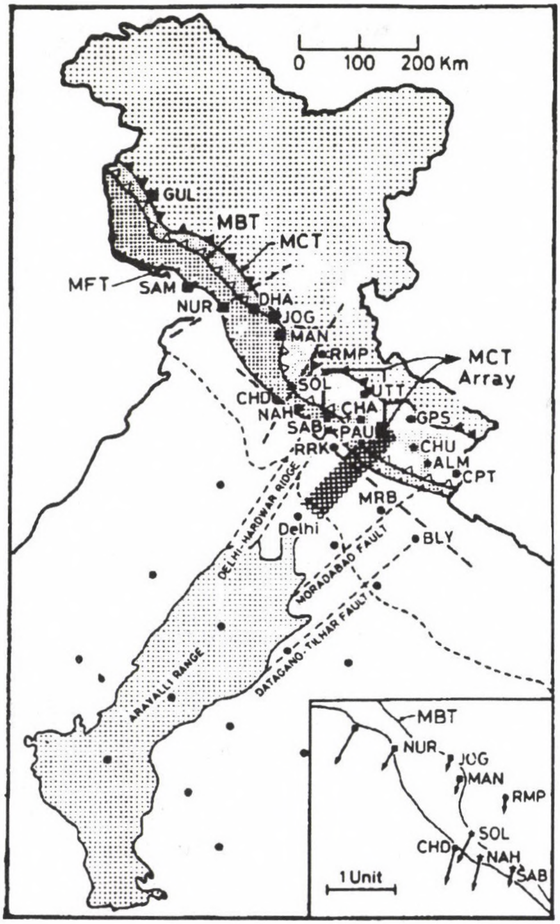


Fig. 1. Map of northern India showing the magnetometer coverage around the frontal thrust zones of Himalaya covered in different magnetometer array studies. Hatched area shows location of Trans-Himalayan conductor. The inset on bottom right gives the real induction arrows corresponding to a period of 82 min for selected stations around the Main Frontal Thrust of Garhwal and Kangra regions

in the crust or upper mantle have proved important pointer of current regional or global tectonic processes whereas the detection of conductive minerals have helped tracing older tectonic structures (Gough 1989). The magnetovariational studies in the Himalaya have continuously received impetus from the All India Co-ordinated

Project on Seismicity and Seismotectonics of the Himalayan region. The results of these investigations which have so far been focussed (Fig. 1) around the MBT, MFT and adjoining Indian shield are summarised in Arora (1990) and Arora and Singh (1992). The present paper describes the first results of a new magnetovariational study undertaken during November-December, 1987 to provide coverage across the Main Central Thrust of the Garhwal Himalaya. The section of the MCT selected for the present study is still tectonically active as evidenced by the occurrence of a large number of moderately strong earthquakes (Gaur et al. 1985). The study area also includes the epicentral track of the recent devastating Uttarkashi earthquake of magnitude 6.2 that rocked the area on October 20, 1991.

2. Layout and operation of the magnetometers

Figure 2 gives the layout of magnetometer sites in relation to various tectonic features. Nine fluxgate magnetometers were placed approximately perpendicular to

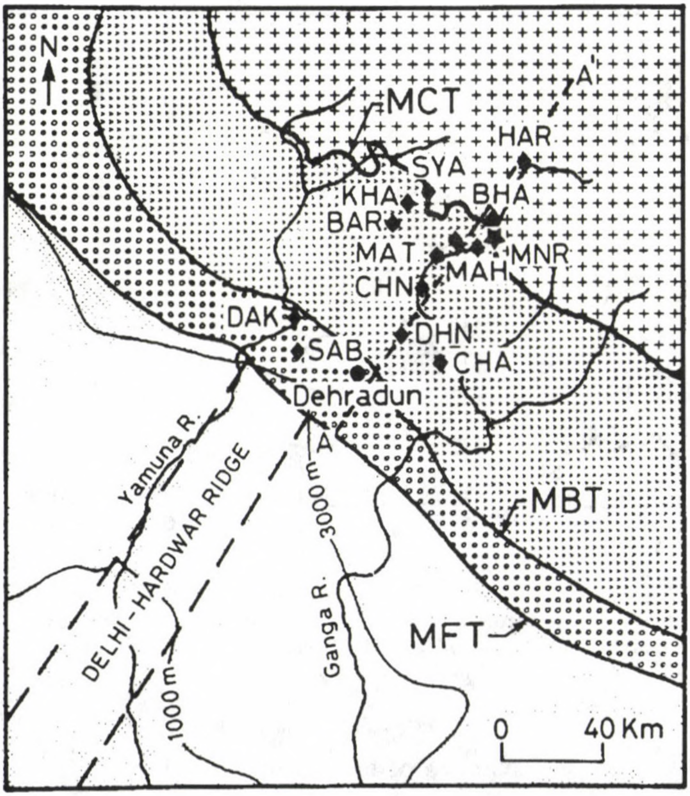


Fig. 2. Layout of the magnetometers across the Main Central Thrust of the Garhwal Himalaya, covered as a part of the present study. The epicentre of Uttarkashi earthquake ($\star$) is also shown. Line AA′ marks the profile referred to in text and figures

the MCT along two lines. Six of these stations *viz.* HAR, BHA, MNR, MAH, MAT and CHN along an eastern line coupled with two more stations, one permanent observatory at Sabhawala (SAB), and another, DHN, covered in earlier survey, provide coverage from the north of MCT to a little south of MBT. Three of the stations *viz.*, SAY, KHA and BAR lie along the western line. Seven of the nine sites covered in this array are common with the microearthquake recording stations operated by the University of Roorkee (Gaur et al. 1985).

The EDA fluxgate magnetometers that were deployed in the present study, recorded time variations in three geomagnetic field components, *viz.* the geographic north (X), east (Y) and vertical (Z) components. These magnetometers which have a measuring accuracy of 0.4 nT, sampled variations at regular interval of 30 sec over a period of about six weeks during November-December, 1987.

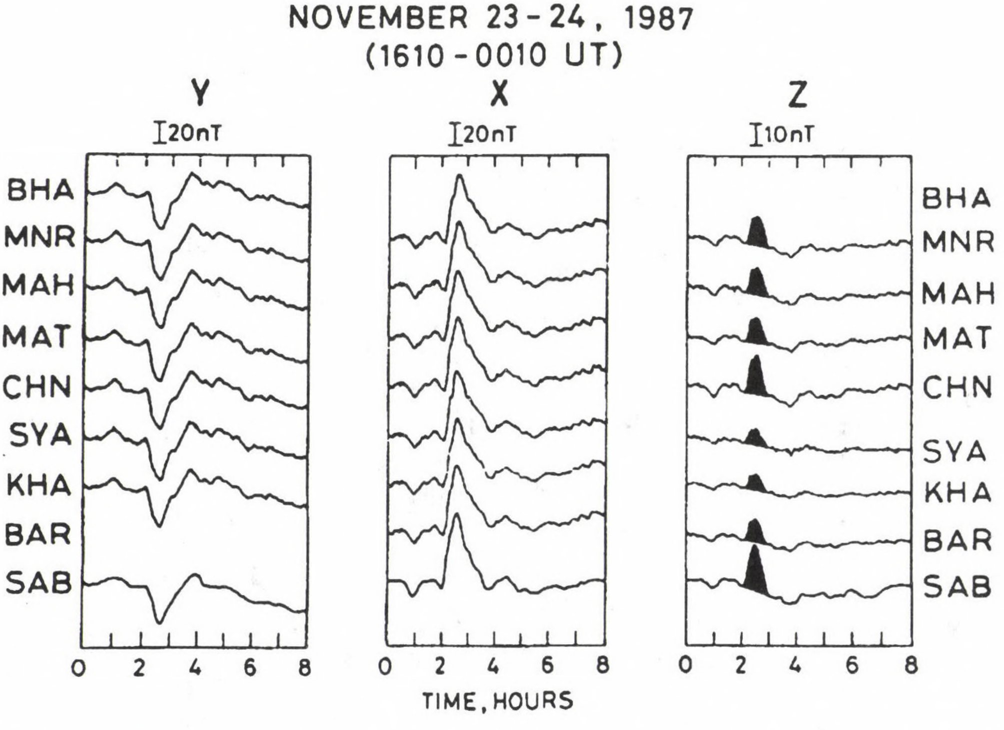


Fig. 3. Magnetograms of a disturbance event for the stations covered during the present MCT array. Note the marginal suppression of Z at MAT over the general increasing trend from northern to southern sites

3. Nature of transient variations

From the continuous stretch of data for a period of over six weeks, some 15 disturbance events were selected and normalized. Figure 3 gives a typical example on the nature of transient variations recorded by the array. The Z variations, which after physical considerations are more prone to sub-surface structures, clearly show a progressive increase in their magnitude from northern to southern stations, both along the eastern and western profiles. One, though not very pronounced but persistent feature, seen on many disturbance events has been the suppression of Z fluctuations at MAT which is later shown to indicate a localized conductivity anomaly a little south of the MCT. The marked similarity of X and Z fluctuations, in-terms of internal currents, suggest a concentration of induced currents along the E-W discontinuity to the south of the array.

4. Transfer function analysis

A variety of disturbance events with a wide range of polarizations and frequency content, after the removal of trend and cosine tapering at both ends of the data section, have been fast Fourier transformed. Fourier components at selected periods were used to calculate vertical-field transfer function (T_{zx}, T_{zy}) summarising the linear relationship between the anomalous vertical field (Z_a) and the normal component of the horizontal inducing (X_n, Y_n) field as expressed below (Schmucker 1970, Beamish 1977).

$$Z_a = T_{zx} \cdot X_n + T_{zy} \cdot Y_n + \epsilon. \tag{1}$$

Here all the quantities are complex and frequency dependent. ϵ represents the uncorrelated part of the vertical field. The transfer functions are evaluated under two simplifying assumptions; firstly the observed Z at each station is considered to be entirely anomalous $(Z_a = Z)$ and secondly the observed horizontal components away from the anomalous zone are taken as a measure of normal inducing field (X_n and Y_n). Ideally, the estimation of X_n and Y_n from the records of the permanent magnetic observatory Sabhawala (SAB) would enable homogeneous estimation of transfer functions for the present, as well as for sites occupied in earlier campaigns. A uniformity would exist in the transfer functions obtained since they are related to the horizontal field components at a single reference site. However, the horizontal components at Sabhawala are found to contain a substantial anomalous field determined by the effect of current channelling (unpublished results of the present authors) and, thus, bias the estimation of transfer functions. In the present study, mean of X and Y variations at Bhatwari (BHA) and Khardi (KHA), which recorded data uninterruptly, were used to calculate inter-station type transfer functions.

Two different approaches are used here for the presentation of the transfer functions. In the first commonly adopted method of presentation, a complex pair of induction arrows are defined for vertical fields responding in-phase (G_R-Real) and in-quadrature (G_I-Imaginary) with the horizontal component with which the vertical field possesses maximum correlation. The induction arrows have a simple interpretation only when the structure is 2-dimensional. In such circumstances, the arrows point at right angles to the strike of the local geoelectric lateral gradient giving rise to an anomalous concentration of currents. When the transfer function possesses an intermediate phase, the separation into real and imaginary components can be artificial (Beamish 1985). The maximum and minimum response functions introduced by Banks and Ottey (1974) prove effective in separating from T_{zx} and

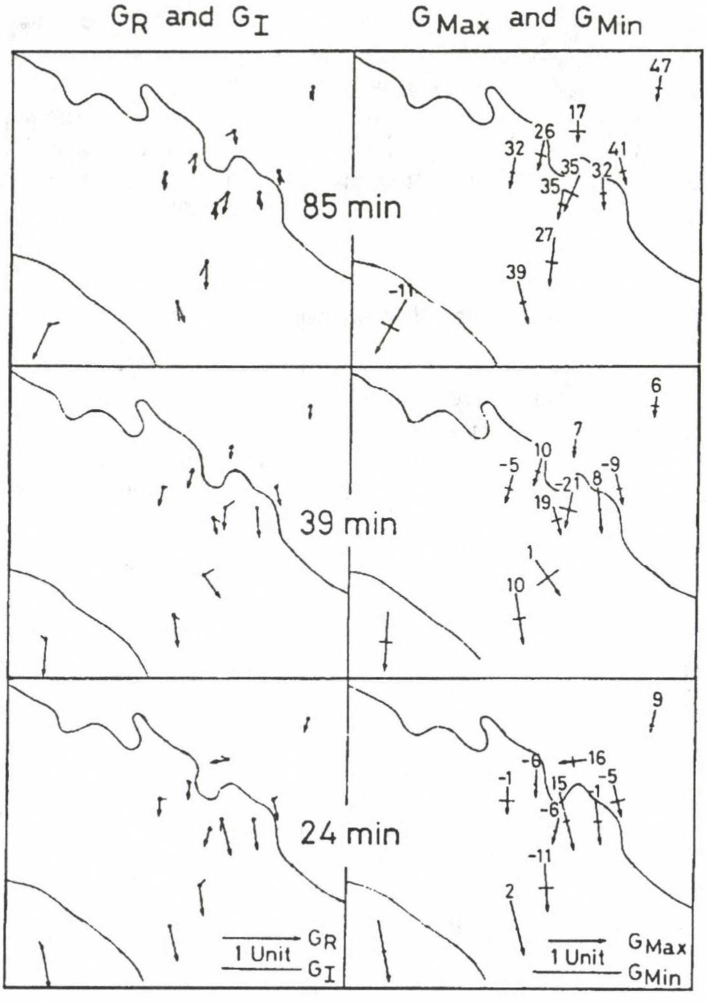


Fig. 4. Real (G_R) and quadrature (G_I) induction arrows as well as maximum ($G_{\max}$) and minimum($G_{\min}$) response functions corresponding to the sites of present study

T_{zy} that part of the response that is compatible with a 2-dimensional assumption. In this approach, Eq. (1) is transformed to a new set of axes such that the modulus of response is maximum along one axis and minimum along the other axis. The presentation and the meaning of the maximum response function, $G_{\max}$, is similar to the real induction arrows. The minimum response ($G_{\min}$) is shown along an axis orthogonal to the azimuth of $G_{\max}$ and the relative strength of $G_{\max}$ and $G_{\min}$ is a measure of how far the two-dimensional assumption is justified (Banks and Ottey 1974). In the present work, some eight sets of induction arrows as well as maximum and minimum response functions in the period range of 16–128 min were calculated and three sets corresponding to periods of 85, 39 and 24 min are displayed in Fig. 4.

5. Physical interpretation of induction effects

The simplest induction arrow pattern is seen at long periods of 85 (Fig. 4) and 128 min. The real induction arrows consistently have southerly orientations and

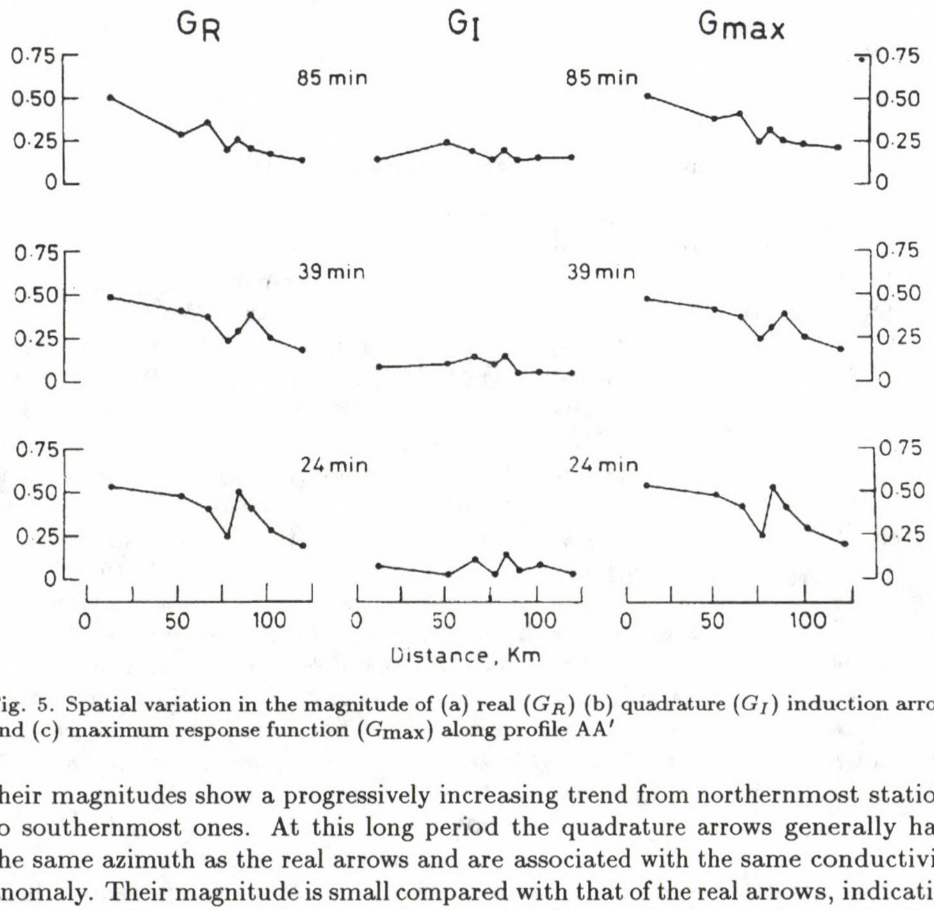


Fig. 5. Spatial variation in the magnitude of (a) real (G_R) (b) quadrature (G_I) induction arrows and (c) maximum response function ($G_{\max}$) along profile AA′

their magnitudes show a progressively increasing trend from northernmost stations to southernmost ones. At this long period the quadrature arrows generally have the same azimuth as the real arrows and are associated with the same conductivity anomaly. Their magnitude is small compared with that of the real arrows, indicating that anomalous vertical field variations are in-phase with the inducing field. Since the induction response is largely in-phase ($G_R > G_I$), there is a high degree of equivalence between the real induction arrows and the maximum response functions ($G_{\max}$). The directional patterns of $G_{\max}$ as well as G_R clearly indicate the presence of a high conductivity zone further south of the study area, aligned along the foothills of the Himalaya.

Consistent with the attenuation of the vertical field, the magnitude of the induction arrow at MAT is marginally suppressed. The effect is more conspicuous on Fig. 5 which displays the spatial behaviour of G_R, G_I and $G_{\max}$ along profile AA', placed roughly at right angle to the MFT/MBT (see Fig. 2 for location) and formed by grouping data of the eastern group of stations. Clearly, the strength of G_R and $G_{\max}$ show a small scale anomaly centred around MAT and MAH, superposed on a long-regional anomaly marked by a general decreasing trend away from MFT. This small wave-length anomaly indicates a narrow conductive structure a little south of the surface trace of MCT. The persistence of this anomaly at all periods examined with a clear dominance at short periods further suggests that the induction effects are related to a body of high electrical conductivity rising from greater depth to a relatively shallow a depth. Perhaps due to the superposition of this local effects, a group of stations located in the immediate vicinity, show substantial $G_{\min}$ which are found to be negligible at most of the sites.

6. Seat of the regional conductivity anomaly

The large magnitude southerly and south-southeasterly directed arrows in the vicinity of the MFT, with gradually decreasing magnitude arrows away from the frontal belt, are pointers of a large-scale conductivity anomaly whose source is located further south of the present study area. The results of the NW India array had shown that along this sector the induction arrows south of the MBT are controlled by the Trans-Himalayan conductor, aligned along the strike of the Aravalli range, beneath the Indo-Gangetic plains (Arora et al. 1982, Arora 1990). At stations of the NW India (CHD and RMP) and Garhwal arrays (SOL and NAH), located far away from the influence of the Trans-Himalayan conductor, as well as at stations (MAN, JOG, NUR, SAM) of the Kangra array, (see Fig. 1 for location of the above referred sites) the real induction arrows (given in an inset of Fig. 1), point at right angles to the main geological grain of the frontal folded belt similar to the present array. The magnitude of the induction arrows without any suggestion of reversal attain large values in close vicinity of MFT and were interpreted by Mahashabde et al. (1989) to be indicative of a strong electrical discontinuity along the MFT. 4–5 km thick wet alluvial sediments in the Indo-Gangetic plains (Raiverman et al. 1983) to the south of the MFT provide natural seat for ehanced internal current flow. Noting the similarity of the geoelectrical cross-section across the MFT to that across the coast line, Mahashabde et al. (1989) suggested that the dominant spatial and frequency characteristic of the anomalous vertical field can be explained by the edge effect of the induced currents flowing in the vast stretch of thick conducting sediments in the Indo-Gangetic foredeep. The gradually decreasing magnitude of the real induction arrows (seen on the data of the present array in Figs 4 and 5), on the resistive side of the electrical discontinuity aligned with the MBT further corroborate the above interpretation.

7. Frequency characteristic of induction effects

A number of analytical and numerical procedures have been developed to study the edge effect of induced currents in electrically thin conducting sheet, taken to represent the ocean (Hewson-Browne and Kendall 1976, Fischer 1979, Weaver 1979).

Beamish (1985) using a matching technique of Hewson-Browne and Kendall (1976), as adopted by Quinney (1979), examined the frequency dependence of G_R and G_I for locations away from the edge of thin sheet. The similarity of the observed induction features to that of edge effect allow a comparison with that predicted by the model calculation of Beamish. His result showed that over a period interval of 100–10,000 sec, the magnitude of the real and quadrature arrows, respectively, show a general decreasing and increasing trend with increasing period. For periods in excess of 1000 sec, a characteristic asymptotic frequency response is observed which is a function of distance from the edge (Fig. 2 of Beamish 1985). The solution also provided that for the period range corresponding to the observational bandwidth of present study, $G_R > G_I$.

Turning now to the observed frequency response, shown in Fig. 6 for selected sites, representing different lithotectonic belts of the Himalaya. As stated earlier the variations at SAB are contaminated by channelling effect. Frequency characteristics representative of the Siwalic block are shown by station DAK, placed similarly in relation to MFT (see Fig. 2 for location). Individual stations show certain characteristic features but on the general trend, approximated by parabolic fit to response, falling and rising behaviours of G_R and G_I with increasing period are consistent with frequency characteristic seen in calculated values from a thin-sheet model (Beamish 1985). However, the observed asymptotic behaviour of G_R and G_I at periods greater than 1000 sec is less marked than predicted by the model for sites located only some tens of km from the edge of the thin conducting sheet. A superposition of induction effects related to some deep structure in the crust or mantle may distort the frequency response solely due to an isolated near surface 2-dimensional electrical structure. However, it is practically impossible to distinguish the deep and surface anomalies by the shape of observed frequency dependence either in amplitude or phase (Rokityansky 1982, p. 268).

8. Estimation of geoelectrical parameters

When studying analytically the frequency response of strong 2-D structures, Rokityansky (1982) and Chen and Fung (1985) noted that there exists a characteristic period, T_0, where the modulus of the real induction arrow is maximum. Across such period, the quadrature arrow passes through zero and its direction flips over by 180° which can be more directly seen as the near parallel and anti-parallel alignment of real and quadrature arrows across the period T_0. If this period (T_0 in sec) can be identified on the frequency response curve observed over dominantly 2-D anomalies, the empirical relation

$$G = 5 \times 10^4 (T_0)^{1.2}, \qquad (2)$$

established by Rokityansky (1982), can be used to calculate longitudinal conductivity ($G = \sigma Q$, Siemens m), determined by the product of conductivity (σ, Siemens/m) and cross-sectional area ($Q = h \cdot L$ in m^2) of the anomalous conductivity body. For a surface sheet-like structure, h and L represent thickness and width of the band of anomalous currents.

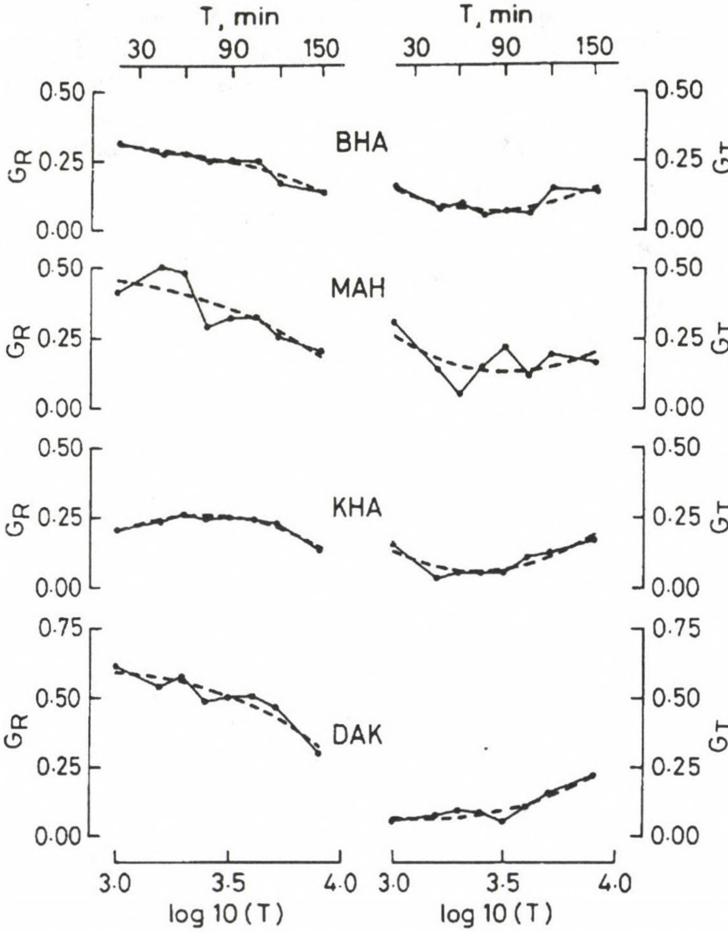


Fig. 6. Changes in the magnitude of real (G_R) and quadrature (G_I) arrows as function of period for selected stations. The broken curve shows parabolic fit to observed response

The frequency response curves in Fig. 6 do not exhibit any clear maximum in G_R, rather G_R stabilizes to a nearly constant high value at periods below 3000 sec (50 min). In contrast, the associated imaginary part, G_I, displays a near zero minimum around the period range of 30–40 min as anticipated for a 2-D anomaly. The lack of true zero in G_I may imply a departure from near-field 2-dimensionality (Beamish 1985). The near parallel orientation of real and quadrature induction arrows at long periods (e.g. for 85 min in Fig. 4) and large deviation in the azimuth at short periods, reaching almost anti-parallel at certain sites (Fig. 4c), further corroborate that the characteristic period is close to 30 min. For $T_0 = 30$ min, the estimated value of longitudinal conductivity from Eq. (2) is about 4×10^8 ohm·m. Further, considering that the band-width of the anomalous current is approximately

80–100 km, corresponding to the width of sedimentary belt in the Indo-Gangetic plains, the expected overall conductance for the conducting body causing observed induction effect is of the order of 4000–5000 Siemens.

9. Geoelectrical block model

To gain some insight whether induction across the tectonic boundaries separating the different lithological belts of Himalaya could explain, even if semi-quantitatively,

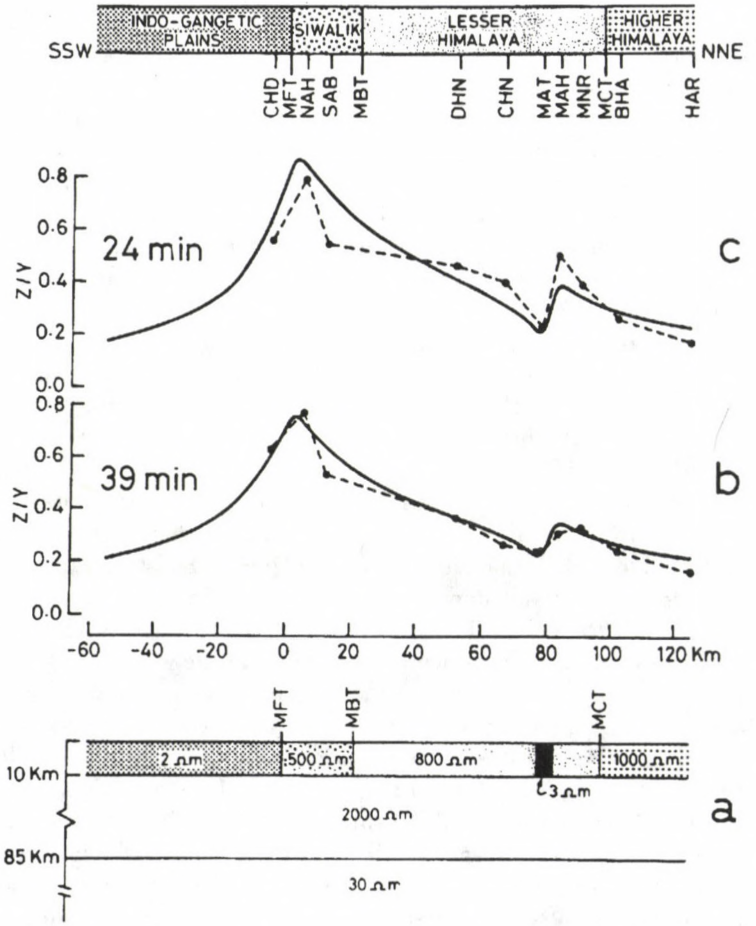


Fig. 7. (a) Proposed block geoelectrical model along a profile AA′ (Fig. 2). Observed and calculated real induction arrows (Z/Y) for periods of (b) 39 min and (c) 24 min are also shown

the spatial characteristics of anomalous vertical fields, a numerical block model has been developed. The 2-D geoelectrical model with its homogeneous axis running parallel to the general strike of the Himalayan foothills is considered to align with profile AA' (see Fig. 2 for location). On profile AA' each latitudinal belt of the Himalayan tectogene is represented by vertical blocks of 10 km height with different electrical parameters (Fig. 7). From left to right, blocks represent the Indo-Gangetic plains, the Siwalik folded belt, the lesser Himalaya and the higher crystalline Himalaya. A narrow block embedded in the lesser Himalaya denotes a localized conductive zone centred around MAT and MAH. The block structure is underlain by two layers simulating respectively the bulk of the resistive lithosphere and the conducting asthenosphere. The electromagnetic induction response and associated induction arrows are worked out using the computer program developed by Červ and Pek (private communication).

The distribution of electrical resistivity in the model was adjusted in a number of steps to improve the fit between the model and observations, although a care was exercised so that the adopted distribution has some semblance to the lithological character of respective blocks. Figure 7b, c show the calculated and observed real induction arrows at periods of 39 and 24 min. Given the simplicity of the model, a quantitative comparison is not warranted. However, a general agreement in the decay pattern of observed and calculated values north of the MFT is clearly apparent. The spatial characteristics of the localized conductivity structure around MAT and MAH are reproduced by a 5 km wide block with an equivalent conductance of 2500 Siemens.

Considering that the observed enhanced vertical field to the north of the MFT is a manifestation of the edge effect of the currents circulating in the sediments or some deep seated structure beneath the Indo-Gangetic plains, the above simple model calculations show that the block simulating these layers would have an equivalent conductance of about 5000 Siemens. This is in fair agreement with that deduced from the relationship between G and T_0. Given that the sediments in the Indo-Gangetic plains in the immediate vicinity of MFT have a thickness of approximately 4–5 km and resistivity log data from deep wells suggest a resistivity of 10–15 Ohm m (Vozoff 1984) for the alluvial deposit. Thus, the sedimentary layer can at best account for a conductance of 250–500 Siemens. This suggests that the bulk of the conductance required to produce the observed induction effect is provided by some deep seated source. A downward step-like structure in the mantle arising due to the thickening of the crust beneath the Himalaya, in accordance with the isostatic model, provide suitable configuration, induction in which would cause an effect similar to the edge effect of sheet current. More quantitative constraints on the geometrical and geoelectrical parameters would require more comprehensive 2-D modelling which at this stage is inhibited by the lack of information on the nature of induction, particularly south of MFT. To generate desired data sets for this exercise, operation of a dense network of magnetometers, extending from the southern end of Indo-Gangetic plains to the higher Himalaya along, the Karnal-Chakrata-Tuni-Sarswati Nagar profile (further west of the present study area) is recenctly accomplished, and an other along the Bisalpur-Tanakpur-Pithoragarh-Askot line in Kumaun Himalaya (east of the present study area) is in progress.

10. Correlation of the localized conductive belt with gravity and seismicity pattern

The signature of a certain structural discontinuity aligned with the mapped conductivity belt, a little south of MCT, is also evident in other geophysical data.

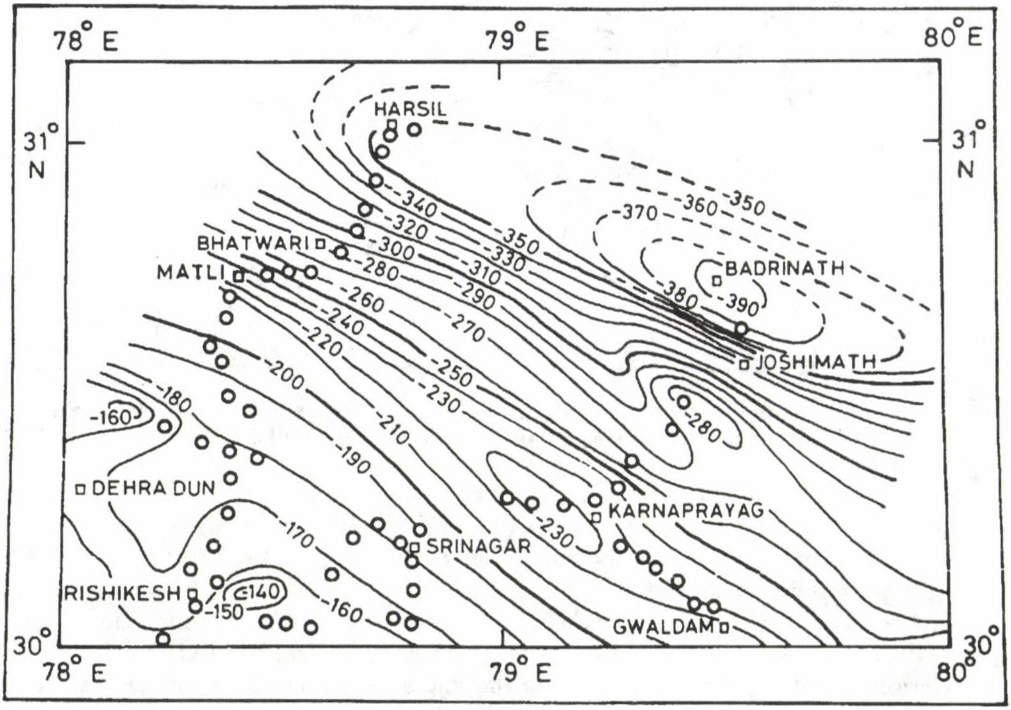


Fig. 8. Gravity anomaly map for Garhwal Himalaya (based on Qureshy et al. 1974). Note the sharp gradient around Matli in the western corner

On the gravity map of the region (Qureshy et al. 1974), shown in Fig. 8, the area of conductivity anomaly is characterized by a large gravity gradient, signifying the sharp density discontinuity. Figure 9 gives the contour plot of the anomalous vertical field associated with the MCT conductor *vis-a-vis* the epicentral map of micro-earthquakes as recorded by a network of seven stations (Gaur et al. 1985). A well defined concentrated belt of hypocentres located to the south of the surface trace of MCT is quite apparent. Khattri et al. (1989) have estimated the width of this seismic belt to be of the order of 10–30 km with epicentres of most earthquakes lying at depths of less than 13 km. It is interesting that the axis of mapped conductive zone, as judged from the line separating contours of high and low anomalous vertical fields in Fig. 9, shows a good positional correlation with the southern edge of the high seismicity belt over its entire length from northwest of the Yamuna valley to southeast of the Alakhananda valley. A geographical correlation of the high seismicity zone in respect of the Trans-Himalayan conductor has been

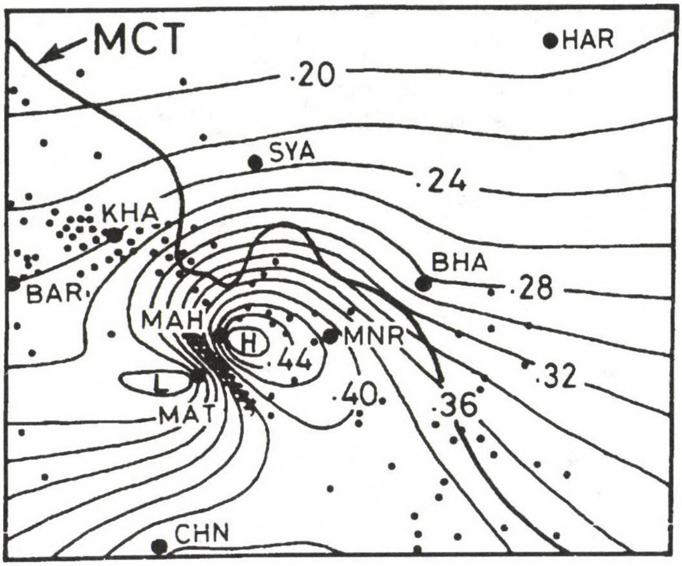


Fig. 9. Contour plots of real part of the induction arrow (G_R) near localised conductive zone *vis-a-vis* the spatial distribution of the epicentres of micro-earthquakes (after Gaur et al. 1985)

earlier noted by Arora and Mahashabde (1987). The noted correlation between the deep conductive structure and the zone of high seismicity in the Himalaya or elsewhere may represent different manifestations of common causes such as some kind of ductile flow at depth (Lilley 1975) or increased lubrication of faults due to the presence of free water released during metamorphic dehydration (Fyfe et al. 1978) or frictional heating. The coincidence of the shallow conductive anomalies with the seismic belt may simply imply a continuity of saline ground water as a result of extensive fracturing in response to stress release.

Ni and Barazangi (1984) using medium size earthquake events occurring in immediate vicinity of MCT concluded that these earthquakes were not related to MCT. Instead, they proposed a planer zone at about 10–20 km depth with an apparent northward dip of about 15°. This planer zone coincides with the Basement Thrust Fault (BTF) separating the underthrusting Indian plate from the overriding lesser Himalayan crustal block (Seeber et al. 1981). The narrow high seismicity belt located just south of MCT may be related to a certain small scale undulation in this dipping plane which can geologically be viewed as an obducted or backward thrusted zone along the BTF. The high density anomaly seen over this part tends to support such a conjecture. This obducted zone may owe its enhanced conductivity to hot saline water released during tectonic reactivation or free water from dehydration processes. If the presence of such conducting fluid could be confirmed from the seismic wave velocity structure, the noted correlation between high seismicity and fluid filled conducting zone may raise an interesting question whether fluids could serve as potential reservoir of strain which is released on exceeding a certain threshold value. The recent Uttarkashi earthquake of October 20, 1991 with its epicentre embedded in the presently mapped zone may be one such example of energy release accumulated in the fluids.

11. Conclusions

The nature of the induction anomalies seen in the frontal belt of the Garhwal Himalaya reveal a regional anomaly, associated with the concentration of induced currents in the Indo-Gangetic plains, with a small wave-length anomaly superposed a little south of the Main Central Thrust. Constrainsts on electrical parameters imposed by the frequency dependence of observed induction effects and by simple 2-D numerical model warrant that the main source of the regional anomaly is some deep seated structure, currents in the sediments of Indo-Gangetic plain contribute little to the observed induction pattern. More detailed 2-D modelling with fresh data from south of MBT is required to confirm the geometrical and geoelectrical parameters of the deep seated conductive structure.

The persistence of induction features associated with a narrow conductive belt south of MCT at all periods is interpreted to indicate a back thrusted or obducted zone of transformed crust above the plane of MBT. As the most likely mechanism for enhanced conductivity is due to conducting fluids in fractured rocks or free water released during progressive metamorphism, it seems likely that increased lubrication of sub-surface fault may account the positional correlation of high conductivity and seismic belts.

Acknowledgements

This work has been carried out as a part of "All India Coordinated Project on Seismicity and Seismotectonics of the Himalayan Region" sponsored by the Department of Science and Technology (DST), Govt. of India. Financial support received under this project is gratefully acknowledged. We thank S S Bist, V J Jacob and Vijay Dhapre for their help in installation and care of the recording instruments. We are grateful to B P Singh for his encouragement and suggestions at all stages.

References

Arora B R 1990: *Earth Planet. Sci.*, 99, 693–716.

Arora B R, Lilley F E M, Sloane M N, Singh B P, Srivastava B J, Prasad S N 1982: *Geophys. J. R. astr. Soc.*, 69, 459–475.

Arora B R, Mahashabde M V 1987: *Phys. Earth Planet. Int.*, 45, 119–127.

Arora B R, Singh B P 1992: *J. Geol. Soc. India* (in press)

Banks R J, Ottey P 1974: *Geophys. J. R. astr. Soc.*, 36, 321–335.

Beamish D 1977: *Geophys. J. R. astr. Soc.*, 50, 311–332.

Beamish D 1985: *Journal of Geophysics*, 57, 207–216.

Chen P F, Fung P C W 1985: *Geophys. J. R. astr. Soc.*, 80, 257–263.

Fischer G 1979: *Proc. IEEE*, 67, 1050–1060.

Fyfe W R, Price A, Thompson 1978: Fluids in the Earth's crust. Elsevier, Amsterdam

Gaur V K, Chander R, Sarkar I, Khattri K N, Sinhval H 1985: *Tectonophysics*, 118, 243–251.

Gough D I 1989: *Rev. of Geophys.*, 27, 141–157.

Hewson-Browne R C, Kendall P C 1976: *Acta Geod. Geoph. Mont. Hung.*, 11, 427–446.

Khattri K N, Chandra R, Gaur V K, Sarkar I, Sushil Kumar 1989: *Earth and Planet. Sci.*, 98, 91–109.

Lilley F E M 1975: *Nature*, 257, 381–382.

Mahashabde M V, Waghmare S Y, Arora B R 1989: *Earth and Planet. Sci.*, 98, 319–326.

Menvielle M, Rossignol R C, Tarits P 1982: *Phys. Earth Planet. Int.*, 28, 118–128.

Ni J, Barazangi M 1984: *J. Geoph. Res.*, 89, 1147–1163.

Quinney D A 1979: *Geophys. J. R. astr. Soc.*, 56, 119–126.

Qureshy M N, Venkatachalam S, Subrahmanyam 1974: *Bull. Geol. Soc. Ann.*, 85, 921–926.

Raiverman V, Kunte S V, Mukherjee A 1983: In: Petroliferous basins of India. Himachal Times Group, Dehradun, 67–92.

Rokityansky I I 1982: Geoelectromagnetic investigations of the Earth's crust and mantle. Springer Verlag, Berlin

Schmucker U 1970: *Bull. Scripps Inst. Oceanogr.*, 13, 1–165.

Seeber L, Armbruster J G, Quittmeyer R C 1981: In: Zagros, Hindu-Kush Himalaya geodynamic evolution. Gupta H K, Delany F M (eds), Geodynamic Series, 3, 215–242.

Valdiya K S 1984: Aspects of tectonics. Focus on South Central Asia. Tata Mc Graw-Hill Publishing Company Limited, New Delhi, 137–155.

Vozoff K 1984: *Tectonophysics*, 105, 399–411.

Weaver J T 1979: *Proc. IEEE*, 67, 1044–1050.

Windley B F 1988: *Phil. Trans. R. Soc. Lond.*, 326, 3–16.

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 35–45 (1992)

CARPATHIAN CONDUCTIVITY ANOMALY AND ITS RELATION TO DEEP STRUCTURE OF THE SUBSTRATUM

E PINNA[1], A SOARE[2], D STĂNICĂ [2], M STĂNICĂ [2]

The present paper offers some new information regarding the geographical position and origin of the Carpathian electrical conductivity anomaly (CECA), in Romania, taking into account the data supplied by the Wiese induction vectors map, by magnetotelluric soundings, as well as by other geophysical data. The complex interpretation of these data allowed us to put forth some hypotheses referring to the origin and the depth at which the CECA is situated.

Keywords: Carpathian electrical conductivity anomaly; magnetotelluric soundings; Wiese induction vectors

Introduction

Since being pointed out for the first time by Wiese (1963) using geomagnetic induction vectors, CECA has constituted, from the point of view of its spatial extension and of the causes that generated it, the object of several studies (Ádám 1974, 1976, Ádám and Pospíšil 1984, Calotă et al. 1971, Jankovsky et al. 1977, Ney 1975, Praus et al. 1980, Ritter 1970, Rokityansky et al. 1975) see Fig 1.

Figure 2 represents a general image of the Wiese induction vectors determined on the basis of standard geomagnetic recordings in over 200 stations covering the whole territory of Romania. The parameters of these vectors (amplitude and orientation) correspond to the inductive response of some transient geomagnetic phenomena included in a range of periods between 5 min to 50 min, therefore reflecting an inhomogeneity in the distribution of the electrical conductivity down to great depths in the Earth's crust and upper mantle.

In contrast to previous works, vectors presented in Fig. 2 have been mostly tested by a processing method that allowed the stabilization of the transfer function in view of homogenizing the external sources. This method included the computation of the amplitude parameters for the induction vectors again so that it was supposed $\Delta Hz/\Delta Hx \equiv \Delta Hz/\Delta Hy \equiv$ const., using narrow-bandpass convolutional filtering in order to get a rigorous selection of the frequency range where field sources are plane-polarized.

It was thus possible to trace on a length of more than 1000 km the axis of CECA's stretch probably associated with the development of a zone in the Earth's crust of higher conductivity. Wiese induction vectors have a divergent orientation at the sides of CECA, both being perpendicular to the axis of the anomalously conducting body.

[1]Earth Sciences Depart., S.Maria 53, 56126 Pisa, Italy
[2]Inst. of Geology and Geophysics, Caransebes 1, 78344 Bucharest 32, Romania

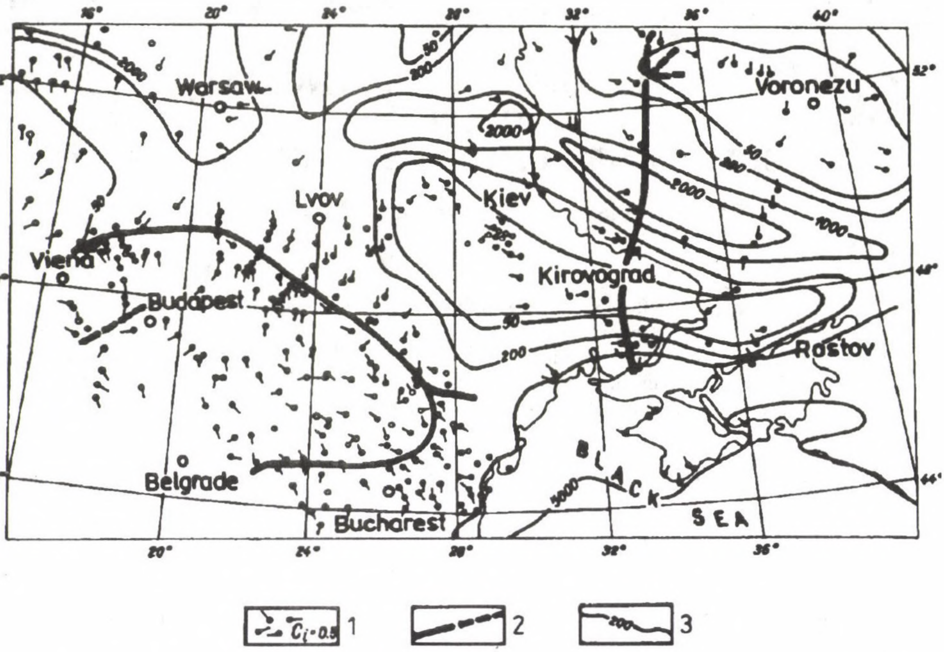


Fig. 1. Wiese induction vectors, electrical conductivity anomalies and S-contours for Central and Eastern Europe (Rokityansky 1975): 1 – Wiese induction vectors (Cw) for periods 15–60 min, 2 – Electrical conductivity anomalies in the lithosphere, 3 – S-countours of total longitudinal conductance of the sediments in mho

As regards the causes which generated the CECA, several hypotheses have been offered, the most important of which are the following:

- Rokityansky et al. (1975) have shown that the CECA may be ascribed to the metamorphic changes of the sedimentary rocks under the influence of temperature and pressure at the depth of 10–25 km;
- Ney (1975) explained CECA as an effect of lithospheric subduction;
- Ádám and Pospíšil (1984) have demonstrated that the CECA may be caused by partial melting of sediments with water release during metamorphism, by the partial melting of crystalline rocks (granite, basic and ultrabasic complexes) or by serpentinization.

For the time being, the causes which generated the CECA still remain in the domain of hypotheses, considering the great depth where the eventual conducting bodies are situated.

The present paper brings forth some more precise data regarding the geographic position of CECA, and also contributes on the basis of new magnetotelluric information, to the elucidation of some quantitative aspects on the depth at which the anomalously conducting body is situated.

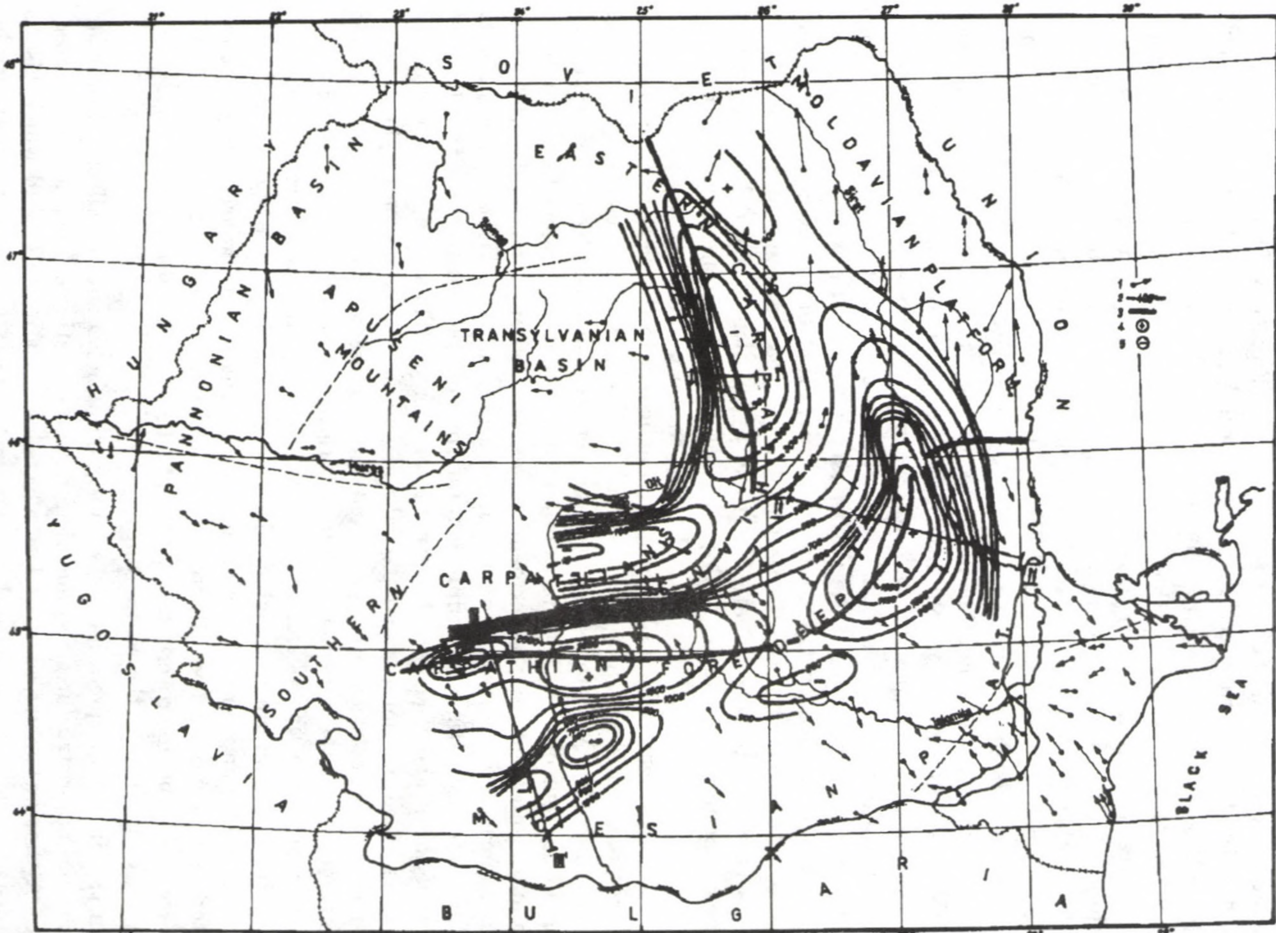


Fig. 2. Wiese induction vectors, CECA and S-contours for Romania: 1 – Wiese induction vectors (Cw) for periods 5–50 min, 2 – S-contours of the total longitudinal conductance of sediments, in mho, 3 – CECA, 4 – Maximum anomaly of S, 5 – Minimum anomaly of S

Results

The results of magnetotelluric researches obtained on 14 profiles that cross the Carpathian Orogen and the neighbouring zones are presented in the form of a map showing the distribution of effective conductance (Fig. 2), MTS curves (Figs 3 and 4) and in the form of cross sections through deep-lying layers (Figs 5, 6 and 7).

In the area of the present development of the crystalline of the Central-East Carpathian Nappes system North of the 46-th parallel, the CECA can be explained by the presence of saline water in fractured rock, or of hydrated minerals, such as amphiboles, created when sediments or oceanic crust are subducted (Camfield et al. 1989). Using petrophysical data Rădulescu and Săndulescu (1973) have demonstrated in the Carpathian Chain two zones with oceanic crust which had developed by means of spreading ever since the Triassic up to early Cretaceous. They underwent later a consumption due to the tectogenetic processes, during Cretaceous and Tertiary, turning into an ophiolithical suture, called "consumption paleoplane". One should exclude the possibility, neither, a simultaneous action of several causes (i.e. ruptural tectonics + the local effects of basic and ultrabasic complexes). This hypothesis about the origin of the CECA was also put forth by Ney (1975), Ádám (1974), and confirmed by the complex interpretation of the geological and geophysical data collected in the Carpathian Orogen (Săndulescu and Visarion 1979, Stănică et al. 1986, Soare et al. 1989) and by supplementary data furnished by 15 magnetotelluric soundings (MTS) achieved along the profile Zetea - Mădăraş - Nădejdea (Fig. 5). The MTS profile I-I' was chosen perpendicularly to the general north-south tendency of the Carpathians, implicitely to the electrical conductivity anomaly, in order to get an optimum record of the electrical currents induced along the structure. The MT processing program performs spectral decomposition of the records, and four different estimates of impedance tensor elements were computed from auto-power and cross-power density spectra. Computation of all of them provided a measure of the total amount of noise present, as indicated by a stability coefficient for the estimates. Additional information, related to the signal-to-noise ratios of the E and H signals, have been considered. Thus, the correction of the results is based on multiple and partial coherences. The sums being the basis of all subsequent calculations were preserved and no impedance values were computed from them only if the multiple coherence with respect to Ex and Ey, and geometric mean of the partial coherences (Ex, Hy) and (Ey, Hx) were greater than 0.7. The two sets of information, which describe current flowing along the structure (E-polarization) and current flowing across the structure (B-polarization) were extracted (Fig. 3 and 4). These figures illustrate the apparent resistivity responses in four frequency bands, covering the interval 0.001–20 Hz with standard errors less than 1% over virtually the whole range.

In this initial interpretation we wish to minimize 2D and 3D effects in order to determine major structures responsible for the observations. Therefore we concentrate on that parameter which is less affected by "static shift" and which also is responsive to the gross structural information. This is the apparent resistivity for the E-polarization mode.

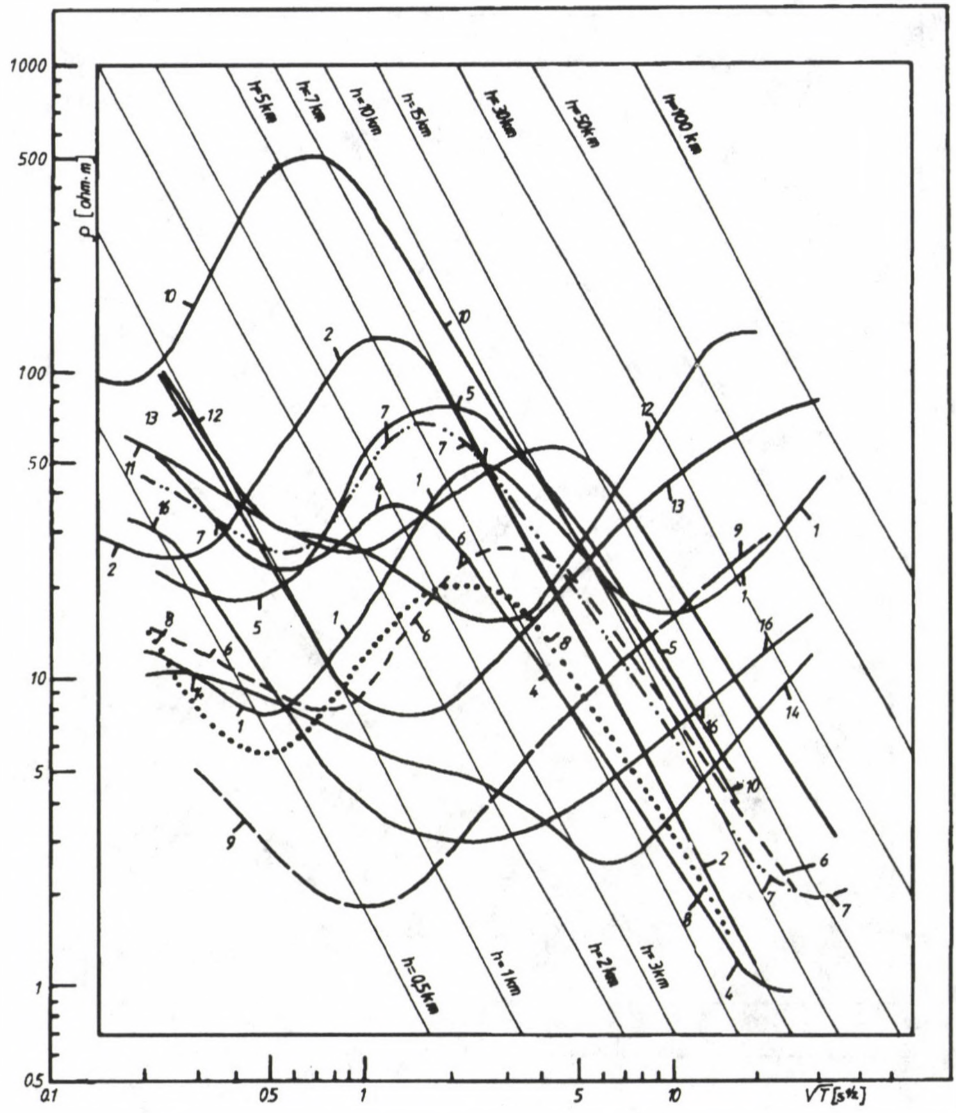


Fig. 3. MTS from all 15 stations along the I-I' profile (E-polarization mode)

The apparent resistivity cross-section of the E-polarization (Fig. 5b) and the expected geological section (Fig. 5c) led to a delimitation between soundings 6 and 9 of a conductive zone ($\varrho < 1$ ohm.m) of anticlinal form whose upper part is estimated to be at a depth of 14 km. This zone stands in the immediate vicinity of the "consumption paleoplane" of the Carpathian Flysch Nappes, and we suppose that it represents the main cause which generated the CECA.

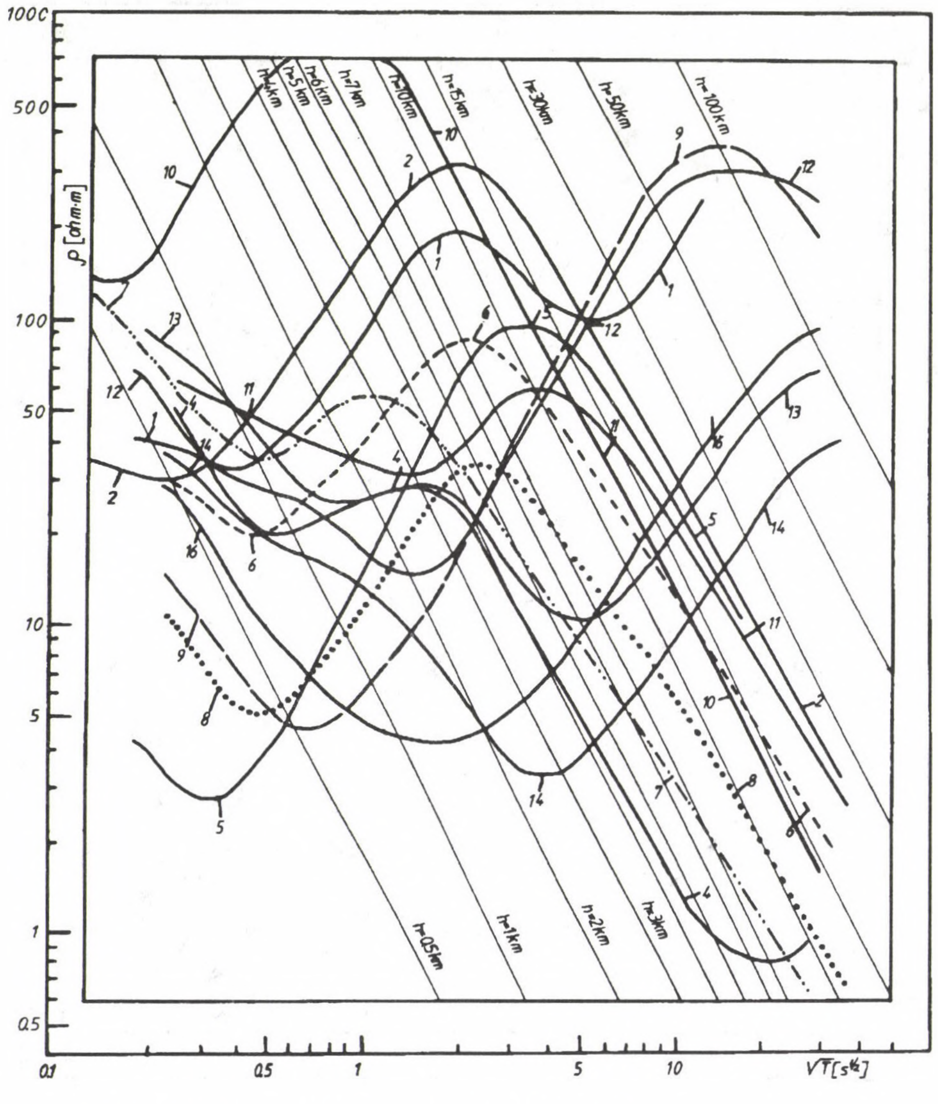


Fig. 4. MTS from all 15 stations along the I-I' profile (B-polarization mode)

Magnetotelluric data (Stănică et al. 1986) also revealed the presence of two types of crust; a thinner one (30–32 km), to the west of the consumption paleoplane, as well as a thicker one (40–45 km) east of it. As to the lithosphere its thickness is also variable: 80–90 km in the west and over 150 km for the East-European Platform underthrust, situated east of the consumption paleoplane.

South of the 46-th parallel, in conformity with the map of induction vectors

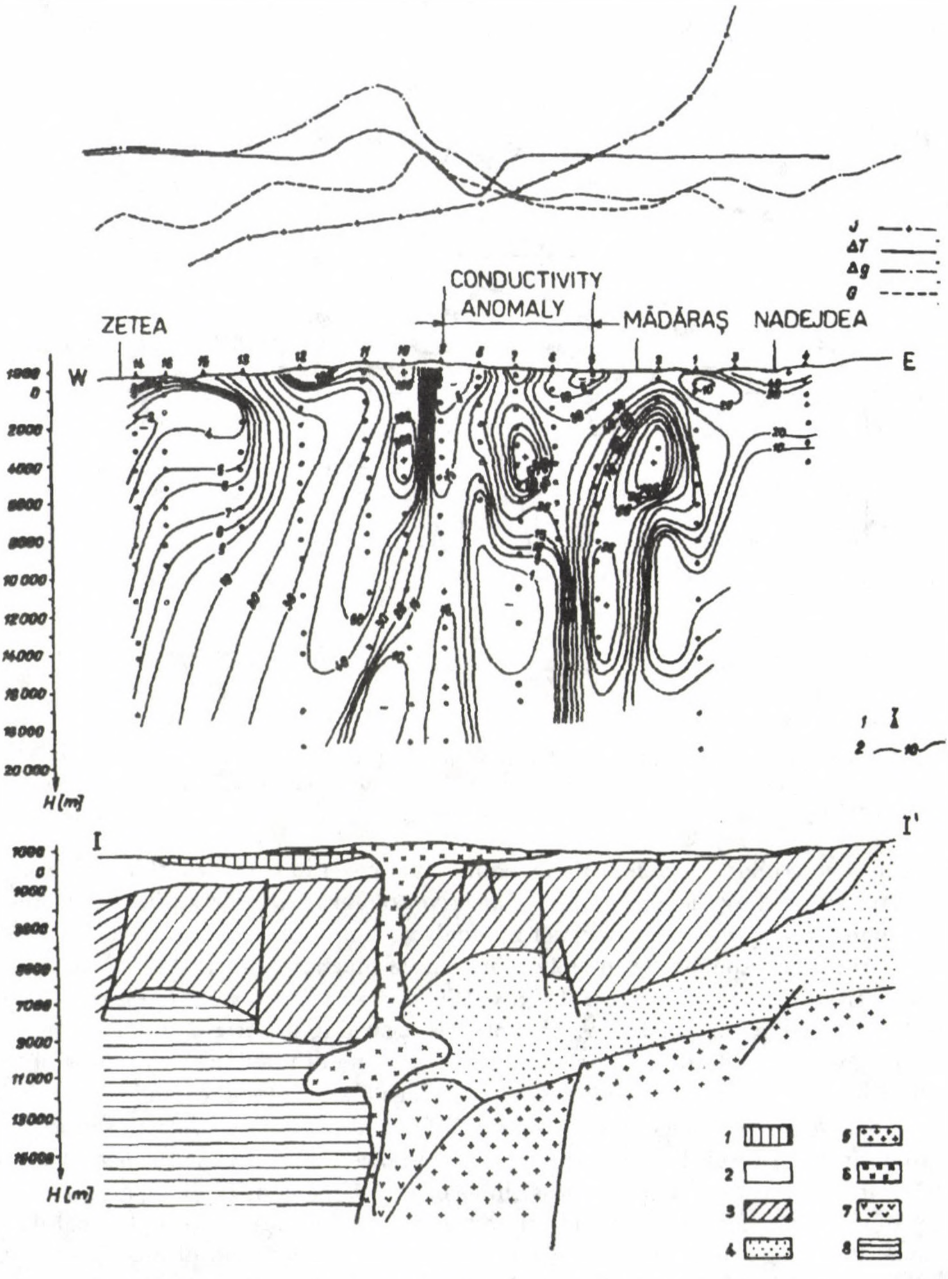


Fig. 5. Magnetotelluric results along the I-I' profile. a) Comparative profiles: J – Telluric parameter, ΔT – Total intensity of the geomagnetic field, Δg – Bouguer gravity anomaly, G – Horizontal gradient of gravity. b) Apparent resistivities (E-polarization mode) plotted as contoured depth cross section: 1 – MTS location, 2 – Isoohm-line. c) Geological cross section: 1 – Volcano-sedimentary formation, 2 – Post-tectonic sedimentary cover, 3 – Central-East Carpathian Nappes system, 4 – Carpathian Flysch Nappes, 5 – East-European underthrust Platform, 6 – Subvolcanic formation, 7 – Oceanic crust, 8 – Socle of the Central-East Carpathian Nappes system

(Fig. 2) the anomaly of electrical conductivity is interrupted and appears again approximately 100 km eastwards; it continues southwards and then westwards along

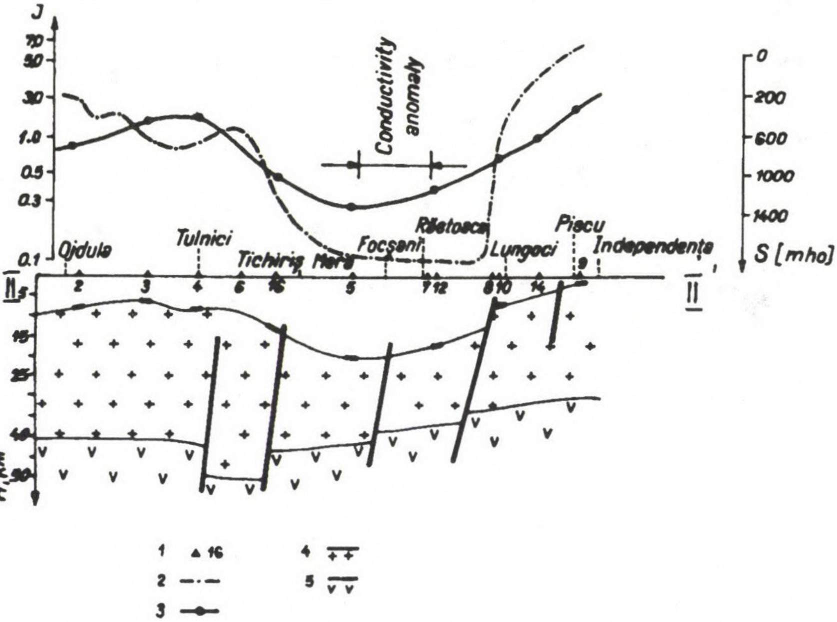


Fig. 6. Magnetotelluric results along the II-II' profile (Stănică and Stănică 1981): 1 – MTS location, 2 – Telluric parameter, 3 – Total longitudinal conductance of sediments, 4 – Crust, 5 – Upper mantle

the axis of maximum thickness of the sedimentary cover in the Carpathian Foredeep, and axis revealed by the high values of effective conductance ($S_{\text{ef}} = 2500$–4200 mho) and minimum values of the Bouguer anomaly.

Information supplied by the interpretation of magnetotelluric data obtained on profiles II-II' (Fig. 6) and III-III' (Fig. 7) (Stănică and Stănică 1981, Visarion et al. 1984), has enabled us to advance a hypothesis regarding the origin of the conductivity anomaly. Since the CECA is situated in the zone of maximum thickness of the sedimentary cover, i.e. 18 km between soundings 5 and 7 (Fig. 6) and respectively 12 km between soundings 31 and 32 (Fig. 7), but also in the vicinity of some areas with a marked ruptural character — e.g. the Vrancea seismic zone (Constantinescu 1978), it could represent the external limit of the collision of the Carpathians with the Moesian Platform. Airinei (1981) advanced on the basis of geophysical data, a hypothesis regarding the existence of a paleo-subduction and of a collision with a slide at the contact zone between the Moesian and Interalpine microplates. According to Savu (1987), the Alpine geological evolution of the Mehedinti Plateau, respectively of the Southern Carpathians Chain, took place in two main stages: in Jurassic-Early Cretaceous and Late Cretaceous. During the first stage an oceanic zone emerged between the Transylvanian microplate and the Moesian one. There-

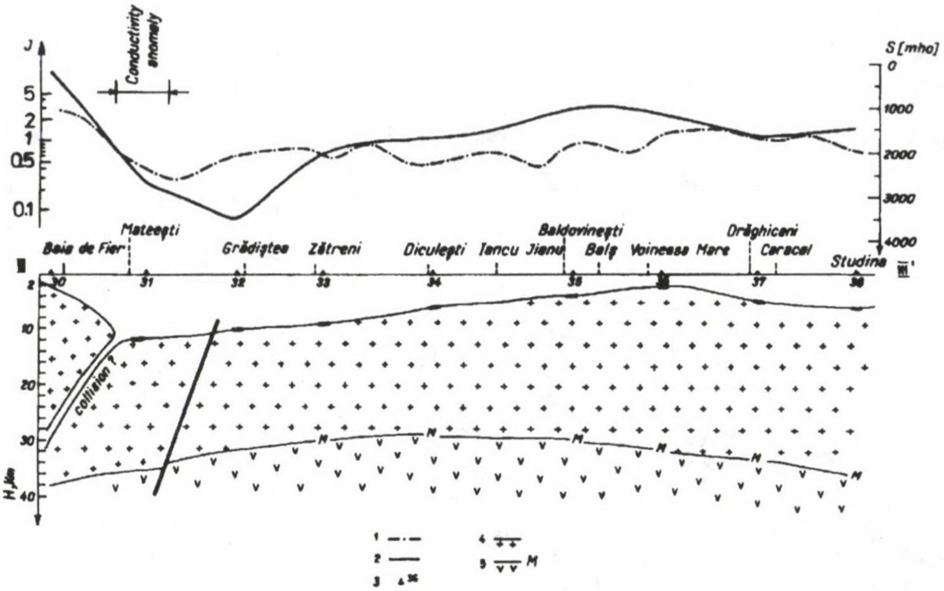


Fig. 7. Magnetotelluric results along the III-III' profile (Visarion et al. 1984): 1 – Telluric parameter, 2 – Total longitudinal conductance of sediments, 3 – MTS location, 4 – Crust, 5 – Upper mantle

fore, during the expansion of the oceanic bottom, a typical oceanic crust was formed (Early and Mid-Jurassic). During the Mid and Late Jurassic the oceanic zone was closed by a process of subduction generating fragments of basic and ultrabasic rocks, elements of crystalline schists, gabbros, basalts and sediments specific for the ocean bottom, hydrosaturated, which may constitute from the point of view of the electrical properties, the source of the anomaly of conductivity in the Southern Carpathians.

The information on CECA has been completed by a study of the distribution of regional magnetic anomalies ΔZa, obtained by filtering the magnetic map ΔZa (Fig. 8). The map thus obtained shows the effect of some bodies situated at depths greater than 10–15 km. These data also indicate a fragmentation of the magnetic isolines into two almost perpendicular segments, corresponding to the external limits of the collision of the Carpathians with the East-European Platform and with the Moesian Platform, respectively.

Conclusions

The CECA in Romanian territory is discontinuous, and the causes which generated it, at least for the one situated south of the 46-th parallel, still belong to the domain of hypotheses which can be accepted or invalidated by future researches of the deep structure.

The primary role among the different mechanisms which could generate the

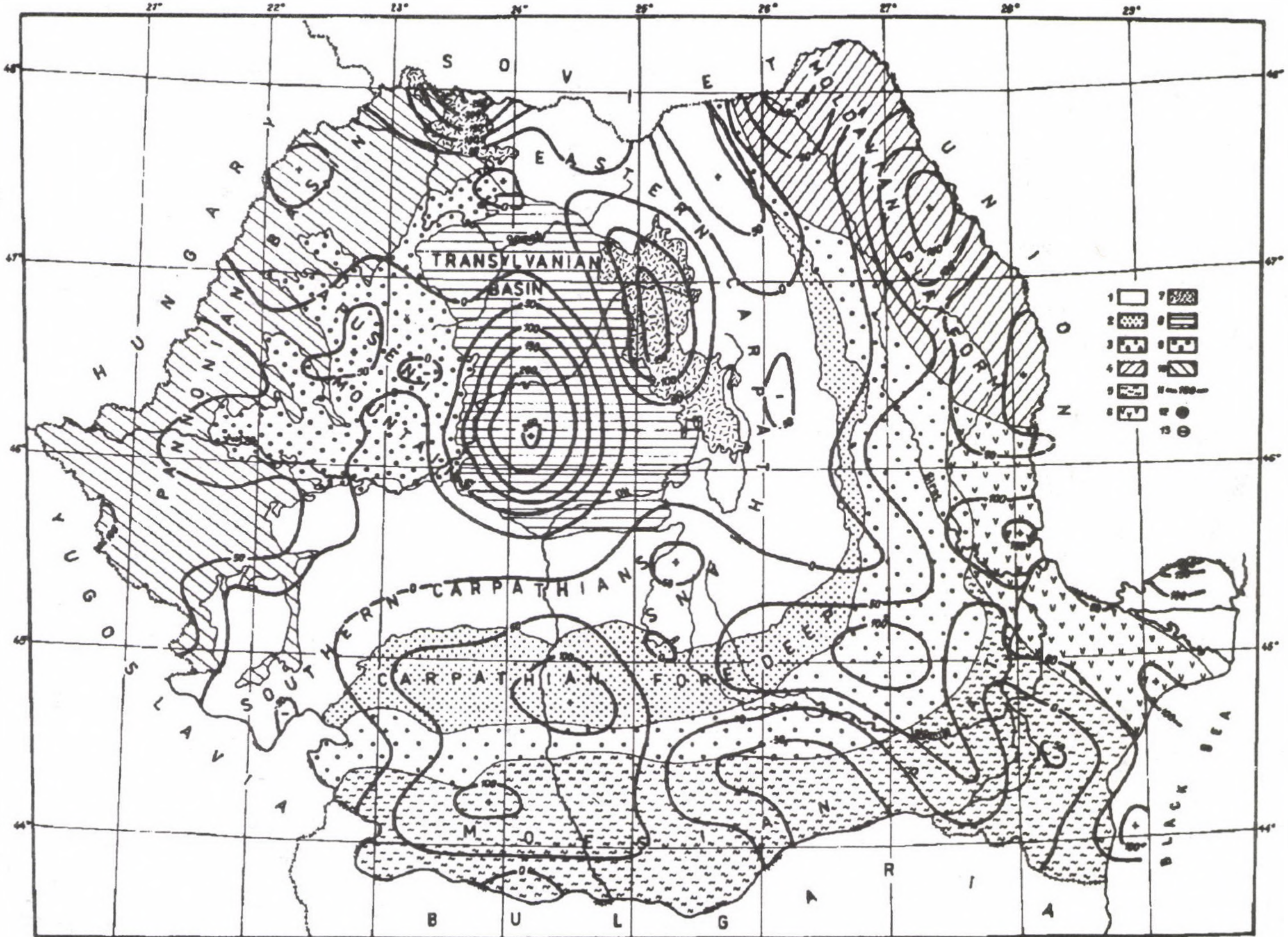


Fig. 8. Magnetic map ΔZa: 1 – Eastern and Southern Carpathians, 2 and 3 – Carpathian Foredeep, 4 – East-European Platform, 5 – Moesian Platform, 6 – North-Dobrogean Orogen, 7 – Volcano-sedimentary formation, 8 – Transylvanian Basin, 9 – Apuseni Mountains, 10 – Pannonian Basin, 11 – ΔZa isoline, 12 – Maximum anomaly of ΔZa, 13 – Minimum anomaly of ΔZa

CECA is played by ruptural tectonics (consumption paleoplanes, sutures, etc.) at the limit between the Carpathians and the neighbouring platforms (East-European to the north-east and Moesian to the east and south).

The geophysical data quantitatively interpreted converge toward a localization of the source of the anomalous conductivity at depths between 10–14 km for the Eastern Carpathians, 18 km in the zone of the Carpathian arc bend and between 12–14 km in the Southern Carpathians.

Acnowledgements

The authors wish to acknowledge the guidance given by Professors A Ádám and O Praus after their reading of the first version of this manuscript, and thank for their constructiv remarks, as reviewers.

References

Ádám A 1974: *Acta Geol. Hung.*, 18, 13–22.

Ádám A 1976: *Acta Geod. Geoph. Mont. Hung.*, 11, 227–285.

Ádám A, Pospíšil L 1984: *Acta Geod. Geoph. Mont. Hung.*, 19, 19–34.

Airinei S 1981: *Rev. Roum. Géol. Géoph.*, 25, 75–81.

Calotă F, Ionescu J, Soare A, Steflea V 1971: *Stud. Cercet. Geol. Geofiz., Seria Geofiz.*, 9, 323–331.

Camfield P A, Gupta J C, Jones A G, Kurtz R D, Krentz D H, Ostrowski J A, Craven J 1989: *Can. J. E. Sci.*, 26, 2385–2395.

Constantinescu L 1978: An extreme Roumanian earthquake and its wider geonomical setting. *Rev. Roum. Géol. Géoph.*, 23, 179–206.

Jankovski J, Petr V, Pěčová J, Praus O 1977: *Acta Geol. Hung.*, 21, 337–351.

Ney R 1975: *Mat. i Prace Inst. Geof.*, 82, 95–110.

Praus O, Pěčová J, Petr V 1980: Present state of the research of the Carpathian geoelectrical anomaly (in Czech.). In: Proceedings of "The 7-th Conference of Geophysicists", Gottwaldov, Sess. S6-complex geophysical interpretations and syntheses. Geofyzika n.p., Brno

Rădulescu D, Săndulescu M 1973: *Tectonophysics*, 16, 155–161.

Ritter E 1970: *Acta Geod. Geoph. Mont. Hung.*, 5, 87–93.

Rokityansky I I, Kulik S N, Logvinov L M, Shuman V N 1975: *Acta Geod. Geoph. Mont. Hung.*, 10, 277–286.

Savu N 1985: *D.S. Inst. Geol. Geofiz.*, 69, 57–71.

Săndulescu M, Visarion M 1979: *Studii Cercet. Geol. Geof. Geogr., Geofizica*, 17, 191–201.

Soare A, Cucu G, Alexandrescu M 1989: *Rev. Roum. Géol. Géoph. Géogr., Géophysique*, 33, 49–56.

Stănică M, Stănică D 1981: *Studii Cercet. Geol. Geof. Geogr., Geofizica*, 19, 41–52.

Stănică D, Stănică M, Visarion M 1986: *Revue Roum. Géol. Géoph. Géogr., Géophysique*, 30, 25–36.

Visarion M, Stănică D, Stănică M 1984: *Acta Geod. Geoph. Mont. Hung.*19, 173–184.

Wiese H 1963: *Geofisica Pura e Applicata*, 56, 101–114.

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 47–64 (1992)

TECTONIC AND GEOELECTRICAL FEATURES OF THE PERIADRIATIC-LINEAMENT (S-AUSTRIA) WITH A GENERALIZATION

A ÁDÁM[1], G DUMA[2], J BERGER[2], J HORVÁTH[1]

A tectonic graben with steep fault at its northern border has been traced by electrically conducting formations along the Periadriatic-lineament in Gail valley and in the Karawanken.

The conducting formations consist of strongly anisotropic graphitic black shales (schists) the high carbon content of which was concentrated by tectonism to the narrow area of the Drauzug Geologic Unit (collision zone).

The low viscosity graphite helped the tectonic movement (escape structure, strike slipe, etc.) along the lineament. The direction of the tectonic forces can be derived from the electric anisotropy.

Geodynamic models by Horváth et al. (1987) are supported if including the role of low viscosity graphitic schists. Perspectives of graphitic formations (as geochemical barrier) for natural resources will also be mentioned in a general summary on the geophysical-geological phenomena of the Periadriatic lineament.

Keywords: audiomagnetics; Bakony-Drauzug Geological Unit; conducting formations; electrical anisotropy; fracture tectonics; graphitic black shales; Periadriatic lineament

1. Introduction

Ádám et al. (1990) reported about a narrow crustal conductivity anomaly (CA) due to black schists in the Alpine collision zone (so-called BAKONY-DRAUZUG GEOLOGICAL UNIT: BDU) which is bounded by long lineaments (Rába and Balaton lines in Transdanubia/Hungary and Periadriatic and DAV lines between the Eastern and Southern Alps in Austria (Fig. 1)). This zone is characterized by strong current channelling in E-W direction as indicated by the distortions in the horizontal magnetic field variations which have been normalized to those measured in Nagycenk observatory (Western Hungary) to illustrate the anomaly (Fig. 2) (Ádám et al. 1986). After a detailed study of the Transdanubian part of the CA, firstly deep MT soundings were carried out in its continuation in the Alps and recently 3 AMT profiles have been completed each crossing in N-S direction the narrow conducting zone as shown in Fig. 1.

The aim of this paper is to demonstrate tectonic information is given by the audiomagnetotelluric results in the Alpine collision zone by indication of near-surface conducting graphitic formations (black schists).

[1]Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H–9400 Sopron, Csatkai E. u. 6–8, Hungary

[2]Zentralanstalt für Meteorologie und Geodynamik, A–1190 Wien, Hohe Warte 38, Austria

Akadémiai Kiadó, Budapest

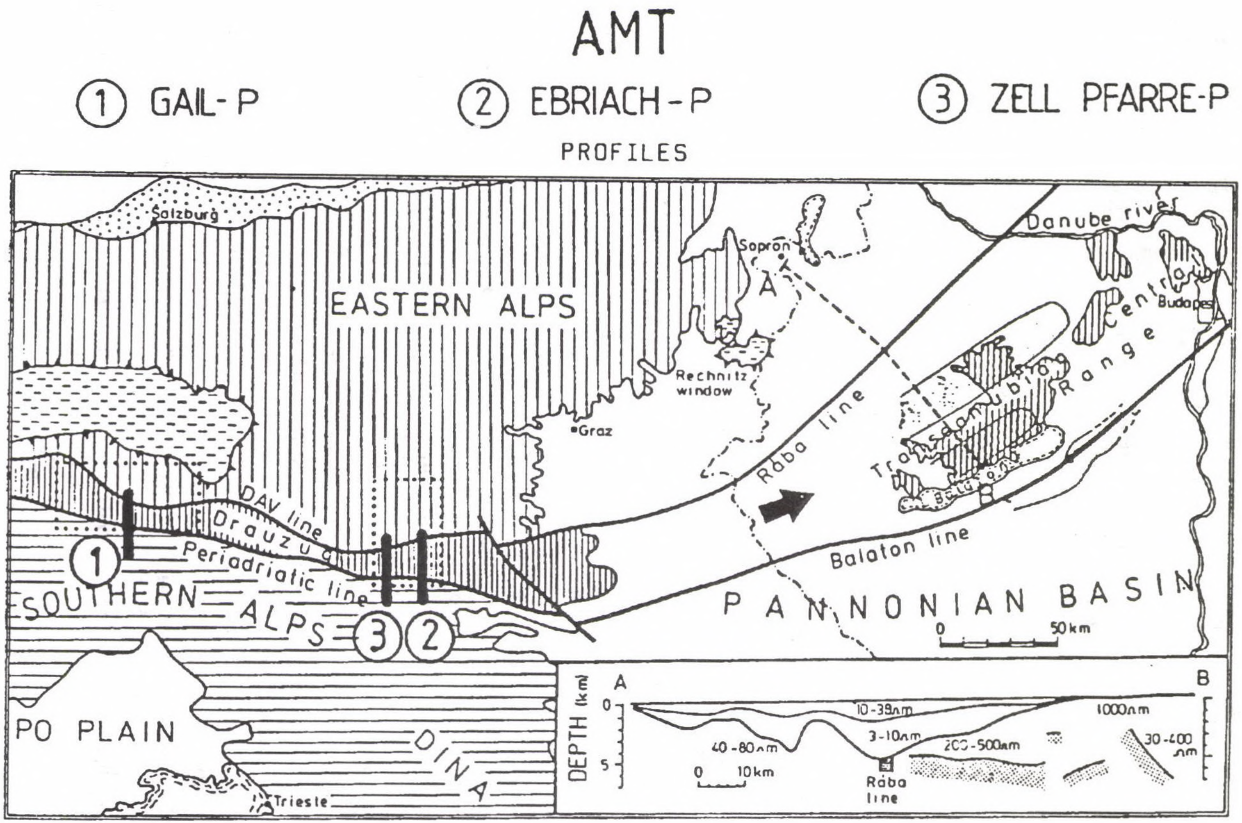


Fig. 1. Tectonic setting of the Bakony-Drauzug Unit (BDU) and high conductivity zones in the Transdanubian Central Range (stippled). Dotted squares indicate the sites of magnetotelluric measurements in the Alps with the AMT profiles (thick lines): 1 Gail profile; 2 Ebriach profile; 3 Zell Pfarre profile. The inset shows the electric structure along the profile A-B. Stippled line gives the position of the conductor (Horváth et al. (1987) and Ádám et al. (1990), respectively)

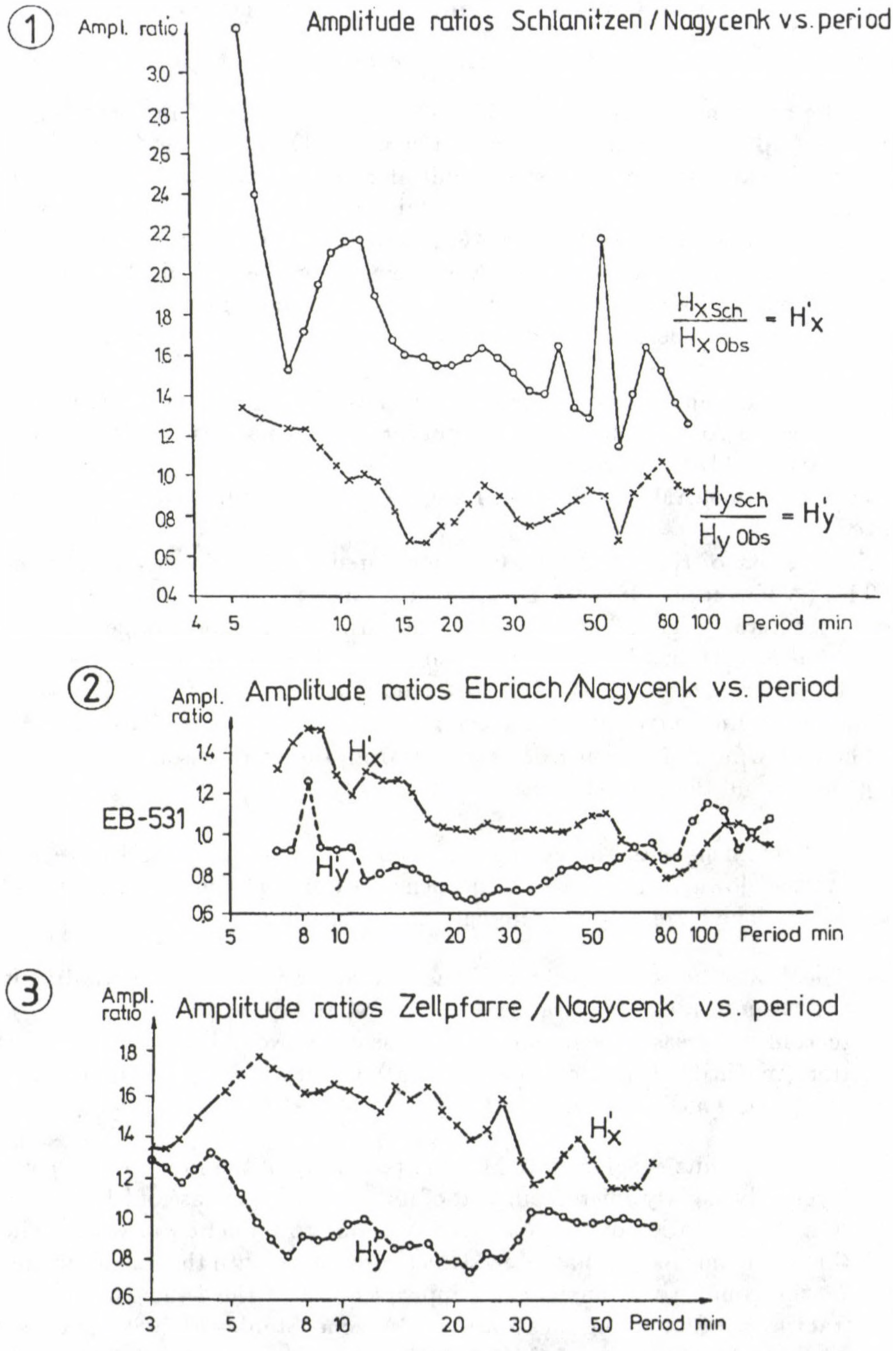


Fig. 2. Amplitude ratios of the time variations of the horizontal magnetic field components measured at stations Schlanitzen (1), Ebriach (2) and Zell Pfarre (3) and Nagycenk (base station) (Ádám et al. 1986)

2. Relation between tectonics and geoelectrical structure

a) Fracture tectonics

In the audiomagnetotelluric profiles No. 1 and 2 measured in 1986 and 1988 in the Alps (Ádám and Duma 1990, Ádám et al. 1990) across the Gail (1) and the Ebriach (2) valleys indicated a steep fault along the Periadriatic lineament by the position of the CA. To study in more detail this fracture tectonics, a new profile was chosen in the Karawanken for AMT soundings where both the geologic and magnetic profiles are known (Seiberl and Steinhauser 1980). This AMT profile No. 3 starts from the Koschuta Hütte, crosses the surroundings of Zell Pfarre, passes near the Hochobir peak and ends in the Sattnitz conglomerate Mts at the Drau river.

The geologic and magnetic profiles are shown in Fig. 3. The magnetic blocks of the so-called Altkristallin have been compressed into a narrow zone between the Periadriatic and Drau lineaments. Under the Periadriatic lineament — in agreement with the results of AMT profiles No. 1 and 2 — the crystalline basement is strongly faulted.

The results of the AMT soundings measured at 12 frequencies between 4.1–2300 Hz (Ádám et al. 1988) are displayed on a quasi-profile (Fig. 4) and on different parameter maps (Figs 5, 6 and 7) as the measuring sites lie on a larger area around the profile No. 3. In Fig. 4 the AMT sites have been projected to a N-S profile with the depth and resistivity of the detected CA as derived from 1-D inversion of Rhomax sounding curves directed mostly in the strike direction (around E-W).

The following conclusions can be drawn from the comparison of the AMT and the geologic and the magnetic profiles:

— In agreement with the geologic profile, the conducting blocks are strongly faulted down by more than 2 km beneath the Periadriatic lineament (Naht-zone). The corresponding anomaly disappears here.

— The lowest resistivity values of the CA appear along the Periadriatic and the Drau (Dráva) deep fractures (0.1 to 64 and 30 to 150 ohmm, resp.). The tectonics decreases the resistivity in these cases twofold: beside the carbonization (maturity of the biological material), the crack and pore volume has been increased which is possibly filled by electrolytes.

— In the Sattnitz conglomerate Mts the resistivity of the more conductive third layer only slightly differs from that of its overburden in case of Rhomax curves. A weak indication of an apparent CA is due to induction or lateral effect in this area; this means that — in strong correlation with the magnetic anomaly — the conductivity anomaly disappears north of the Drau lineament (deep fracture). Concerning this narrow CA zone Stanley's (1989) statement is confirmed. *"This type of CA is a long, narrow zone in which high grade carbon occurs, having been concentrated by tectonism ... between two converging crustal components."*

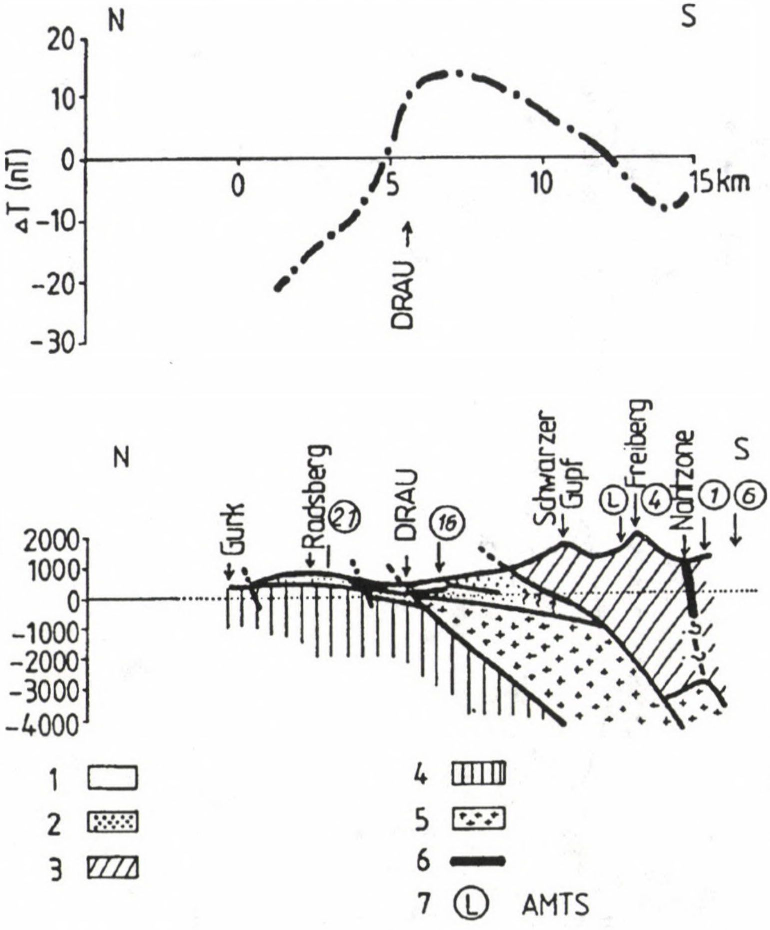


Fig. 3. The results of magnetic measurements and the corresponding N-S geologic profile (see the names), ΔT = anomaly of the total magnetic field (after Seiberl and Steinhauser 1980). 1 - Holocene terrace gravel, 2 - Sattnitz conglomerate, 3 - Mesozoic rocks of the Karawanken, 4 - "Altkristallin", 5 - Magnetic Altkristallin, 6 - Rock sequence of the Periadriatic lineament (Nahtzone), 7 - AMT site

The mentioned peculiarities are confirmed and completed by the areal variations of the geoelectric parameters shown in Figs 5, 6 and 7.

In Figure 5 the Rhomax directions are displayed with the depth values of the CA calculated by 1-D inversions of the Rhomax and the Rhomin extreme curves.

The Rhomax directions are median values, similarly to those parameters chosen for illustration in Figs 6 and 7. The special Austrian railway frequency (: 16 2/3 Hz) and other noises originating from the short-circuit and overloading of the electrical network of 50 Hz strongly disturb the AMT measurements causing a lot of false values in their nearfield zone. Therefore the median values of the on-line measured

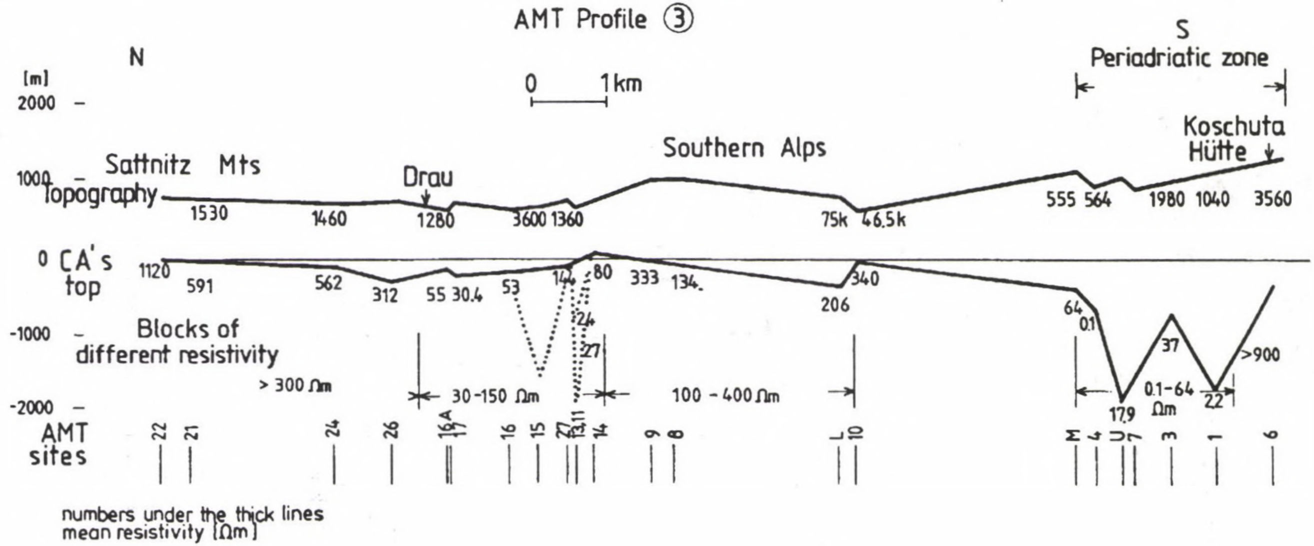


Fig. 4. AMT profile No. 3 with topography, top of the conducting formation and its resistivity calculated from Rhomax curves as well as the resistivity of its overburden

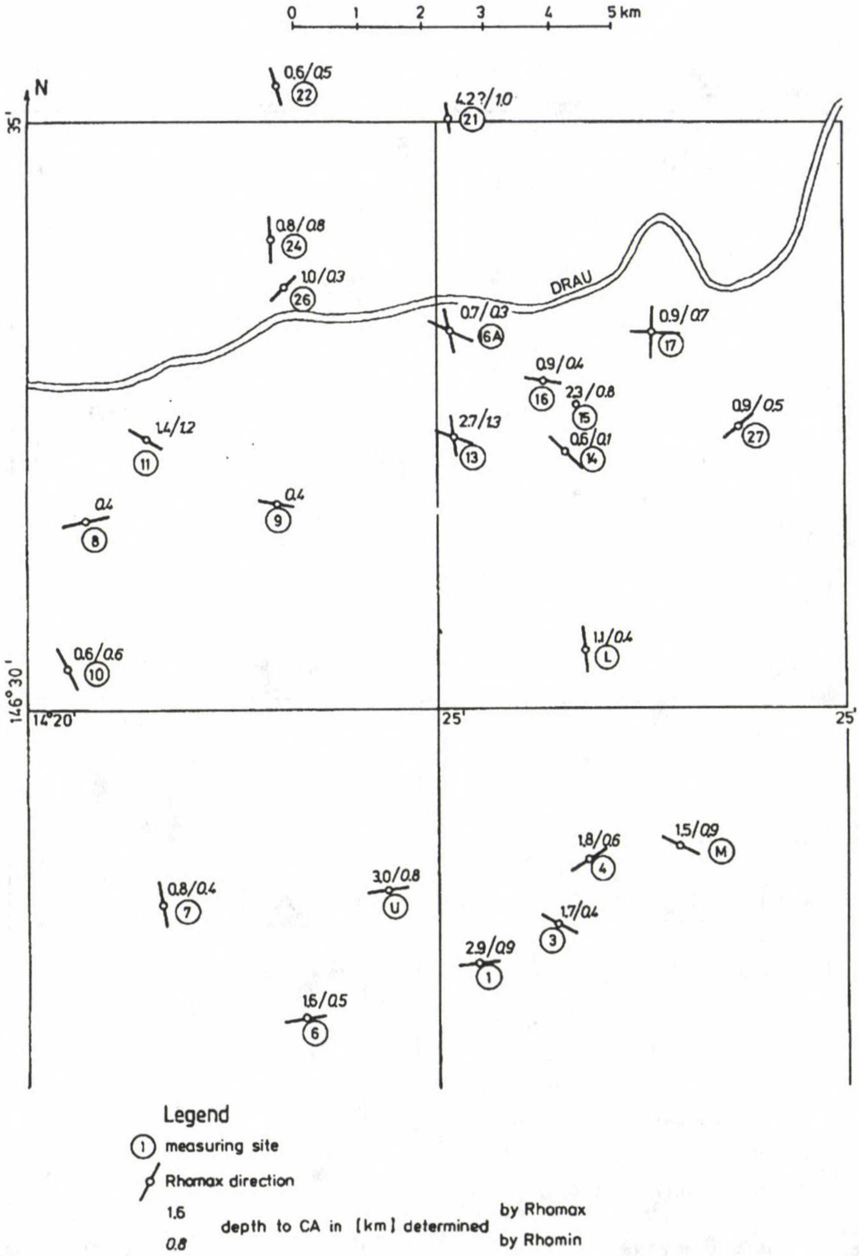


Fig. 5. Map of the average Rhomax directions (and rotation of the Rhomax direction in a few points given by its direction limits) and depth to the CA calculated by 1-D inversion of Rhomax and Rhomin AMT sounding curves

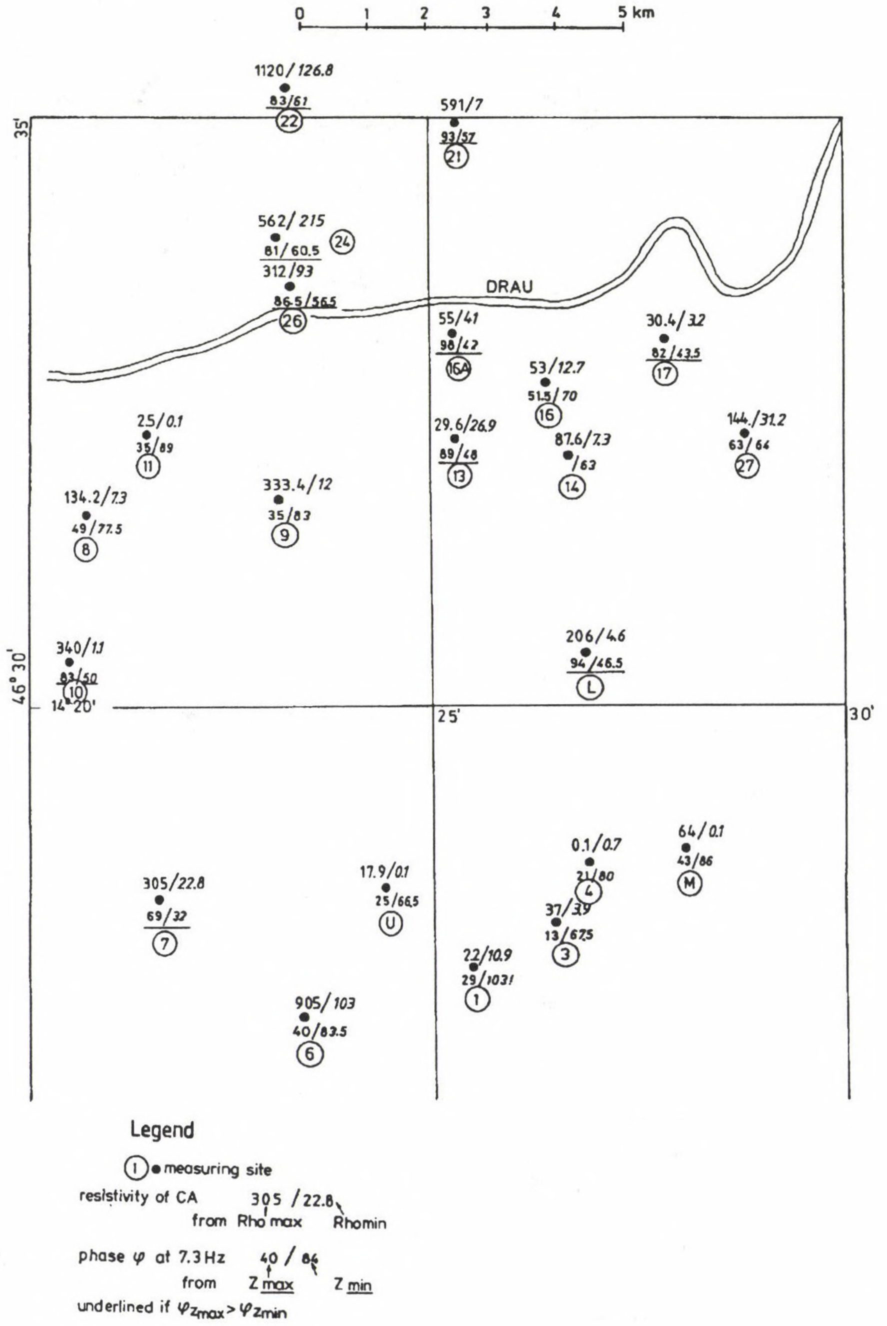


Fig. 6. Map of the resistivity of the CA and phase values at 7.3 Hz

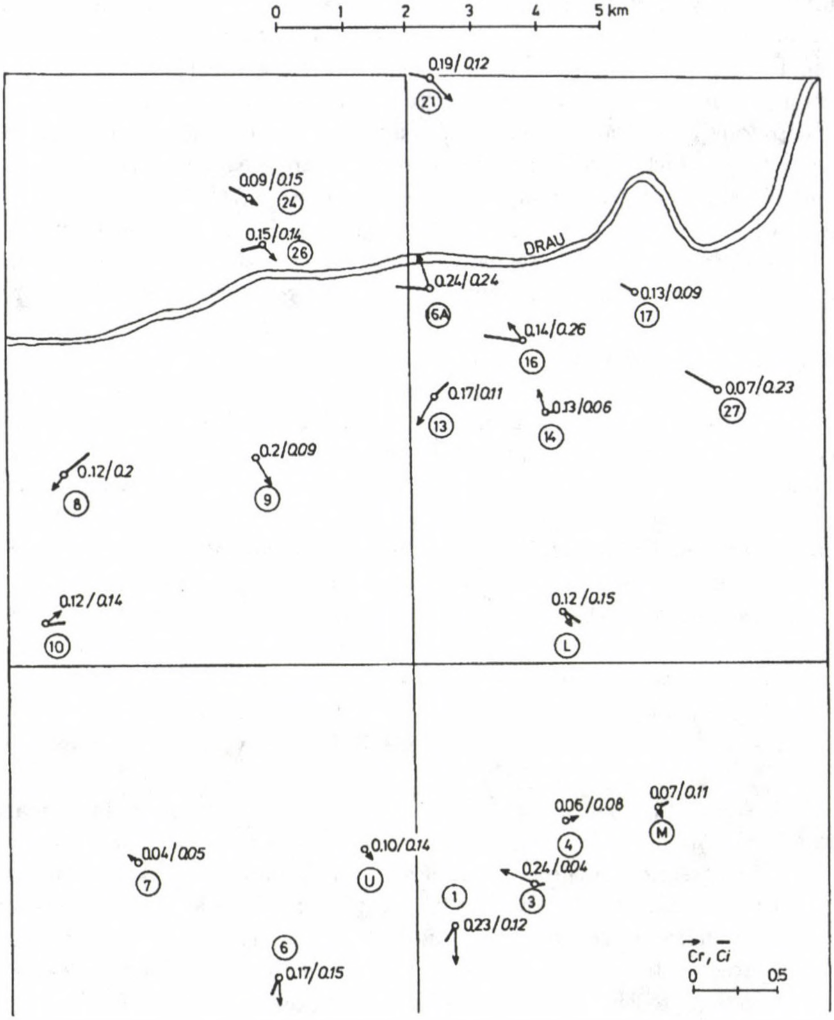


Fig. 7. Map of the real (C_r) and imaginary (C_i) induction vectors at 7.3 Hz

parameter, as a kind of robust estimation can help to find the more reliable values. These noise effects have been (Ádám et al. 1989) and will be more rigorously studied in the future.

In Fig. 5 two main directions can be separated: one lies in ± 10–$15°$ near E-W, the other near N-S direction. The former approximates the strike direction of the Karawanken, the latter one is perpendicular to it and at the same time to the strike of the faults and fractures of the "Altkristallin" blocks, first of all, to the Periadriatic and Drau lineaments. The depths to the top of the CA calculated from Rhomin curves are generally lower, so if they are used in AMT sites lying above the fault planes with N-S Rhomax direction, the height of the fault is greater. The great crystalline block has been lifted beneath "Schwarzer Gupf" (see geological profile in Fig. 3) by the faults along the Periadriatic and Drau line.

AMT sites in the middle part of the area studied i.e. No. 1, U, 13, 21 represent a N-S deep zone according to depths of conducting basement measured here. This zone divides the block of E-W strike into western and eastern parts.

Figure 6 illustrates the changes of 4 parameters in this area. The first two parameters are the resistivities of the CA calculated from Rhomax and Rhomin curves by 1-D inversions. These values support with their areal distribution the conclusion drawn from the AMT profile: the lowest resistivity values are connected to the lineaments.

It is interesting to note the special behaviour of the phase (φ) values at 7.3 Hz (in the second row) depending on the Rhomax direction. The difference between $\varphi_{z_{max}}$ and $\varphi_{z_{min}}$ is negative in case if the Rhomax direction is near E-W i.e. the strike of the Karawanken and positive in case if the Rhomax direction is perpendicular to the strike, i.e. the AMT site lies above the fault plane. That means that the lower phase value appears always in about E-W direction. Its cause has to be studied by numerical modelling, nevertheless, it should be noted that the phase values have been distorted much more than the Rho values.

Figure 7 shows both the real (C_r) and imaginary (C_i) parts of the induction vectors at 7.3 Hz. Their lengths are given numerically, too. The directions of disturbed vectors are not shown due to their uncertainty.

The smallest C_r and C_i values have been observed along the Periadriatic lineament (at sites 4, 7 U, M). In these points a definite vector direction cannot be given.

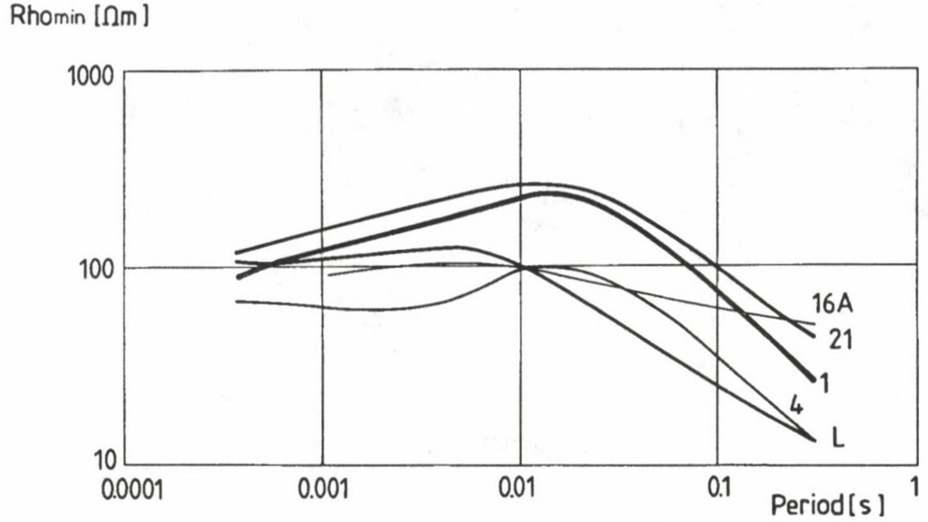
On both sides of the Drau river the C_r real induction vectors point towards the conducting fracture of the Drau lineament i.e. there is a rotation of 180° here in their direction. The direction of the imaginary vectors C_i, nevertheless, does not change along the profile crossing the Drau. Southward from this Drau zone, the geoelectric conditions are characterized by a systematic change in the C_r direction. The C_r vectors are directed towards the conducting zone of the Periadriatic lineament.

In Figs 8a and b Rhomax AMT sounding curves are shown from AMT sites where a definite structural change is reflected by the measurements i.e. from points No. 1, 4, L, 16A and 21.

The scatter of the resistivities measured by an on-line AMT instrument (Ádám et al. 1988) is shown in Fig. 8c. At some frequencies e.g. at 13 Hz, 410 Hz the near-field distortion is too high.

b) Relation between the micro- and macroanisotropy of the AMT resistivities and tectonics

In the western part of the BDU and Gail valley (No. 1 AMT profile), the conducting black shales (graphitic schists) crop out with a very steep N-S dip and are arranged in narrow E-W conduction zones. These zones have been indicated by the high frequency airborne EM measurements of shallow penetration in agreement



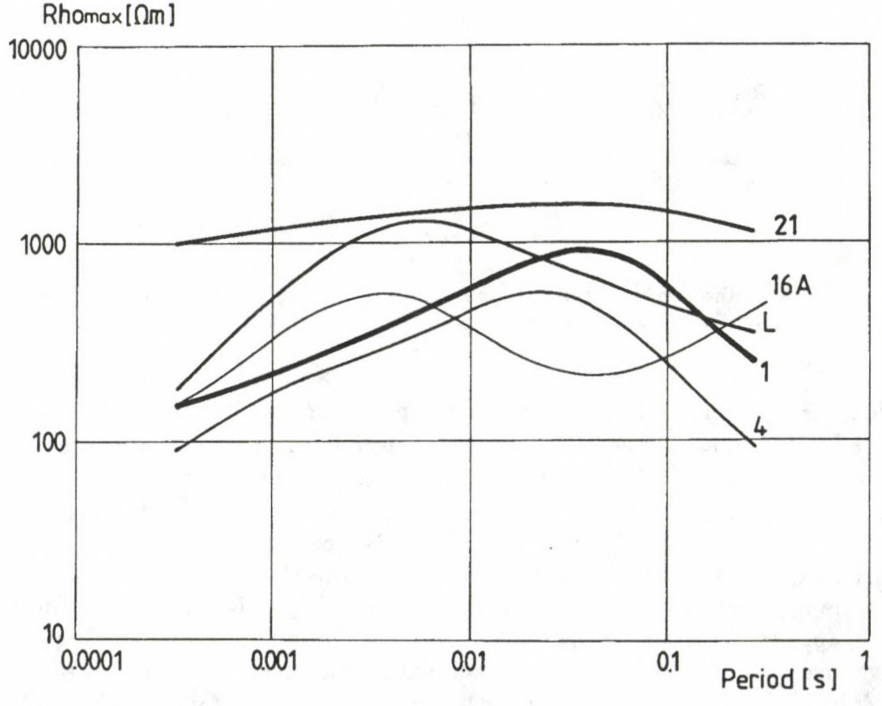


Fig. 8. a, b. AMT sounding curves measured at AMT sites (1, 4, L, 16A, 21) characterizing the change in the block structure

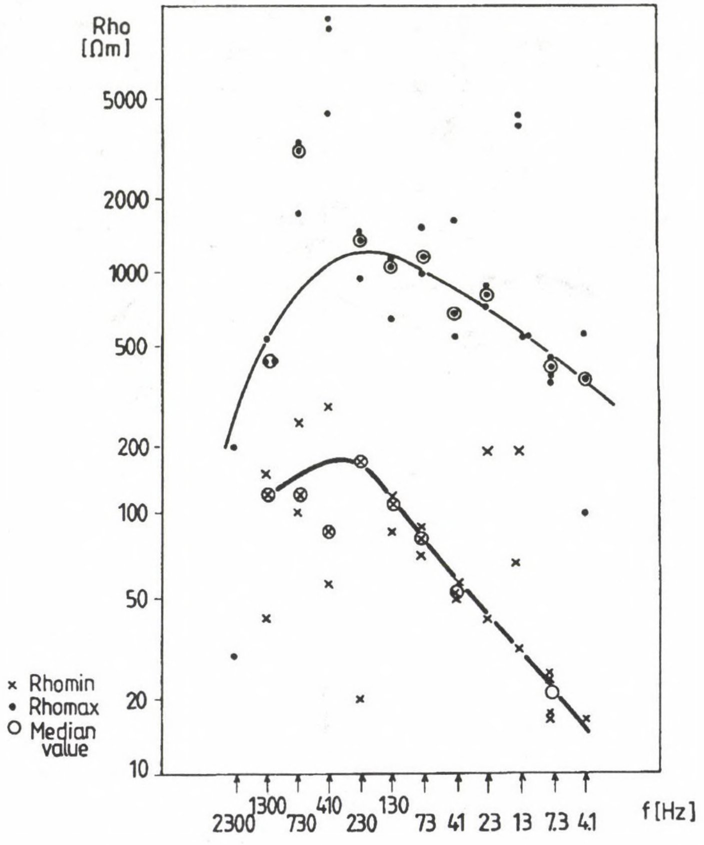


Fig. 8. c. AMT Rhomax and Rhomin sounding curves of AMT site L showing the data scatter obtained by an on-line AMT instrument (Ádám et al. 1988)

with the results of AMT soundings (Duma pers. communication).

The anisotropy direction in the black shales due to the schistosity (microanisotropy) is N-S (x) and E-W (y). As the strike of the 2-D narrow conducting zones is also E-W, the macroanisotropy cannot be separated from the intrinsic i.e. microanisotropy. Both kinds of anisotropy can be deduced from the effect of northward collision forces, as is shown in Fig. 9a after Kázmér and Kovács (1985) together with the conducting block structure of the Gail valley (Fig. 9b) and the measured $\lambda = \varrho_x/\varrho_y$ simple anisotropy (Fig. 9c). The AMT anisotropy with the higher resistivity measured in N-S direction (x) can be the consequence of the N-S direction of the collision forces due to the northward movement of the Adriatic promontory of the African plate.

To demonstrate the great intrinsic or microanisotropy of the graphitic schists along the Periadriatic lineament AMT sounding curves (with their 1-D layer sequences) are shown as examples from the region of the Ebriach valley (No. 2 AMT

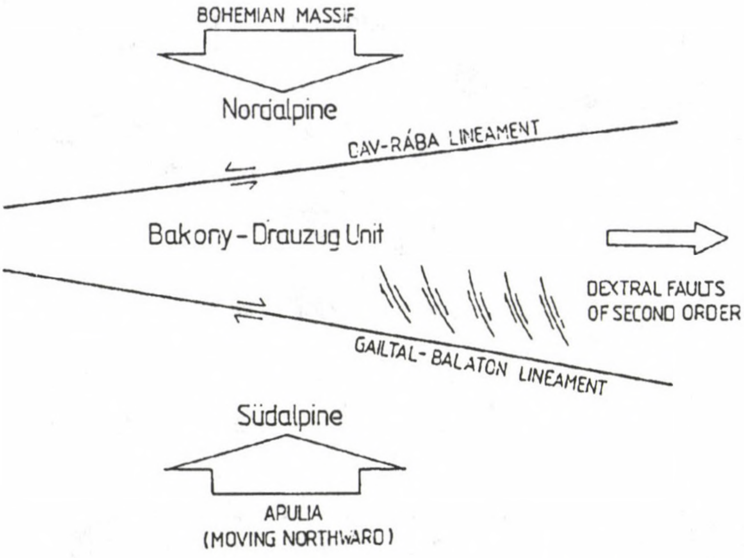


Fig. 9. a. Forces in the Alpine collision zone after Kázmér and Kovács (1985)

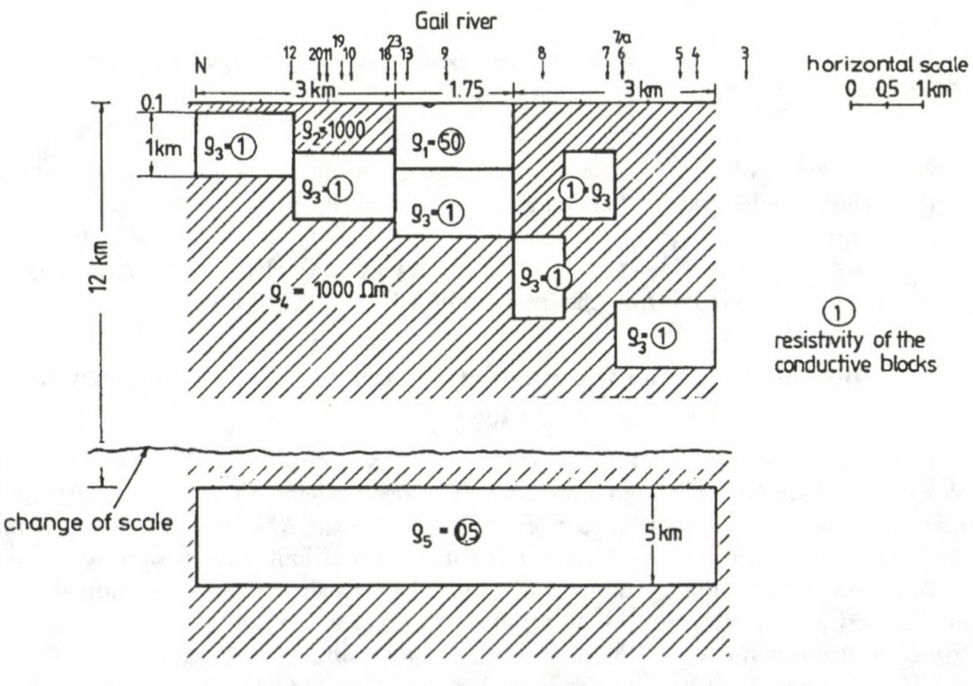


Fig. 9. b. 2-D block model of the graphitic schists in the Gail valley. The cause of anomaly at depth of 10–20 km can be ionic conduction

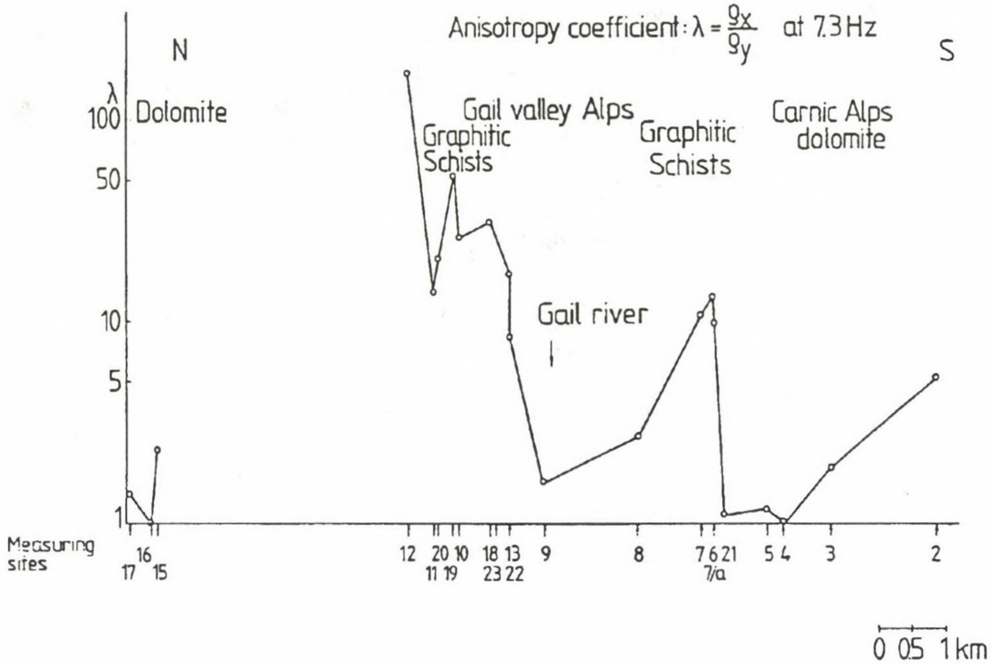


Fig. 9. c. Great northward (x) directed resistivity increase i.e. $\lambda = \varphi_x/\varphi_y \gg 1$ in connection with the collision forces (see Fig. 9a)

profile). They have been measured on a Paleozoic sequence (Fig. 10) according to Ucik's (1968) geologic map. This anisotropy reflects the effect of the same collision force.

In the AMT profile No. 3, the macroanisotropy due to electric inhomogeneity is determinant in points lying on the great faults. Here also $\varrho_x > \varrho_y$.

3. Differences between the conducting structures in the eastern and western part of the BDU

As it was shown, the conducting graphitic blocks lie near the surface in narrow E-W zones and are faulted up and down by only some kms beneath the Periadriatic and Drau lineaments in the western part of the BDU along the AMT profiles Nos 1, 2 and 3. These deep fractures may have conducting roots of ionic conduction according to the deep MT sounding (Ádám et al. 1986). Nevertheless this supposition should be checked as it is very difficult to separate the real indication of deep conductors from the induction effect of the near surface conducting formations.

The narrow conducting zones have been well pointed out by the AMT and magnetic anomalies.

In the Eastern part of the BDU lies the Transdanubian CA between the Rába and

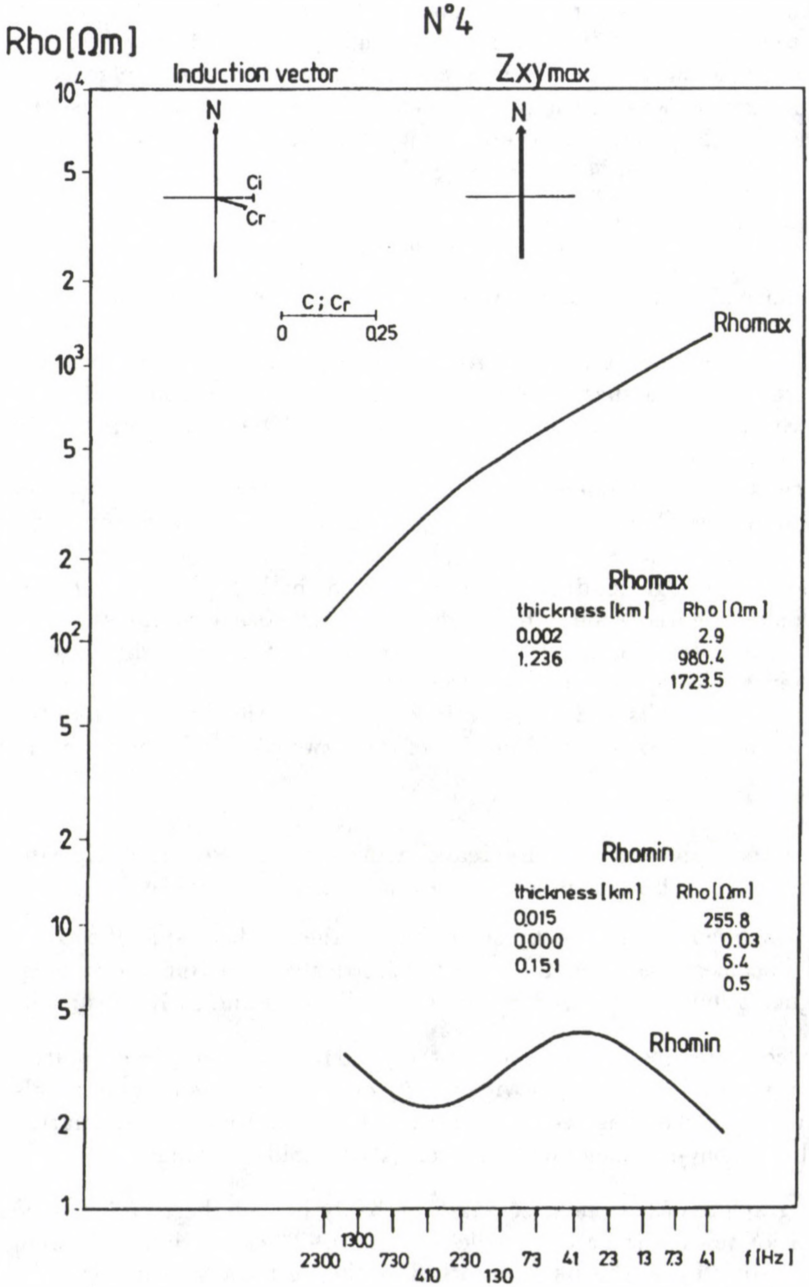


Fig. 10. Extreme values of AMT resistivity *vs.* frequency; the direction of the impedance maximum ($Z_{xy\max}$) and the direction of real (C_r) and imaginary (C_i) parts of induction vector at 7.3 Hz measured at AMT point 4 along the No. 2 AMT profile (Paleozoic schist). The layer sequences fitted to the measured resistivity values are also given

Balaton lines in a much larger area as shown in Fig. 1. The conducting zone widens eastward together with BDU not only horizontally, but vertically too, dropping from a depth of about 6 km to deeper than 10 km on its southern part.

It is questionable what is the tectonic cause of this increase in dimensions of the CA in Transdanubia. One has to add to the lateral extension a rotation of about 35° of the structure towards (E)NE-(W)SW.

4. Tectonic models

As we have already discussed (Ádám et al. 1990), the role of the formation of graphite — a material of low shear strength or viscosity — can also be considered in association with motions during the Alpine orogeny, e.g. in the "escape structure" of BDU according to the speculations of Kázmér and Kovács (1985) (see thick arrow in Fig. 1 which shows the escape direction of Bakony Mountains during the middle Eocene–late Oligocene).

Horváth et al. (1987) following the ideas of Kázmér and Kovács (1985) outlined two tectonic models (Fig. 11). They emphasized the role of low viscosity graphite shales/schists in both models.

In Model 1 the high conductivity layer formed the bottom of the thrust sheet during the late Cretaceous and late Eocene to early Oligocene compressive deformations. In Model 2 the graphite formations occurred in the strike slip zones (dikes) when the renewed compression during the late Eocene and early Oligocene initiated regional strike-slip faults and the mentioned escape of the Bakony Mountains towards East. In our view the combination of these two models is the most probable one.

5. Main tectonic and geophysical features of the Periadriatic-Balaton lineament – a summary and conclusion

— Tectonic graben bordered by steep fault at the northern side of the Periadriatic lineament has been followed by electrical conducting formations using magnetotellurics in NW Transdanubia, Gail valley and in Karawanken, too.

— The conducting formations consist of strongly anisotropic graphitic black shales (schists) containing a wide variety of metallic (magnetic) minerals. The high grade carbon has been concentrated by the tectonism to the narrow area of the Bakony-Drauzug Geologic Unit (BDU, collision zone).

— The graphite of low shear strength probably helped the tectonic movement (collision, escape structure, strike slip, *etc.*). The direction and strength of the tectonic forces can be approached by the electrical anisotropy.

— The conductivity anomaly region has been built up by E-W (in the Alps) or NE-SW (Transdanubia) oriented conductive stripes at different depths (Fig. 1). Each stripe has a sequence of isolated conductive dikes (blocks) (Ádám and Varga 1990) which cannot be separated at some km depth by any

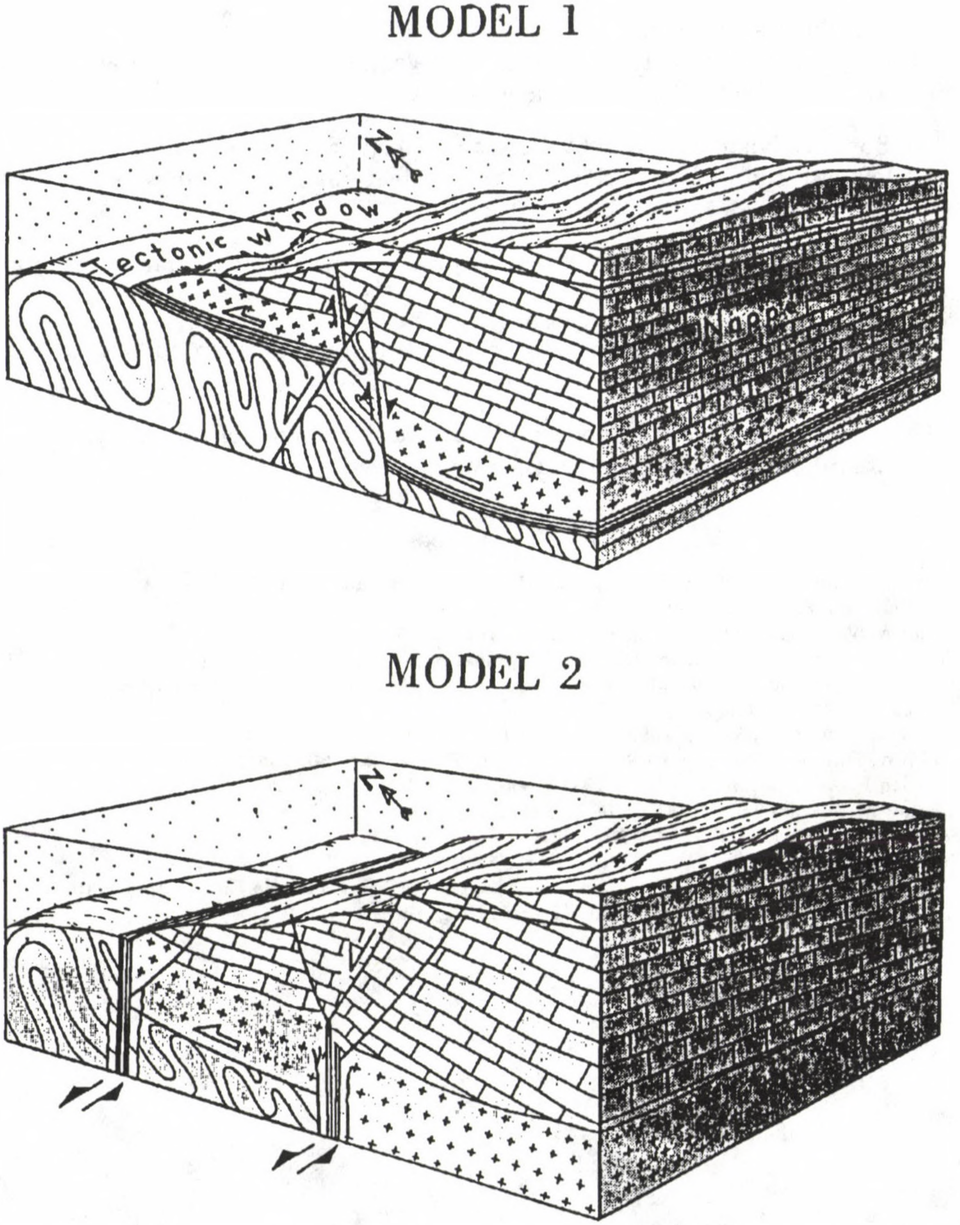


Fig. 11. Tectonic sketch by Horváth et al. (1987) to show two alternative models for structural development of the Transdanubian Central Range (thrust (1) and strike slip (2) model)

geophysical technique (seismic, MT *etc.*), therefore they appear as a subhorizontal layer.

— The dimensions of the BDU are extended in Transdanubia not only horizontally but vertically, too, i.e. there is a general extension in the area of the Trandanubian conductivity anomaly (CA).

— Both geodynamic models by Horváth et al. (1987) may be verified by help of low viscosity graphite schists. The combination of these models is the most probable one.

— Around the lineament there is a narrow magmatite belt (granite, tonalite, diabase, andesite, *etc.*).

— The natural resource perspective of graphitic formations are to be studied as geochemical barrier in ore generation.

— A research borehole proposed to be drilled in the area of the smallest depth (3 km) of the Transdanubian CA (Mór graben).

References

Ádám A, Duma G 1990: In: Mechanics of Jointed and Faulted Rock. Rossmanith H P (ed.) Balkema, Rotterdam, 149–156.

Ádám A, Varga G 1990: *Phys. Earth Planet. Int.*, 60, 80–88.

Ádám A, Duma G, Gutdeutsch R, Verő J, Wallner Á 1986: *Journal of Geophysics*, 59, 103–111.

Ádám A, Tiikkainen J, Horváth J, Hjelt S E, Varga J, Saastamoinen R, Verő J 1988: *Acta Geod. Geoph. Mont. Hung.*, 23, 63–72.

Ádám A, Szendrői J, Verő J 1989: *Gerlands Beitr. Geophysik*, 98, 25–36.

Ádám A, Duma G, Horváth J 1990: *Phys. Earth Planet. Int.*, 60, 155–162.

Horváth F, Ádám A, Stanley W D 1987: *Rend. Soc. Geol. It.*, 9, 123–130.

Kázmér M, Kovács S 1985: *Acta Geol. Hung.*, 18, 71–84.

Seiberl W, Steinhauser P 1980: *Mitt. österr. geol. Ges.*, 71/72, 291–298.

Stanley W D 1989: *Phys. Earth Planet. Int.*, 53, 228–239.

Ucik F H 1968: In: "700 Jahre Markt Eisenkappel" (Marktgemeinde Eisenkappel-Vellach ed.). Verlag des Geschichtsvereines für Kärnten, Klagenfurt, 142–160.

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 65–86 (1992)

A CRITICAL REVIEW OF MAGNETOTELLURIC STUDIES IN DIVERSE TECTONIC AREAS IN ARGENTINA, CHILE AND ANTARCTICA

M A Muñoz[1], H G Fournier[2], M Mamani[2], E Borzotta[2]

Magnetotelluric soundings carried out during the past fifteen years in Argentina along the cratonic region of Buenos Aires, the geothermal area in the provinces of Tucumán and Santiago del Estero and in sites surrounding the Argentinian bases in the Antarctic Peninsula (Marambio and Matienzo) were interpreted through 1D models. An attempt is done presently to review these studies in the frame of 2D models as to better know the resistivity structure of each area and to point out differences between stable and tectonically active regions. Results are exposed concerning thickness of sediments, thermal anomalous layers in the crust and depth to the conducting layer in the mantle (electric asthenosphere), considering the distortion of curves and coast and island effects. Preliminary models of the cratonic area in eastern Argentina are compared with those established for southern Chile in the active transition zone of the Pacific type. In the case of the Antarctic Peninsula in the border of the Weddell Sea the review is mainly intended to determine the morphology of the sedimentary basin and the conducting layer in the zone of the nunataks Foca.

Keywords: Antarctic Peninsula; Argentina; Chile; distortion effects; electric asthenosphere; geothermal fields; magnetotellurics

Introduction

During the past fifteen years magnetotelluric soundings have been carried out in diverse areas of Argentina, including the Antarctic territory. In some cases analysis of data was not optimal and interpretation procedure not complete. Particularly, MT soundings over extended areas were interpreted by means of 1D models and mainly through E-polarization mode curves. Full analysis of both polarization curves and phase diagrams seem nowadays very difficult to be done. In the past few years MT studies were extended to southern Chile, and high distortions of apparent-resistivity curves have been observed. An attempt is exposed here to review MT interpretation on the basis of available data and in the frame of 2D models. Distinct results from very diverse tectonic areas in the continent South America and Antarctica are compared taking into account new MT resistivity models and other geophysical results. For soundings in Argentina 2D modelling was carried out using the finite element program by Wannamaker et al. (1987); for soundings in Chile the programs by Kisak and Silvester (1975) and Wannamaker et al. (1987) were used.

[1]Departamento de Geofísica, Universidad de Chile, Casilla 2777, Santiago, Chile

[2]Instituto de Investigaciones Aplicadas de Ciencias Espaciales, Casilla 131, 5500 Mendoza, Argentina

Akadémiai Kiadó, Budapest

MT soundings in Argentina

2D modelling of MT soundings made in areas 1 and 3 of Fig. 1 will be presented. Area 1 corresponds to a cratonic region of an age of approximately 2 billion years and

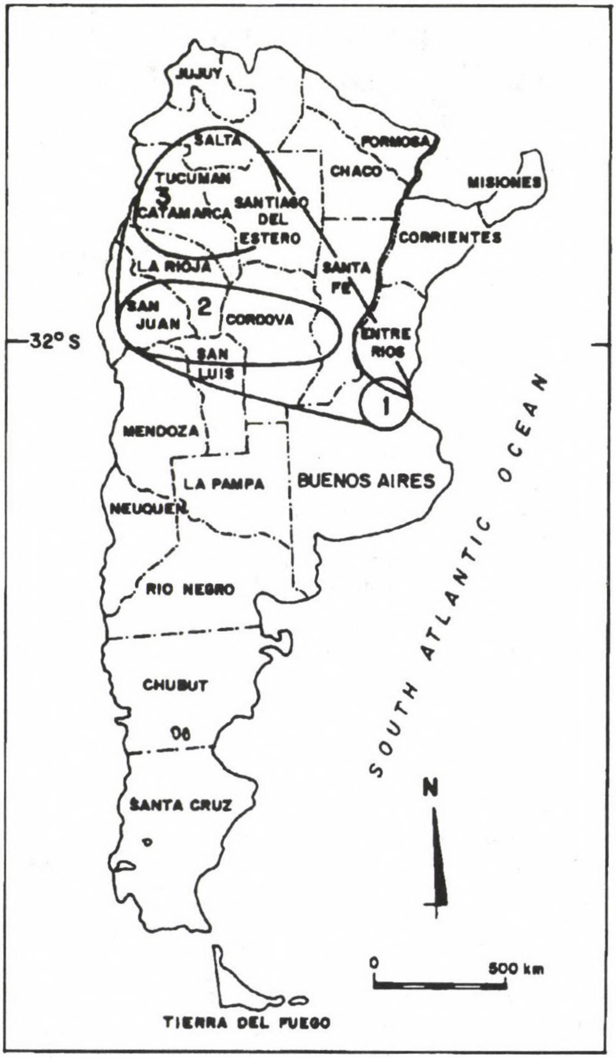


Fig. 1. Map of Argentina. Magnetotelluric soundings in Areas 1 and 3 are reviewed in this paper

area 3 is an anomalous thermal area. Apparent-resistivity curves and 1D modelling and interpretation were presented by Fournier (1981), Febrer (1981) and Baldis et al. (1983, 1985). Also, 2D modelling of MT soundings carried out in the Antarctic Peninsula (nunataks Foca) near the Argentinian scientific base Matienzo will be performed; previous analysis of these soundings are contained in Del Valle et al.

(1983) and Fournier et al. (1989).

Geothermal area of NW Argentina

A cross section along the NS direction of the geothermal area of NW Argentina which approximates the thickness and resistivity layer parameters found in 1D models of the sounding curves is shown in Fig. 2 (Baldis et al. 1983). A hot layer of about 165 km length in the NS direction is encountered in the 1D modelling. The

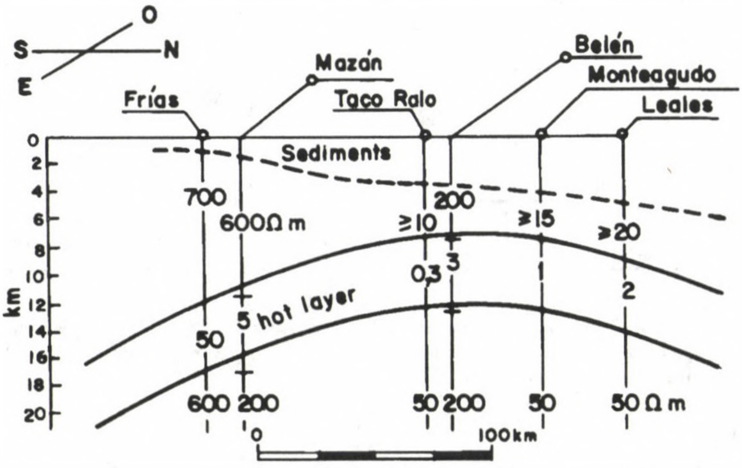


Fig. 2. A cross section of the geothermal area in NW Argentina based upon 1D model results (taken from Baldis et al. 1983)

resistivity of layers down to 12 km is seen to increase towards the south reaching 600–700 ohmm at Frías sounding site. Near thirty variations of model parameters within the 2D model shown in Fig. 3 were performed to compute the response of the E-polarization mode. The resistivity structure of layers are not well determined because the few number of sounding sites do not enable to further constrain the field of model parameters. In any case, a crude fitness of responses of 2D models to observations has been attained (Fig. 4) for F, TR and MA sounding sites; on the contrary, responses of a 2D model synthesizing 1D results as those shown in Fig. 2 are far to fit the observation curves. In both cases the approximate RMS errors are given in Fig. 4. The effect of lateral heterogeneities is observed in all the cases considered, the sections of different resistivity having large influence on each other. 2D modelling cannot fit the observation curves of M and B sounding sites located at about 100–200 km to the west of the NS principal sounding area. 3D effects which can distort the apparent resistivity curves are surely acting in these cases; in some special situations this perturbation can be stronger in the E-polarization mode (Park et al. 1983; Wannamaker et al. 1984). Results for L are marginal. The lower boundary of the 2D model-structure (not shown in Fig. 3) was put at 188 km depth — the last layer having a resistivity of 100 ohmm. As seen in the

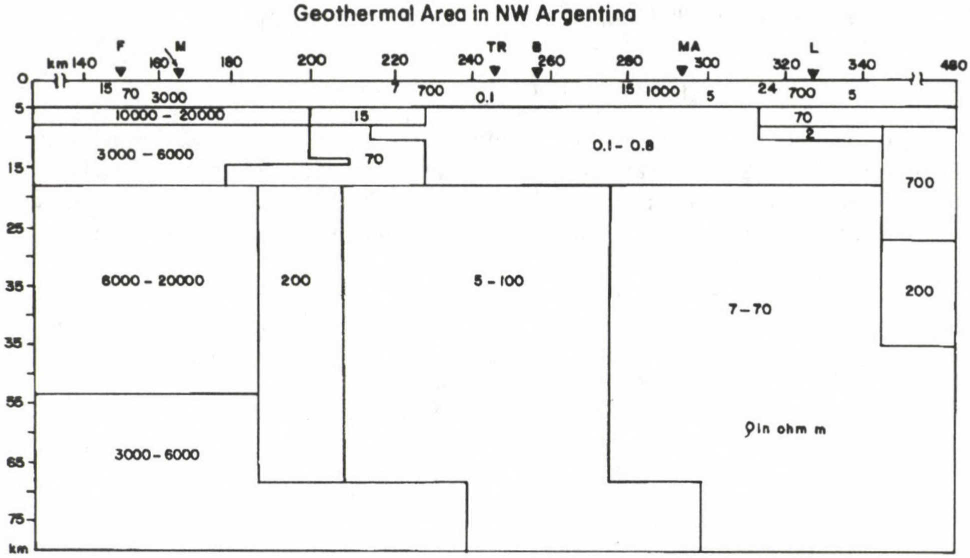


Fig. 3. A 2D model of the geothermal area in NW Argentina (vertical scale is exaggerated). The electrical structure of the shallower part is less defined (stepping numbers give a possible resistivity sequence depending on depth)

2D model (Fig. 3) the hot layer of low resistivity is reduced to an extension of about 110 km. The 14 km maximum thickness of this layer is found to be at Taco Ralo (TR). The top of the layer is at nearly 4 km descending to 10–15 km in the northern zone. The southern zone (Frías sounding site — F) is characterized by a much higher resistivity (3000–20000 ohmm) in the first tens of kilometers than in the 1D model-structure (50–700 ohmm, Fig. 2). The 2D model results are in agreement with geothermal and seismic observations. In the area of low resistivity the temperature gradient in water wells is generally of about 100°C per km; many hot springs are known in this area and thermal stations are built particularly at Taco Ralo (TR). New magnetotelluric soundings northward of this area have confirmed the existence of a geothermal anomaly (Osella et al. 1991).

Seismicity of the area is shown in Fig. 5 (Barazangi and Isacks 1976); a gap in seismic activity in the western border of Argentina has been observed precisely in the sounding site area including the zone from Taco Ralo (TR) to Leales (L). Frías (F) sounding site is located in the border of the zone where seismic activity reinitiates with foci at intermediate depths.

The Buenos Aires Craton

Thickness of layers and resistivity upon 1D results at sounding sites in the Buenos Aires craton are shown in Table I (Baldis et al. 1983, 1985). Only two soundings have been made in the area and a great diversity in tectonic blocks

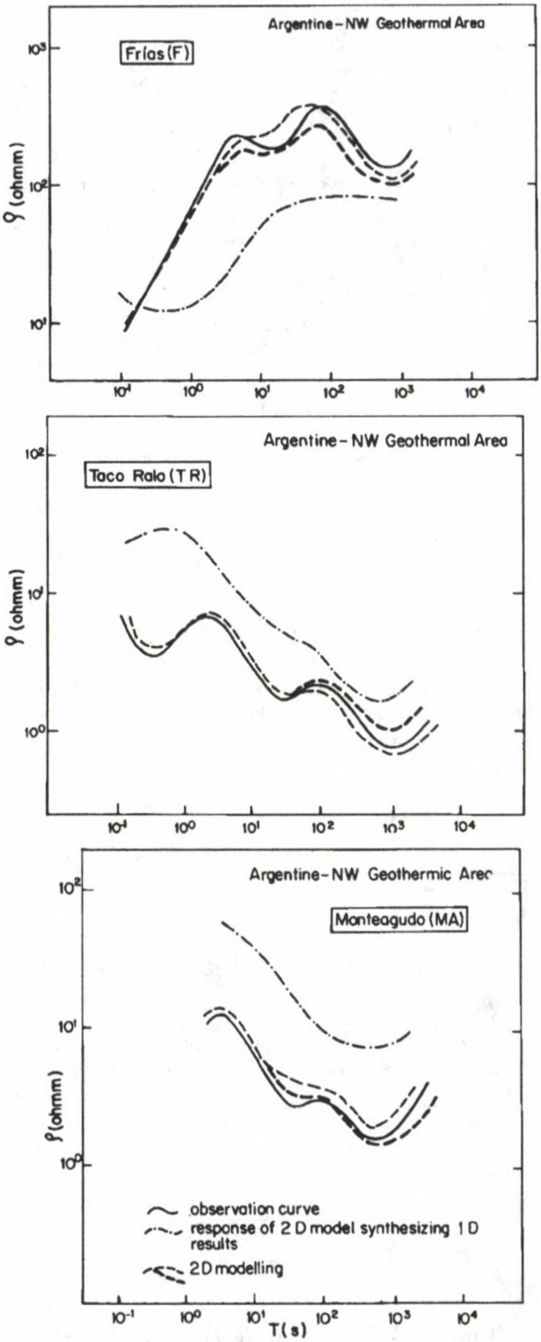


Fig. 4. Geothermal area in NW Argentina. E-polarization curves. Fitness of responses of 2D models to observations: RMS=25% (F); RMS=28% (TR); RMS=17% (MA). Response of 2D model synthesizing 1D model results: RMS=75% (F); RMS=31% (TR); RMS=447% (MA)

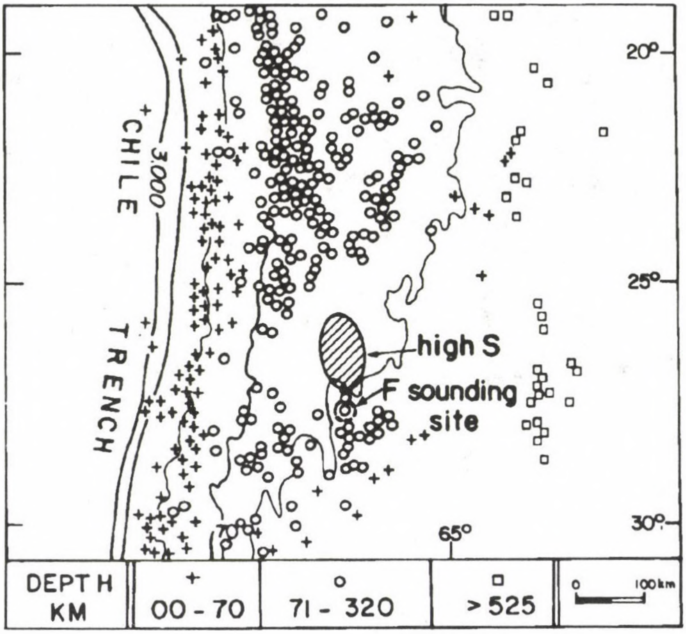


Fig. 5. Seismicity of the area including the geothermal anomaly in NW Argentina (after Barazangi and Isacks 1976). F: Frías sounding site. Also shown is the zone of high conductance (S: 2500–27000 Siemens from 0 to 75 km depth)

characterized by resistivities of 10000 ohmm (Zárate) and 500 ohmm (Chivilcoy) and thickness of about 400 km has been supposed (Table I). The response of a 2D model synthesizing 1D results is in general accordance with observation curves in the

Table I. Thickness of layers and resistivity at sounding sites in the Buenos Aires craton upon 1D results (from Baldis et al. 1983, 1985)

Zárate (Z) Thickness (km)	(6 Layers) Resistivity (ohmm)	Chivilcoy (Ch) Thickness (km)	(2 Layers) Resistivity (ohmm)
0.001	10	3	0.7
0.008	1.5	$\geq$ 400	$\geq$ 500
0.14	10	–	–
0.1	3.3	–	–
400	$\geq$ 10000	–	–
∞	$\leq$ 9	–	–

E-polarization mode (Fig. 6). The screening effect of the thick sedimentary basin below Chivilcoy is surely distorting the apparent resistivity curve at this site but the former diversity in resistivity of large volumes of matter of two tectonic blocks should imply a different trace of the B-polarization curve; this is in contradiction with the unique trace obtained for orthogonal directions of the telluric lines in the field-as given in Baldis et al. (1983).

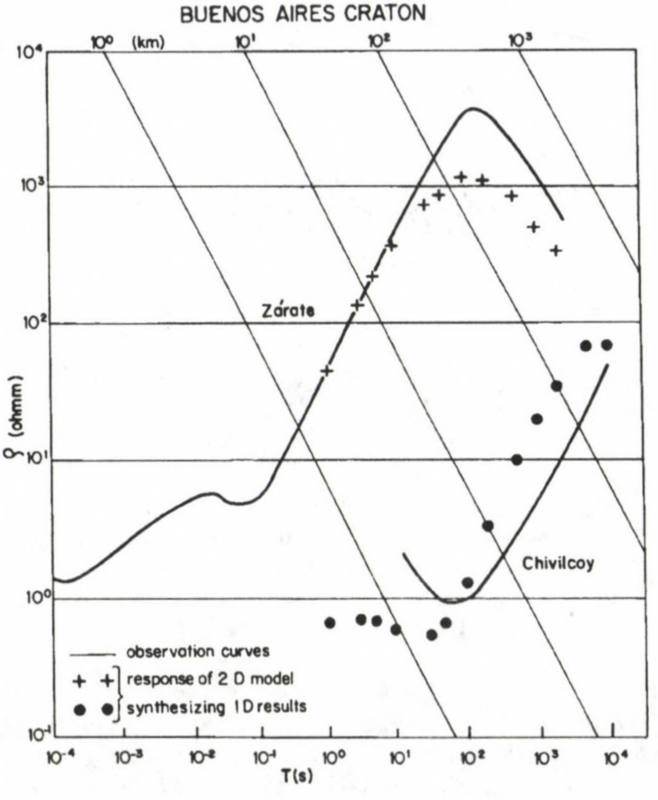


Fig. 6. Buenos Aires craton. E-polarization curves. Response of a 2D model synthesizing 1D results. Sounding sites: Zárate (Z); Chivilcoy (Ch)

Free-air gravity anomalies of large wave-length in the Buenos Aires craton (Cerrato et al. 1980) seem also not to hold the hypothesis of thick blocks of very different resistivities — that may imply a diverse nature of rocks (including density) through the upper crust down to the upper mantle at 400 km depth. In any case, free-air anomalies are more positive in the area surrounding Chivilcoy sounding site (Fig. 7) which is characterized — accordingly to 1D results — by lower resistivity at depth than Zárate.

The Buenos Aires craton may be characterized by large resistivity layers extending to depth but former considerations and lack of other large period sounding sites in the area do make worth not continuing a 2D modelling in this case.

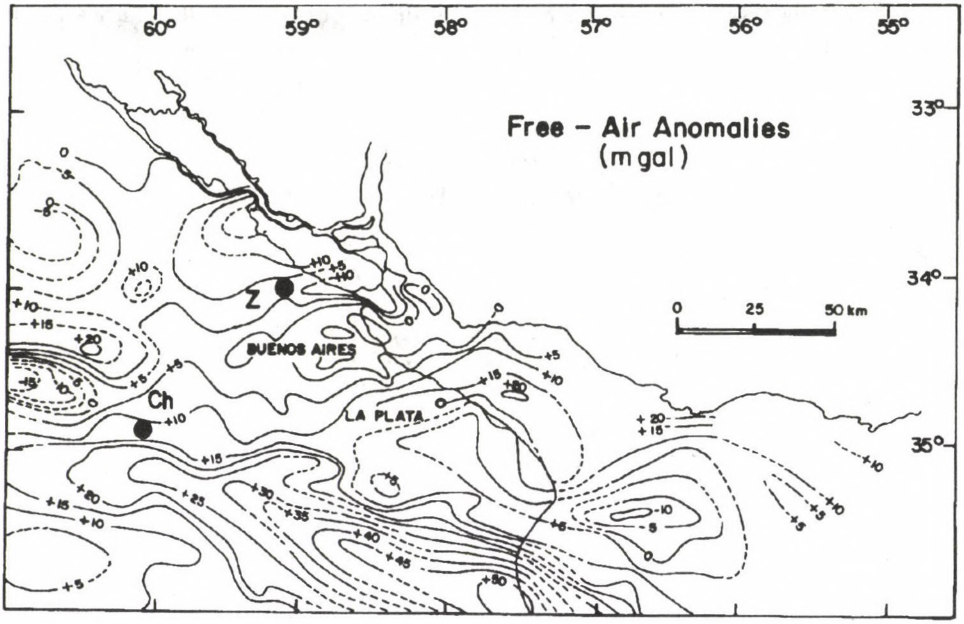


Fig. 7. Free-air gravity anomalies of large wave-lenght in the Buenos Aires craton (taken from Cerrato et al. 1980). Z and Ch: location of magnetotelluric sounding sites

Nunataks Foca and Robertson Island

The sounding sites located around the Matienzo scientific base in the nunatak Larsen are shown in Fig. 8. The nunataks are basaltic islands — generally small — outcropping through the Larsen ice barrier. The nunataks Foca are a modern geomorphology feature made of subaerial basalts with AR/K radiometric ages ranging between 4 ± 1 my (Pliocene) and 0.2 my (Recent). These nunataks are located 300 km east of a Mesozoic subsidence zone in the northwestern edge of the South Shetland Islands (Fig. 9). Subsidence stopped in Early Tertiary. Consumption of oceanic crust was replaced during the Pliocene by rifting and basaltic volcanism on the southeastern margin of the South Shetland Islands — Deception and Bridgman Islands are active volcanoes. Description of soundings, 1D model results and tectonic setting of the environment are contained in Del Valle at al. (1983) and Fournier et al. (1989).

1D model results (Fournier et al. 1989) are shown in Table II. The two shallower layers in the area of nunataks Pedersen (P) and Larsen (MT) represent the working hypothesis concerning ice thickness of the Larsen Barrier and a layer of salt water with resistivity of 0.25 ohmm beneath the ice layer.

2D models have been considered to better approximate the electrical structure of the area. The model chosen is shown in Fig. 10; as in the other cases, the shallower layer boundaries and precise resistivity structure are not well determined

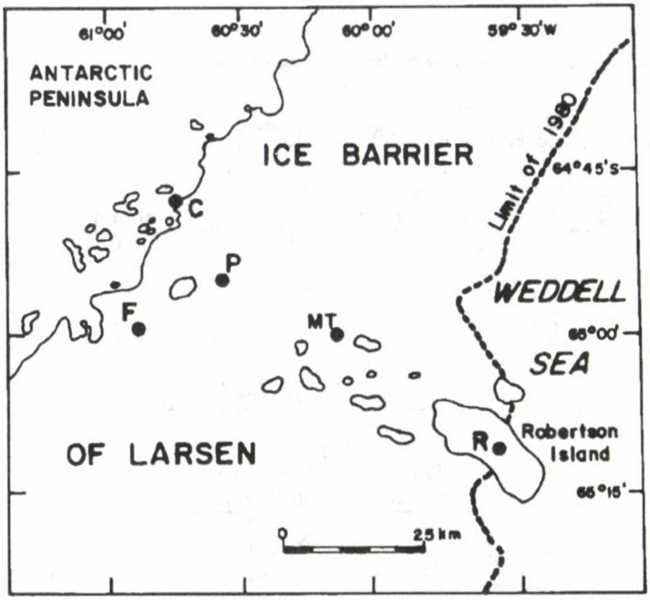


Fig. 8. Sounding sites in the zone of nunataks Foca and Robertson Island. P: sounding at 5 km NE of Pedersen nunatak; MT: sounding at 3 km south of Larsen nunatak (Matienzo base); R: sounding on Robertson Island; F: sounding east of Cape Fairweather, on the Larsen Barrier (taken from Fournier et al. 1989)

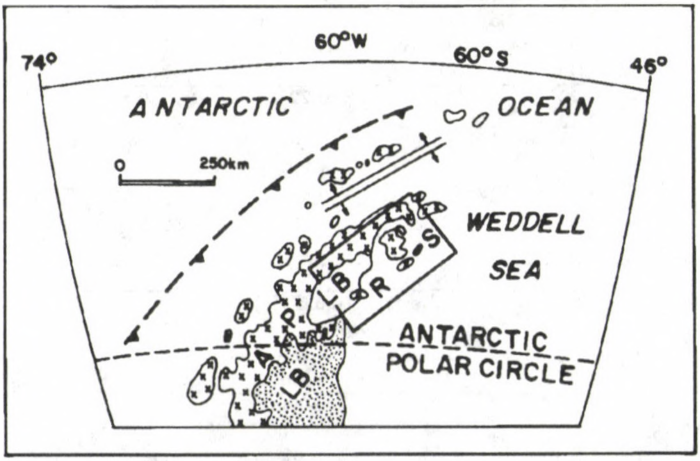


Fig. 9. General view of the region including the Antarctic Peninsula. AP: Antarctic Peninsula; LB: Larsen Barrier; R: Robertson Island; S: Seymour Island. Black triangles indicate a Mesozoic subsidence zone (taken from Fournier et al. 1989)

Table II. Thickness of layers and resistivity at sounding sites in the zone of nunataks Foca (Antarctic Peninsula) (from Fournier et al. 1989)

Pedersen nunatak (P) (6 Layers)		Larsen nunatak (MT) (5 Layers)	
Thickness (km)	Resistivity (ohmm)	Thickness (km)	Resistivity (ohmm)
0.17	1000	0.225	1000
0.02	0.25	0.08	0.25
1.6 s.l.	≥ 25	0.45 s.l.	2
1.9 s.l.	2	28	≥ 250
12	≥ 100	∞	5.5
∞	0.5		

s.l.: sedimentary layer

Robertson Island (R) (6 Layers)	
Thickness (km)	Resistivity (ohmm)
0.2	500
0.035	0.6
1.5 s.l.	≥ 60
0.55 s.l.	2
20	≥ 400
∞	7

Table III. Thickness and resistivity of layers of the 2D model down to 5.26 km at sounding sites in the zone of nunataks Foca (Antarctic Peninsula)

Pedersen nunatak (P) (5 Layers)		Larsen nunatak (MT) (6 Layers)	
Thickness (km)	Resistivity (ohmm)	Thickness (km)	Resistivity (ohmm)
0.22	1000	0.22	1000
0.02	0.3		
1.64 s.l.	25	0.08	0.3
		1.56 s.l.	2
1.9 s.l.	2	0.55	250
1.5	100	1.35	3000
		1.5	1000

s.l.: sedimentary layer

Robertson Island (R) (5 Layers)	
Thickness (km)	Resistivity (ohmm)
0.22	3000
0.04	0.6
1.6 s.l.	60
0.55 s.l.	2
2.85	3000

due to the scarce number of soundings. Also, the electrical structure at depth is provisional, the more accurate determinations being till about 35–45 km depth (Fig. 10). The effect of ocean water (coast effect) has been considered by including

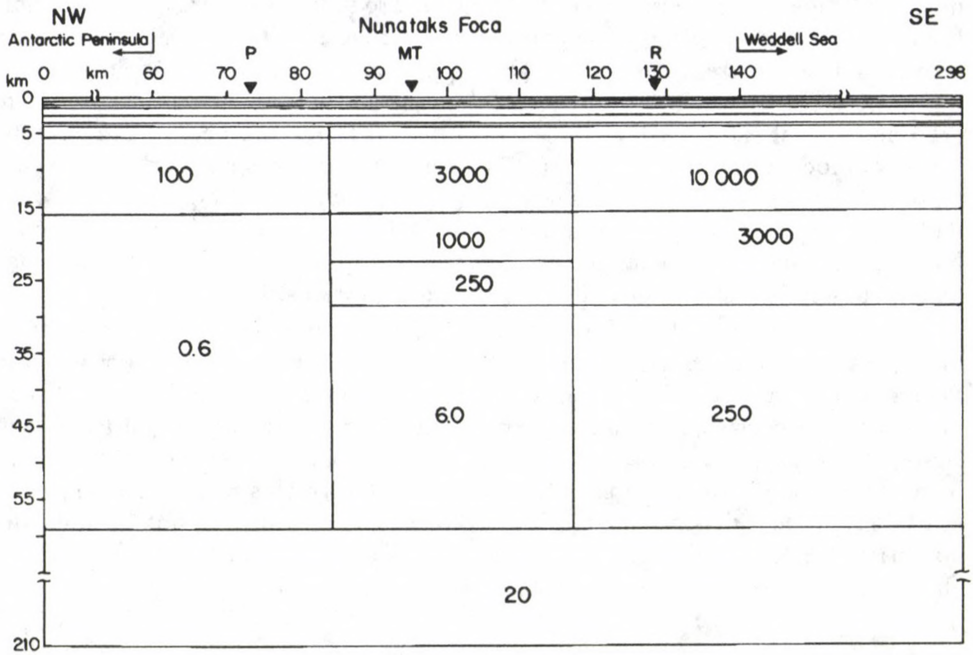


Fig. 10. 2D model of the zone of nunataks Foca and Robertson Island. Sounding sites: P: Pedersen nunatak; MT: Larsen nunatak; R: Robertson Island. Resistivity values are in ohmm. Vertical scale is exaggerated (See in Table III the electrical structure of the shallower layers)

a 158 km maximum width water layer (Weddell Sea); bathymetry has been inferred from Fig. 2 of Lawver et al. (1991) taken from the tectonic maps of the area (British Antarctic Survey 1985). Different kinds of distortion effects have been presented in the literature: Coast effect, Island effect, Peninsula effect, Channeling effect (e.g., Rokityansky 1982, Jones 1983, Berdichevsky and Zhdanov 1984, Rikitake and Honkura 1985) — surely these effects are very important in the studied area.

Thickness and resistivity of layers of the 2D model down to about 5 km are shown in Table III. Fitness of responses to observations are presented in Fig. 11. In the case of Pedersen (P) and Robertson Island (R) sounding sites, 2D model response for periods larger than 5–10 s almost coincide with response of a 2D model structure that synthesizes 1D results, even if high resistivity layers of 3000 and 10000 ohmm — absent in the 1D sections — are encountered through the 2D modelling (see Table II and Fig. 11). Also, a result concerning the sedimentary accumulation in the region can be advanced. Geological and magnetotelluric studies have established the existence of an extended sedimentary basin in the region (e.g., Del Valle et al. 1983, Del Valle and Fourcade 1986, Fournier et al. 1989); 1D resistivity models of observations carried out in the Marambio scientific base located in Seymour Island (Fig. 9) give a total thickness of sedimentary accumulation of 6.7 km under this area and of 3.7 km under sounding site F (Fig. 8) located east of Cape Fairweather, on the Larsen Barrier (Fournier et al. 1989), the thickness and resistivity of sedimentary layers in the 1D models of P, MT and R sounding sites are shown in Table II (s.l. layers). 2D modelling corroborates the former results, but sedimentary accumulation in the area of nunatak Larsen (MT) appears to be thicker than what was established by means of the 1D model — the two thickness values are: 1560 m (2D) and 450 m (1D). Also, as can be seen in Fig. 11 around periods of about 1s, fitness of model response to Larsen (MT) observation curve is not very large, and a better fitness may be provided if a sedimentary accumulation of higher conductance under this site could be entered into the electrical structure of the area of nunataks Foca. Then, under this area, the thickness of the sedimentary accumulation may be not so variable as proposed through 1D model sections.

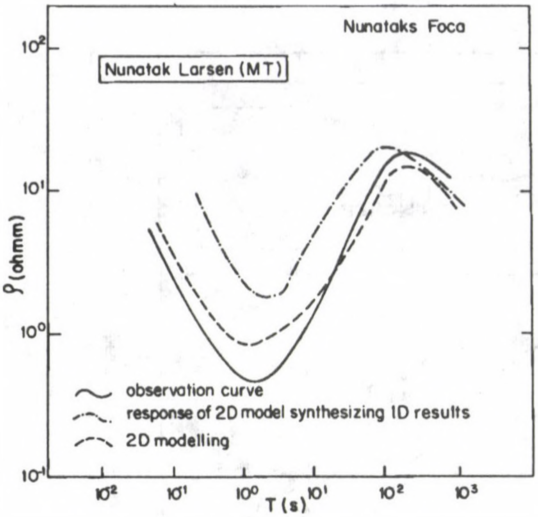
As can be observed from Table II and Fig. 10, a variable conductive layer — most possibly molten rock — is found at nearly 16 km depth in the area of nunatak Pedersen (P) and at 29 km depth in the area of nunatak Larsen (MT). Both 1D and 2D models are consistent in this determination. However, in the case of Robertson Island, there is no such a conductive layer upon results of the 2D modelling at least within the 60 km upper structure; in contrast with this a conductive layer at nearly 22 km depth is encountered in the 1D modelling which do not consider the distortion due to the coast effect of the Weddell Sea — this effect is stronger on the observations done in the sounding site in Robertson Island.

MT soundings in Southern Chile (39° S latitude)

Distortion effects on the E-polarization mode of apparent resistivity curves of magnetotelluric soundings caused by large heterogeneities in the crust and upper mantle in southern Chile are observed in soundings carried out between the Andes Cordillera (Villarrica volcano) and the coastal range (S. José de la Mariquina), (Fig. 12).

We are not dealing here with distortion effects related to shallow heterogeneous structures, but with distortions which are observed in the long period range of magnetotelluric soundings. High values of apparent resistivities in both modes of polarization were observed in the magnetotelluric soundings in the area of Villarrica volcano (see the curves corresponding to Metreñehue (ME), Playón (PL) and Jaramillo (JA) sounding sites in Fig. 13). High apparent resistivity curves have been obtained in many other regions of the Earth — e.g., Rhine graben, Kola Peninsula, Southern Karelia, Bohemian massif, Planchez du Morvan, Antarctica (see Rokityansky 1982). Diverse tectonic environments and geodynamical processes may afford the distinct features of the electromagnetic field in these regions.

1D models of observations done in the zone of Villarrica volcano and preliminary 2D models of the electrical structure of the region are proposed and a synthesis of other geophysical results is presented (Muñoz et al. 1990a, 1990b).



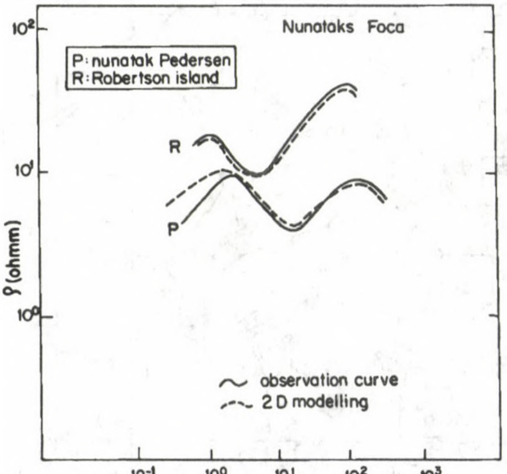


Fig. 11. Nunataks Foca and Robertson Island. E-polarization curves. Fitness of 2D model response to observations: RMS=28% (MT); RMS=18% (P); RMS=17% (R). Response of 2D model synthesizing 1D model results: RMS=51% (MT); RMS=18% (P); RMS=17% (R)

1D models of soundings in the zone of Villarrica volcano

The observation curves for the E-polarization mode of soundings sites in the zone of Villarrica volcano are shown in Fig 13; the observation curves for both modes of ME sounding site are shown in Fig. 14. To approach the problem of distortion a 1D model construction has been followed by considering for large periods the shifted

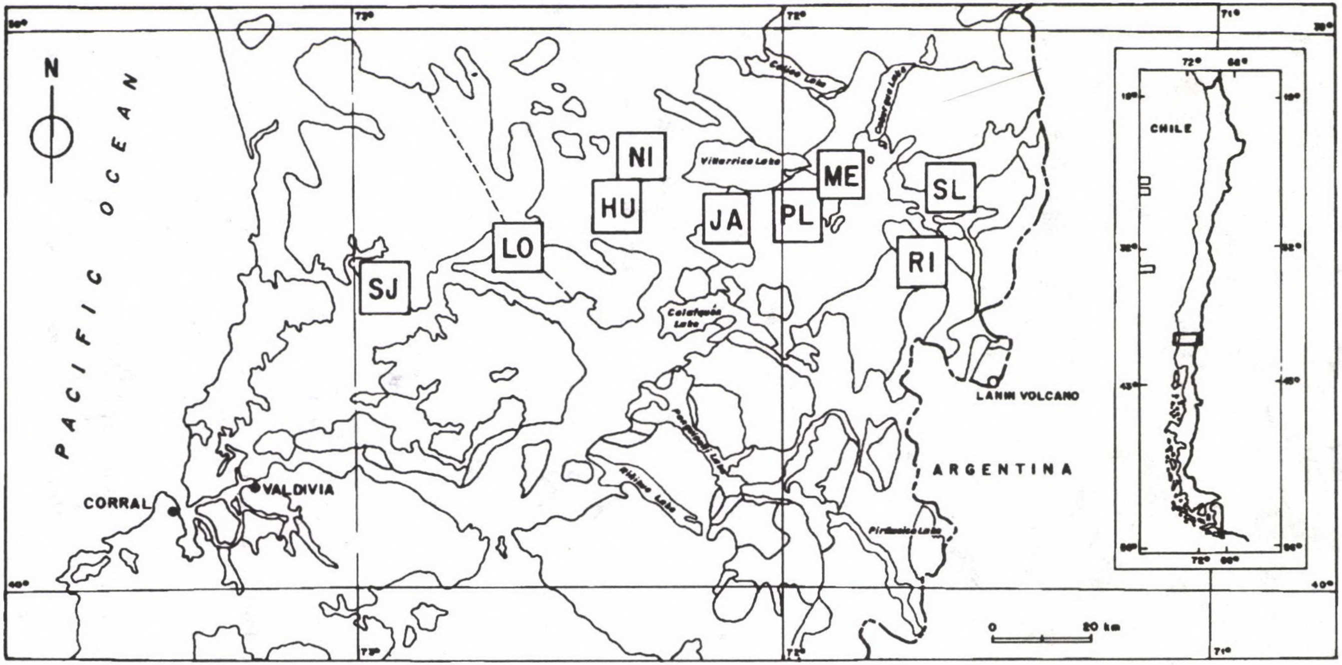


Fig. 12. Location of sounding sites in Chile at 39° S latitude. SJ: San José de la Mariquina; LO: Loncoche; HU: Huiscapi; NI: Ñancul; JA: Jaramillo; PL: Playón; ME: Metreñehue; RI: Rinconada; SL: San Luis, JA, PL and ME and RI are the sounding sites nearest to Villarrica volcano

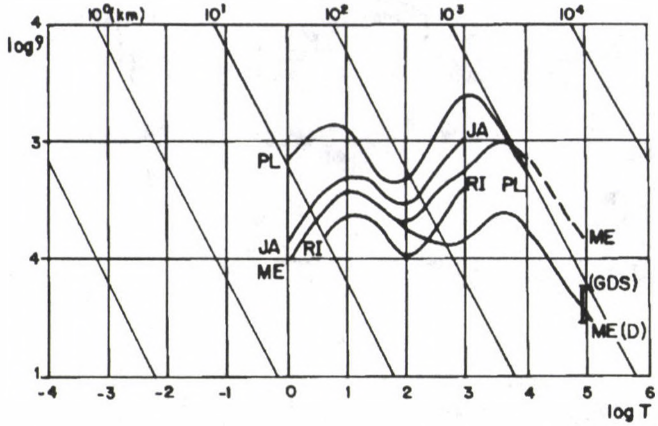


Fig. 13. E-polarization curves for sounding sites near Villarrica volcano. ME-ME(D) corresponds to the tying in of the ME curve to the geomagnetic global model value at T=24h (taken from Muñoz et al. 1990a)

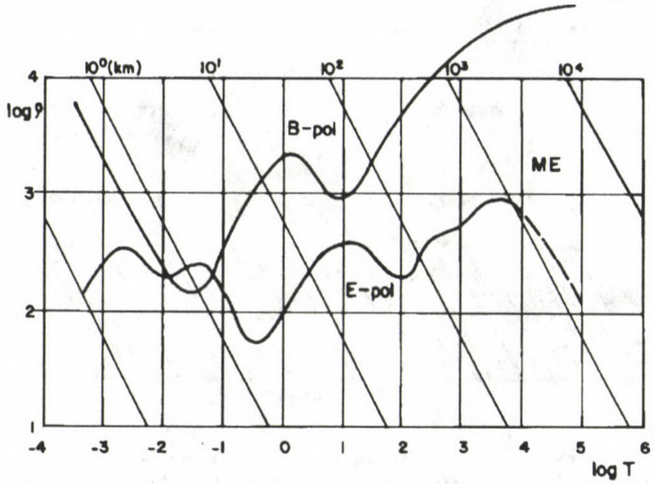


Fig. 14. E-polarization and B-polarization curves of ME sounding site (taken from Muñoz et al. 1990a)

segment of the ME curve fitting the geomagnetic global model value at T=24h (Fig. 13). 1D models give a conductive layer for the three sounding sites considered (Fig. 15) in the depth range from about 40 to 100 km; for ME an increased resistivity (600 ohmm) between 100 and 160 km is also observed. In the boundary at 500 km resistivity has the value of 2 ohmm which manifests the tying in of the distorted curve to the geomagnetic global model value.

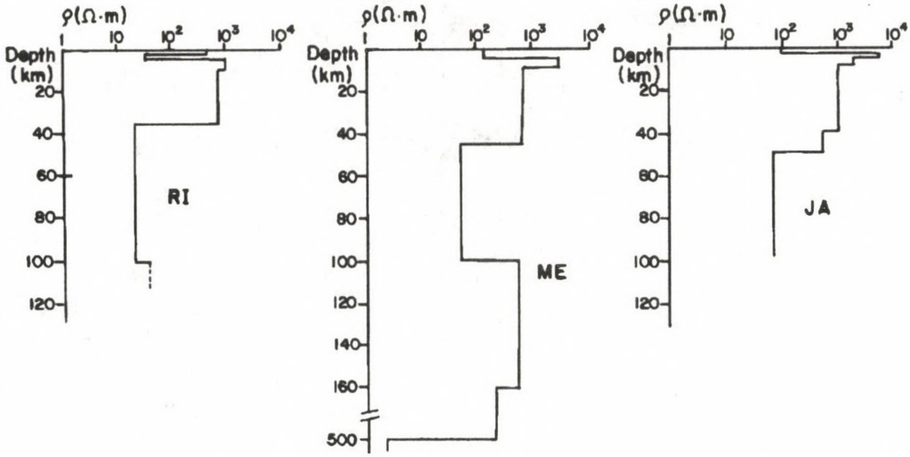


Fig. 15. 1D models for observations in soundings sites near Villarrica volcano. RI: Rinconada; ME: Metreñehue; JA: Jaramillo (taken from Muñoz et al. 1990a). These models consider the tying in of the curves to the geomagnetic global model value at T=24h (see Fig. 13)

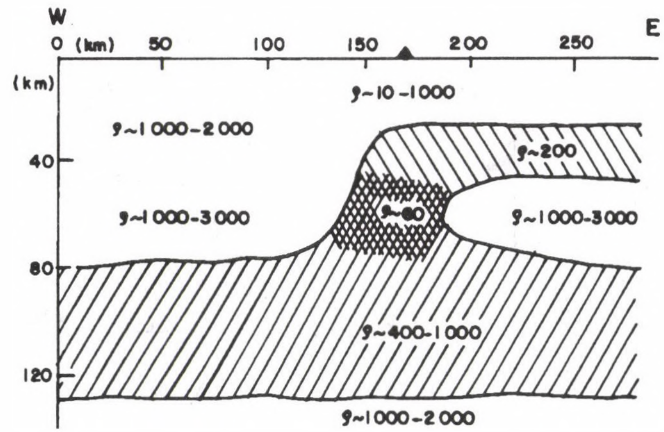


Fig. 16. Schematic cross-section of the area surrounding Villarrica volcano based upon 2D test models considering lateral electric heterogeneities in the continental crust and upper mantle. Resistivity values in ohmm. Vertical scale is exaggerated. Black triangle indicates the position of Villarrica volcano (taken from Muñoz et al. 1990a)

Distortion effects due to heterogeneities in the crust and upper mantle continental structure

1D models of observations in the area of Villarrica volcano are interesting because they give a preliminary picture of the shallower structure of the volcanic area, but they are misleading when a more global context is envisaged. The determination of the earth's structure is tightly connected to the investigation of geodynamical processes, and this is enhanced in the case of active tectonic regions as the one studied presently in southern Chile.

Distortion effects at long periods may be partially due to heterogeneities in the crust and upper mantle. For large periods, the reference area at the ground surface and the effective volume — which characterizes the region where the magnetotelluric field is effectively formed — grow larger, and effects of lateral heterogeneities can be neglected no longer (Rokityansky 1982, p. 202). Two-dimensional model structures extending hundreds of kilometers in the horizontal direction from the

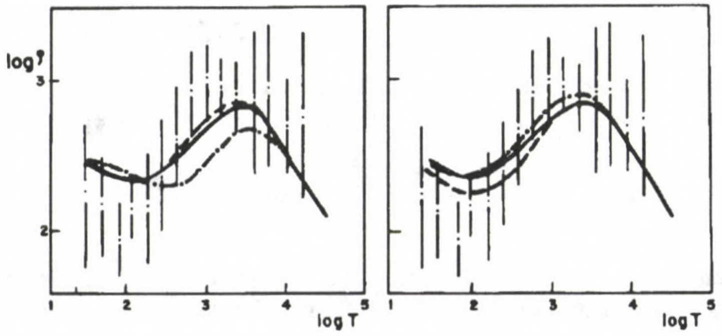


Fig. 17. An example of response curves of 2D test models compared with ME data. Dispersion bars: data. Solid, broken and dot-dash curves: response of 2D models used to construct the schematic cross section of Fig.16 (taken from Muñoz et al. 1990a)

sounding area at Villarrica volcano have been assumed as to study the effect of lateral heterogeneities on the distortion at large periods. Nearly a hundred model parameter variations were performed (Muñoz et al. 1990a) resulting a schematic and preliminary cross-section as the one shown in Fig. 16; an example of the fitness to observations is shown in Fig. 17. Briefly, 2D modelling shows that the conductive zone with its upper boundary at 45–60 km depth in the area of Villarrica volcano is surrounded by rocks of higher resistivity of about 2×10^3 ohmm; the ultimate layer beneath 130 km depth may reach a resistivity value of about 10^3 ohmm.

Distortion effects including the coast effect and conductivity anomalies in the oceanic crust and mantle

As we are dealing with long periods of the electromagnetic field — the effective volume of its formation is very large — it is necessary to consider the crust and mantle structure of the Pacific Ocean along the first hundreds of kilometers of extension to obtain a more real model from the magnetotelluric observations beween the Andes Cordillera and the coastal range.

An increasing apparent resistivity in the E-polarization mode is observed starting at nearly 500 s in the sounding sites in Villarrica volcano and in Huiscapi (HU) (Figs 13 and 18). In S. José de la Mariquina (SJ, Fig. 18) and increase in apparent resistivity starts at 10^3 s from a minimum value of about 25 ohmm. This minimum

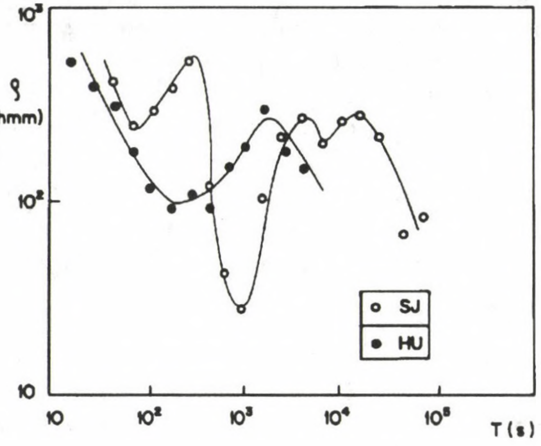


Fig. 18. E-polarization curves for soundings sites SJ and HU

is not present in apparent resistivity curves of sounding sites which are not near the coast. This observation may indicate that the effect of ocean water and — more

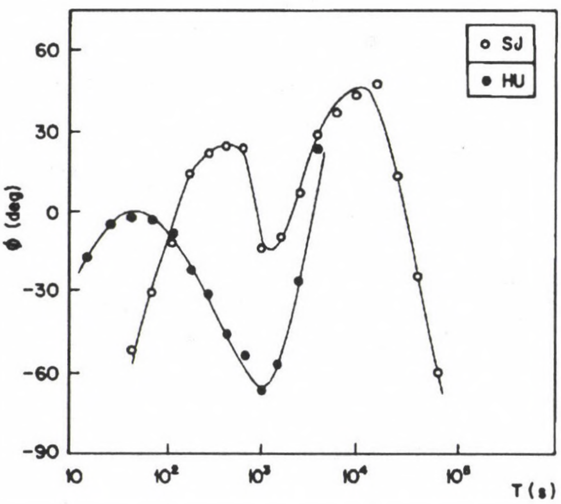


Fig. 19. Phase curves for sounding sites SJ and HU (E-polarization mode)

largely — the structure of oceanic crust and mantle are strongly affecting the MT measurements in SJ that lies at 28 km from the Pacific Ocean. Phase curves corresponding to HU and SJ (Fig. 19) are somewhat consistent with apparent resistivity observations but further research is needed to investigate the effects of ocean water in soundings near the coast.

In order to construct a more reliable model of the earth's structure of continental Chile at latitude 39° S, a scheme including oceanic structures has been assumed and subjected to nearly 80 model parameter variations (Muñoz et al. 1990b). Modelling assumes a very distinctive resistivity pattern of the oceanic lithosphere and asthenosphere. The conductive layer under the ocean has been characterized by resistivity values of about 0.1–100 ohmm with its upper level at a depth of nearly 45 km. The preliminary cross-section obtained is shown in Fig. 20. Even if the principal features shown in Fig. 16 are conserved, to take account of the coast effect

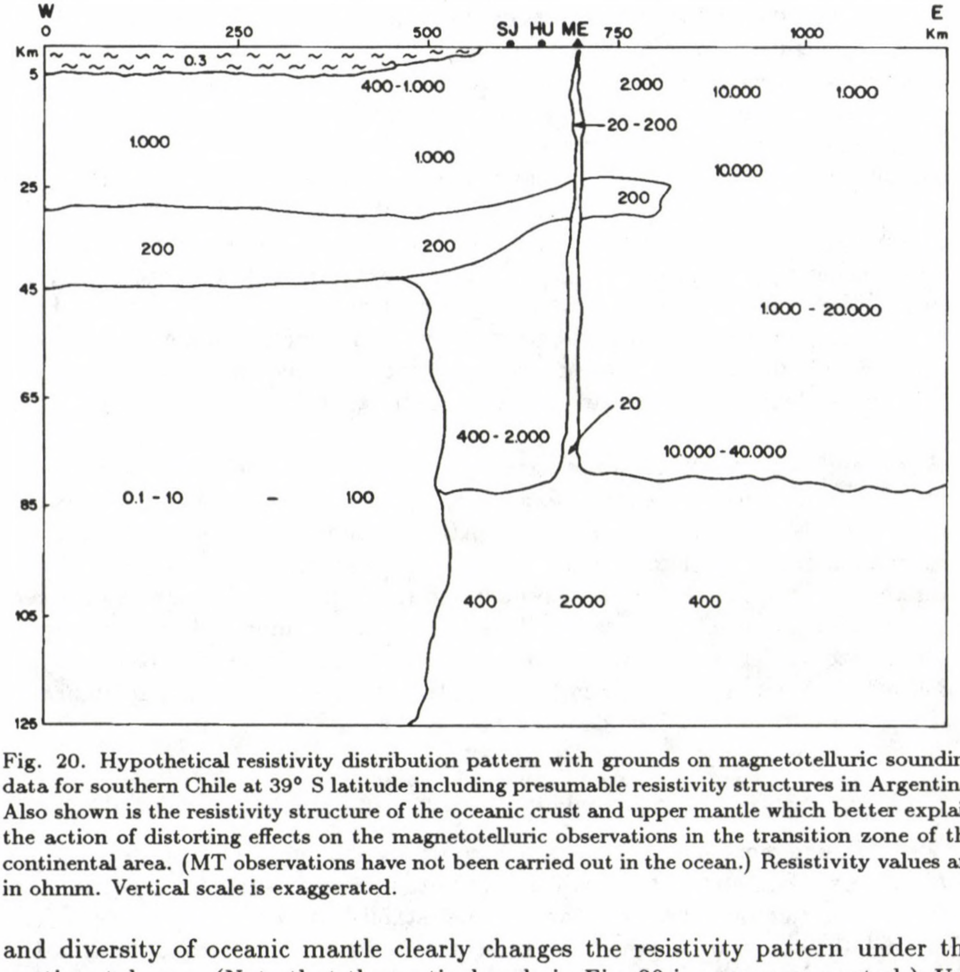


Fig. 20. Hypothetical resistivity distribution pattern with grounds on magnetotelluric sounding data for southern Chile at 39° S latitude including presumable resistivity structures in Argentina. Also shown is the resistivity structure of the oceanic crust and upper mantle which better explain the action of distorting effects on the magnetotelluric observations in the transition zone of the continental area. (MT observations have not been carried out in the ocean.) Resistivity values are in ohmm. Vertical scale is exaggerated.

and diversity of oceanic mantle clearly changes the resistivity pattern under the continental area. (Note that the vertical scale in Fig. 20 is very exaggerated.) Under the area of Villarrica volcano and to the east (Andes Cordillera and Argentina) a resistive lower crust has to be considered; in the eastern region the transition to the upper mantle is also characterized by a high resistivity. The source of magma of Villarrica volcano seems to be concentrated at depths between 70 and 80 km but there is not a simple connection with an 'asthenospheric layer' in the upper mantle. The case may be that the asthenosphere in the transition zone of the continent is subjected to cooling by the subsiding oceanic lithosphere — thermal modelling of Tohoku region (Japan) and central Andes carrried out by Honda (1985) has shown this perturbing interaction. Thermal pulses originated in the middle mantle and geochemical reactions through the oceanic subsiding lithosphere may be the source of magma formation under Villarrica — longitudinal and transverse faults may afford these processes. Estimates of the degree of coupling at the plate interface thrust between latitudes 37° and 46° S in Chile (Spence 1987) may partially support the former hypothesis. A general agreement with other geophysical and with geochemical studies in the area — gravity, morphology of the Wadati-Benioff zone, isotope geochemistry 87Sr/86Sr, SB index of melting — as discussed in Muñoz et al. (1990a) is greatly conserved by this scheme.

Modelling results of MT soundings in southern Chile have shown that the coast effect depends upon the depth of the conductosphere — the level of the earth, at about 450–670 km depth, where a significant drop in resistivity to 80–1.0 ohmm is observed in most cases (Rokityansky 1982, Rikitake and Honkura 1985, Schultz and Semenov 1990). The scatter of data in this case does not allow a reliable determination of the upper level of the conductosphere, but different resistivity values in the range of 20–0.1 ohmm strongly control the coast effect over a large period range. These results are in agreement with the analysis followed by Osella and Duhau (1988) concerning the control of coast effect by conductosphere depth and its dependence on distance of electromagnetic sounding sites from the coastal range.

Very high apparent resistivities are observed in the B-polarization mode in soundings from Villarrica volcano to the coastal range (see for example the ME curve in Fig. 14). The overrise of curves in this mode is attributed to the buildup of charges at resistivity interfaces (Price 1973). A tectonic area as southern Chile characterized by a complex fracture system and the Liquiñe-Ofqui megafault is certainly manifesting itself in this anomalous B-polarization mode. It is necessary in the future to model the "recovery zone" (or "adjustment distance" - Ranganayaki and Madden 1980, Ádám and Szarka 1986) to obtain a comparative reference to which size results of the 2D modelling code.

Conclusions

Even if quality of data is not good, 2D modelling of magnetotelluric observations in Argentina, Chile and Antarctica has proved to introduce a more real parameterization of the electric structures than the one established through 1D models. The 2D model-structures are in greater agreement with other geophysical data than the 1D model sections.

2D modelling has corroborated the existence of an extended geothermal area in northwestern Argentina (Tucumán - Santiago del Estero) and, in turn, has limited the southward extension of this area; the south boundary of the geothermal anomaly coincides with the zone of reinitation of seismic activity which presents a gap through the low resistivity electric structures of the geothermal area.

In the zone of nunataks Foca and Robertson Island (Antarctic Peninsula) coast effect was introduced into the 2D modelling of some few soundings in the area. This has reorganized the morphology of the sedimentary basin as well as it has improved the determination of a conductive layer (possibly partially molten rock) at variable depth (16 to 60 km) from the eastern edge of the Antarctic Peninsula towards the zone of Robertson Island (Weddell Sea). A combined effect of ocean water east of the sounding sites, an ocean water layer of about 20–80 m of thickness under the Larsen ice barrier and the low resistivity sedimentary layer is acting on the electromagnetic observations in this area. Magnetotelluric studies determining the morphology of the sedimentary basin and the structure of the conductive upper mantle in this area should be important for studying the process of ocean floor opening.

In the zone of Villarrica volcano (Chile), 1D models give a conductive layer between 40 km and 100 km depth. 2D resistivity models assumed to investigate the incidence of lateral heterogeneities in the area show that this conductive layer — with upper boundary at 45–60 km depth in these models — may be surrounded by rocks of higher resistivity of about 2000 ohmm. Then the conductive volume may be an anomaly under the zone of Villarrica volcano. Further 2D modelling of observations carried out between Villarrica volcano and the Pacific coastal range including the coast effect and assuming conductivity anomalies in the oceanic crust and mantle has produced a provisional resistivity pattern of continental structures at 39° S latitude. The source of magma of Villarrica volcano seems to be a zone between 70 and 80 km depth but there is not a simple connection with an 'asthenospheric layer' in the upper mantle. Also, the Andes Cordillera and the eastern region towards the Atlantic Ocean may be characterized by a resistive lower crust. The resistivity pattern of the Pacific transition zone including active volcanoes is surely very complex and shows great differences with cratonic regions in Argentina. The Buenos Aires craton, for example, may be characterized by very large resistivities of the crust and upper mantle (about 10^4 ohmm) — but the small number of soundings and contradictory observations do not allow a further insight on this problem. Also, at the present stage of the research, hypothetical magma chambers in the crust have not been individuated under the zone of Villarrica volcano; besides the quality of data, the discovery of such magma chambers by using the magnetotelluric method may depend on the topography of the region, the type of magma and the extended resistivity distribution pattern.

References

Ádám A, Szarka L 1986: *Acta Geod. Geoph. Mont. Hung.*, 21, 177–199.

Baldis B, Demicheli J, Febrer J, Fournier H, García E, Gasco J C, Mamani M, Pomposiello M C 1983: *Acta Geod. Geoph. Mont. Hung.*, 18, 489–499.

Baldis B, Febrer J M, Fournier H G, Gasco J C, Ghidella M, Mamani M J, Pomposiello M C, Vaca A 1985: Structures très diverses de la croûte et du manteau supérieur dans la moitié Nord de l'Argentine, avec découverte d'une zone géothescale. C.R. Acad. sc. Paris, t. 301, Série II, 1245–1250.

Barazangi M, Isacks B L 1976: *Geology*, 686–692.

Berdichevsky M N, Zhdanov M S 1984: Advanced Theory of Deep Geomagnetic Sounding. Elsevier, Amsterdam

British Antarctic Survey 1985: Tectonic Map of the Scotia Arc, scale 1:3.000.000. Cambridge

Cerrato A A, Masciotra J L A, Núñez O A, Ignazzi J C 1980: Densificación de líneas gravimétricas en la zone sud de la provincia de Buenos Aires. Contribuciones a la Geodesia Aplicada, Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires

Del Valle R A, Fourcade N H 1986: La cuenca sedimentaria pos-Triásica del extremo nororiental de la Península Antártica. Contr. 323, Int. Antártico Argentino, Buenos Aires

Del Valle R A, Demicheli J, Febrer J M, Fournier H L S, Gasco J C, Irigoin H, Keller M A, Pomposiello M C 1983: Résultats des premiers sondages magnétotelluriques (M-T) effectués au large de la bordure NE de la Péninsule Antarctique. C.R. Acad. Sc. Paris, t. 296, 1189–1194.

Febrer J 1981: La anomalía geotéscica del área central de Noroeste Argentino. Dpto. de Geofísica, Centro Espacial San Miguel, Argentina

Fournier H G 1981: Rapport de mission en Argentine 1980. Centre de Recherches Géophysiques, Laboratoire de Géomagnetisme, Paris

Fournier H G, Demicheli J, Gasco J C, Febrer J M, Del Valle R, Keller M A, Pomposiello M C, Borzotta E 1989: *Cold Reg. Sci., Technol.*, 17, 49–59.

Honda S 1985: *Tectonophysics*, 112, 69–102.

Jones A G 1983: *Geophys. Survey*, 6, 79–122.

Kisak E, Silvester P 1975: *Comput. Phys. Communic.*, 10, 421–433.

Lawver L A, Della Vedova B, Von Herzen R P 1991: *J. Geoph. Res.*, 96, 2019–2038.

Muñoz M, Fournier H, Mamani M, Febrer J, Borzotta E, Maidana A 1990a: *Phys. Earth Planet. Int.*, 60, 195–211.

Muñoz M, Mamani M, Fournier H, Borzotta E, Moyano C, Maidana A 1990b: Distortion of tensor magnetotelluric apparent resistivities caused by large heterogeneities in the crust and upper mantle at latitude 39° S in Chile. 10th Workshop on Electromagnetic Induction in the Earth. Abstracts (5.3.), Ensenada, México

Osella A M, Duhau S 1988: *Geoacta*, 15, 91–99.

Osella A M, Pomposiello M C, Maidana A, Borzotta E 1991: Electromagnetic studies in the Tucumán Basin. XX General Assembly IUGG, IAGA Abstracts, Vienna

Park S K, Orange A S, Madden T R 1983: *Geophysics*, 48, 1402–1405.

Price A T 1973: *Phys. Earth Planet. Int.*, 7, 227–233.

Ranganayaki R P, Madden T R 1980: *Geophys. J. R. astr. Soc.*, 60, 445–457.

Rikitake T, Honkura Y 1985: Solid Earth Geomagnetism. Terra scientific Publ. Co., Tokyo

Rokityansky I I 1982: Geoelectromagnetic Investigation of the Earth's Crust and Mantle. Springer-Verlag, Berlin

Schultz A, Semenov V 1990: An improved view of global deep mantle conductivity. 10th Workshop on Electromagnetic Induction in the Earth, IAGA Working Group I–2. Abstracts (5.8), Ensenada, México

Spence W 1987: *Rev. Geophys.*, 25, 55–69.

Wannamaker P E, Hohmann G W, Ward S H 1984: *Geophysics*, 49, 1517–1533.

Wannamaker P E, Stodt J, Rijo L 1987: PW2D-Finite element program for solution of magnetotelluric responses of two-dimensional earth resistivity structure. Earth Science Laboratory, University of Utah, Research Institute, Publ. ESL-158.

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 87–101 (1992)

RESULTS OF MAGNETOTELLURIC EXPLORATION FOR GEOTHERMAL RESERVOIRS IN HUNGARY

Z NAGY[1], I LANDY[1], S PAP[1], J RUMPLER[1]

The paper demonstrates initial results of magnetotelluric (MT) prospecting for high enthalpy geothermal reservoirs in the Pannonian Basin.

Tectonic zones in high heat flow regions of South-East Hungary, accompanied by carbonate rocks in the Mesozoic basement underlying a 3 km thick Tertiary are considered potential geothermal reservoirs. The existence of the geothermal system has been confirmed by a heavy steam eruption in a hydrocarbon exploratory drilling on December 16, 1985.

GKV carried out a MT survey in the vicinity of the steam eruption to test its resolving power in detecting conductivity anomalies due to a geothermal reservoir. The MT survey discovered zones of low resistivity in the basement.

2D MT numerical modeling and joint interpretation of MT and seismic results confirmed that zones of low MT resistivity in the basement coincide with tectonic zones detected by seismic data. Another steam well located in such a zone has also confirmed this model assumption.

Keywords: electrical conductivity anomaly; high enthalpy geothermal reservoir; magnetotellurics; Pannonian Basin; tectonic zone

Introduction

The temperature gradient in the Pannonian Basin is generally elevated. The temperature of formation water, however, is low, less than 100°C in young Pliocene sedimentary rocks. Thus, it has no importance for direct electric energy production.

Theoretical studies came up to the conclusion that high enthalpy geothermal reservoirs can be present in the Preneogene basement rocks at medium depth range (approx. 4–5 km) within the Pannonian Basin. Those may be suitable for direct electric power production (Stegena 1985, 1990).

Stegena et al. (1991) suggested aspects to select potential areas for deep, high enthalpy geothermal reservoirs in the substratum of the Pannonian Basin. According to this conception high enthalpy reservoirs may be created in that part of the high heat-flow regions of the Pannonian Basin, where remarkable secondary porosity occur due to faulting or karstification in the basement rocks accompanied by a temperature limit higher than 150°C.

Following the selection criterion suggested by Stegena, a sketch-map has been constructed about potential goethermal areas in the Pannonian Basin in Hungary, on the basis of several thousands hydrocarbon and water exploratory wells. Hundreds of seismic profiles are available to determine the location and direction of the main tectonic lines as well (Fig. 1).

[1]MOL Hungarian Oil and Gas Plc. Geophysical Exploration Co. (GKV), H-1068 Budapest, Gorkij fasor 42, Hungary

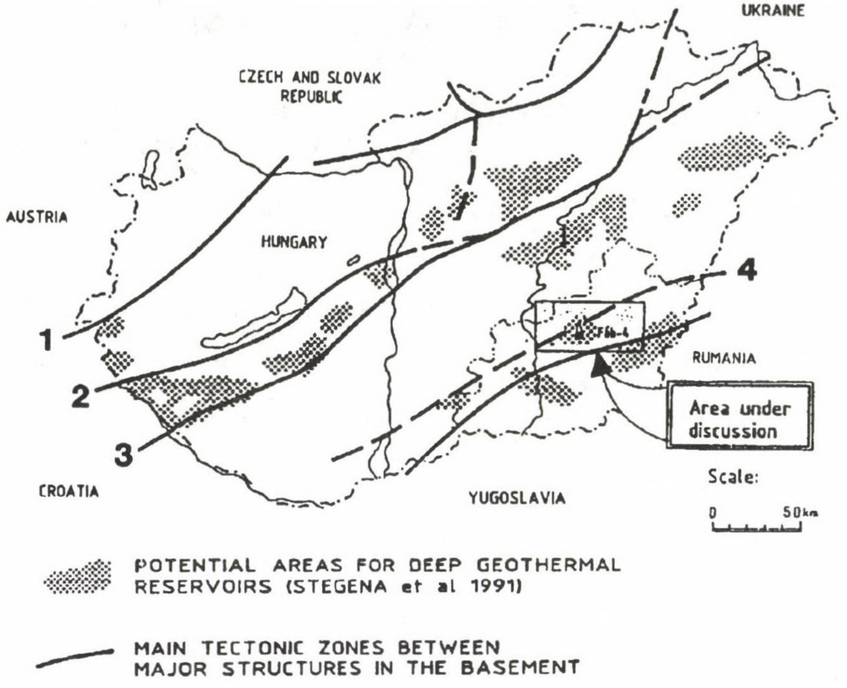


Fig. 1. Sketch map of the karstified and/or tectonically fractured rocks occurring in the substratum of the Pannonian Basin with temperature more than 150°C (after Stegena et al. 1991). Legend: 1. Rába Line, 2. Balaton Line, 3. Middle Hungarian Fault Belt, 4. Békés Line (after Haas 1987)

The existence of the supposed deep geothermal system in South-East Hungary has been proven since December 16, 1985. Then a heavy steam and water eruption came out of a hydrocarbon exploratory drilling: Fáb-4, near the village Fábián-sebestyén. The borehole penetrated a seismic structure located in a tectonic zone on a potential goethermal area (Fig.1). The drilling stopped in Lower Triassic sandstone at 4239 m depth. The breakout occurred from the 3800–4000 m depth range, where Middle-Triassic dolomites had been penetrated. Measurements showed a temperature of 202°C in 4226 m depth from the surface.

Killing the well took forty-one days. Although the drilling verified unexpectedly the presence of an existing high enthalpy deep geothermal reservoir, the borehole itself became technically unfit for any further operation, including the utilization of heat power.

An actual need has been created by the facts mentioned to prompt prospecting activity for geothermal reservoirs in that region. A recommendation for a primary geothermal exploration project has been outlined in a study, supported by the Central Office for Geology, Hungary, emphasizing the necessity of geoelectric measurements (Rumpler 1987). During the past decade several comprehensive papers and case histories were published on the special advantages and results of geoelectric

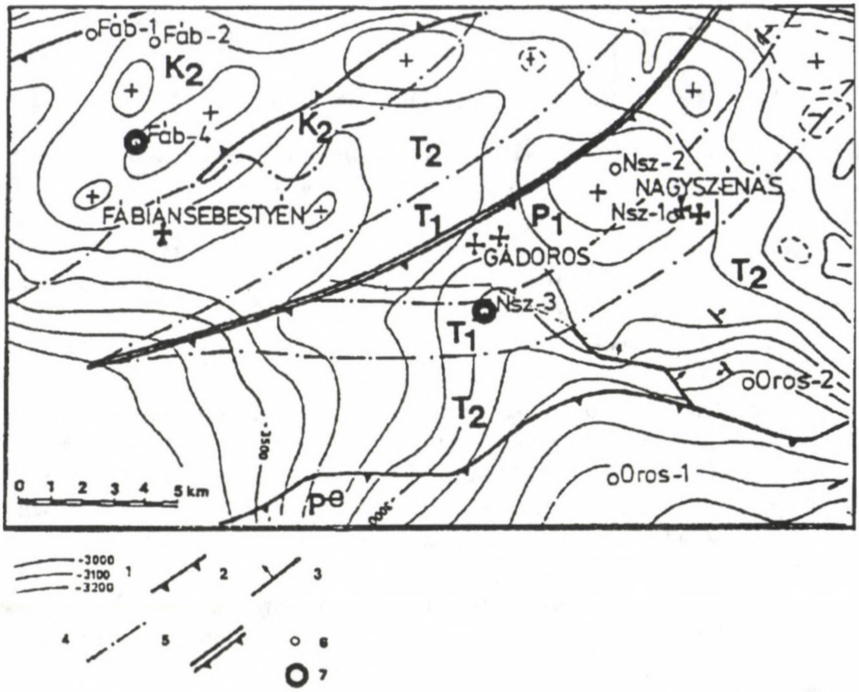


Fig. 2. Structural map of the Fábiánsebestyén and Nagyszénás areas based on the results of previous seismic exploration and deep drilling data. Legend: 1. Depth contours (m) of the Pre-neogene basement top, 2. Overthrust, 3. Normal fault, 4. Lithological boundary, 5. Békés Line, 6. Deep borehole, 7. Steam well. K_2 — Upper Cretaceous fractured unit, P_1 — Lower Permian Quartzporphyr, P^e — Precambrian- Paleozoic Metamorfites, T_1 — Lower Triassic Sandstone, Conglomerate, Mudstone, T_2 — Middle Triassic Dolomite-breccia

methods in prospecting for geothermal reservoirs (Stanley et al. 1977, Wright et al. 1985, Otten and Mussmann 1985, Hughes and Maas 1987, Thanassoulas 1991). Perhaps the most important physical property change due to the presence of a hydrothermal system, other than elevated temperature and heat-flow, is the change in electrical resistivity of the rock-fluid volume. The electrical conductivity of ionic conductors increases largely with temperature. The hydrothermal mineral assemblages produced by the thermal fluids significantly alter the physical properties of the reservoir rocks in the fracture zones. That is why the conductivity of the host rock of the geothermal field also increases. Thus, most geothermal systems have an associated zone of anomalously low resistivity which may be detectable by different surface electrical methods.

The Hungarian Geophysical Exploration Co. (GKV) was entrusted with an electrical test survey in the vicinity of the Fáb-4 steam well for the purpose to detect electrical conductivity anomalies due to the geothermal field, as well as with a further exploration for geothermal reservoirs in a neighbouring area near Nagyszénás. In the Nagyszénás area a hydrocarbon exploratory barren well discovered also some steam influx previously, but the well has been closed down by cement plug.

Magnetotelluric test results

The "efficiency" of the electrical methods used in geothermal exploration depends upon the specific target sought. Figure 2 is a structural map of the Fábiánsebestyén and the neighbouring Nagyszénás areas compiled on the basis of the previous seismic results and deep drilling data. Depth contours of the top horizon of the Preneogene basement rocks are indicated in the map together with faults and other tectonic zones. The overthrust zone marked by double thick line can be identified with the "Békés Line". The structural map shows that the total thickness of the Tertiary age sediments exceeds 3 km.

Thus, considering the deep reservoir and the complex tectonics of the study area, the tensor magnetotelluric sounding method was selected as a technique for the test survey. Because the case history of the Fábiánsebestyén geothermal field is reviewed by Stegena et al. (1991) including the details of the MT survey which are published by Nagy et al. (1990) as well, this paper is confined to a brief description of the test MT survey and results.

In the Fábiánsebestyén geothermal area, GKV carried out a detail survey in 1988 comprising 17 MT sites in a 3 square km area surrounding the Fáb-4 steam well, to test the resolving power of the MT method. Another 11 MT sites were located by longer profiles running along the traces of two seismic sections surveyed previously. Thus, a comparison with seismic results obtained by the same traces has also been available.

The detail, high resolution MT survey discovered a deep zone with increased conductivity in the Mesozoic basement rocks, as it had been expected. Figure 3 is a vertical resistivity pseudosection running through the Fáb-4 steam well in S-N direction. Contour lines of the MT apparent resistivity values obtained by Bostick (1977) transformation are indicated on it *vs.* Bostick depth values on the vertical axis (Goldberg and Rotstein 1982, Jones 1983). The anomaly of the decreased resistivities detects clearly the tectonic zone penetrated by the Fáb-4 drilling at the depth. A further anomalous part of the resistivity pseudosection is recognized by the sudden lateral change with decreasing values of the Bostick resistivities below the bottom of the Fáb-2 drilling hole, which has been discontinued in Miocene age formation at a shallower level. That anomaly can be interpreted correctly by the associated low resistivity volume of the fractured host rock and fluid of the geothermal field, located in a wide vertical extent ranging over several km depth in the basement.

This model assumption has been supported by 2D plane wave numerical model calculation.

2D MT modelling results

For the interpretation of the resistivity pseudosections using Bostick resistivity values, 2D finite element MT model calculations have been accomplished for both

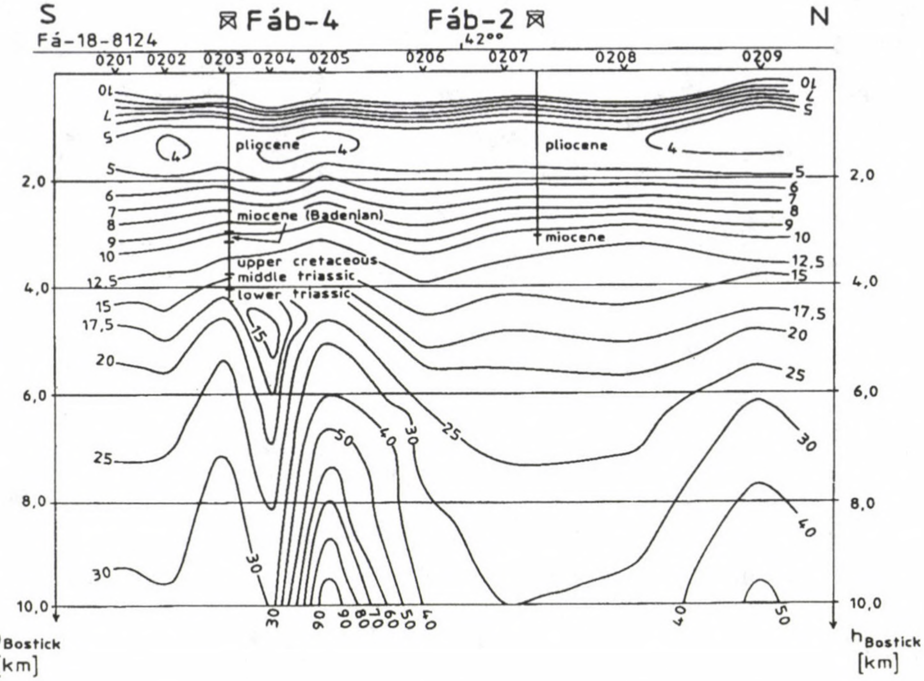


Fig. 3. Vertical magnetotelluric pseudosection through the Fáb-4 steam well, using resistivity values obtained by Bostick transformation

TE and TM modes.

The 2D model studied consists of a three-layer structure plus a low resistivity inclusion ($\varrho_3 = 1$ ohmm) embedded in the second layer, the high resistivity ($\varrho_2 = 200$ ohmm) basement. The first layer represents the sedimentary overburden ($\varrho_1 = 10$ ohmm) with various thickness. The low resistivity inclusion is located between 10–12 km depth range with confined extent in lateral direction. It represents the fractured rock and fluid volume of the reservoir in the basement. The low resistivity asthenosphere ($\varrho_4 = 1$ ohmm) underlain the high resistivity second layer in 30 km depth is not shown in Fig. 5.

In Fig. 4 the calculated curves of MT apparent resistivity are plotted for both TE and TM modes *vs.* square root of period. Reference positions of the calculated pairs of TE and TM mode curves — which are plotted in Figs 4a, 4b and 4c — have been marked in Fig. 5.

Figure 5 shows the resistivity distribution for the 2D model structure as well as contours of the Bostick resistivities calculated for the TE mode plotted against Bostick depth.

The results of 2D modelling have led to the following conclusions:

1. TE mode MT data are influenced strongly by the presence of the low resistivity

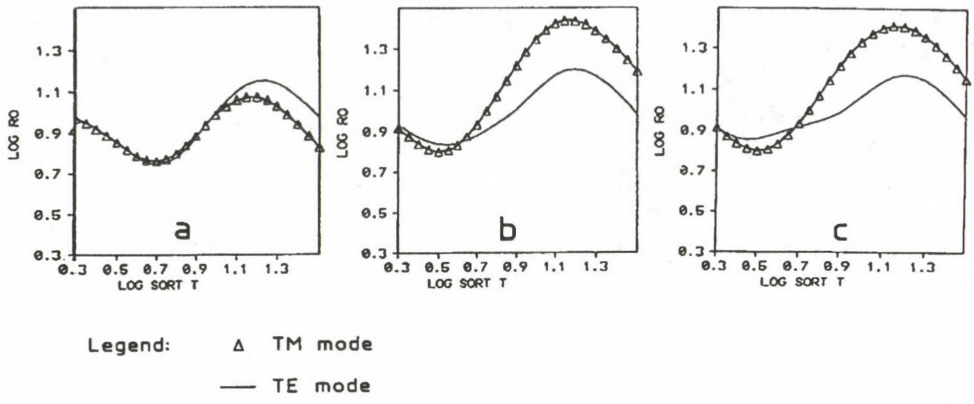


Fig. 4. TE and TM mode MT apparent resistivity curves calculated by 2D finite element numerical modelling for the case of the resistivity structure shown on Fig. 5, a, b and c refer to the corresponding sites marked on Fig. 5

inclusion in a wider environment of the target. However, a significant notch in the TE model Bostick-resistivity distribution is recognized in the vertical pseudosection over the centre of the target.

The modelling results are in full accordance with experimental data of the field test survey.

2. The TM mode data are influenced by the total conductance and the resistivity of the overburden as well as by the resistivity of the basement. Almost no influence of the low resistivity inclusion is visible in the TM mode data.

Therefore, TE mode Bostick resistivity data inform on the presence and location of a deep conductor, like the geothermal reservoir is. It can be expected that TM mode data provide further information on the sedimentary overburden and the electrical properties of the upper part of the basement as well.

These methodological recommendations have been utilized during the further MT survey for geothermal reservoirs accomplished by GKV in the Nagyszénás prospecting area.

MT prospecting for geothermal reservoirs in the Nagyszénás area

GKV carried out further MT prospecting for geothermal reservoirs in the neighbouring Nagyszénás area, in 1990 based on the experience of the test survey.

The aim of the MT survey was to verify the existence of a presumable geothermal system in the vicinity of a barren hydrocarbon exploratory well Nsz-3 in connection with some steam influx had been observed in the well before the finishing operation. The well was closed down by cement plug after it.

The detail MT survey covered a 70 square km area with 98 MT sites which formed a grid approximately. High quality measured data were provided in 2×5

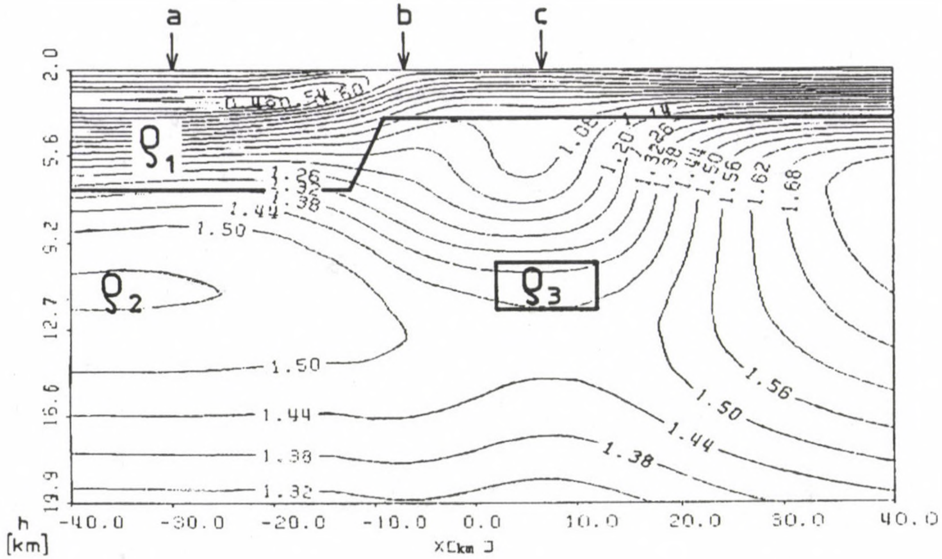


Fig. 5. Resistivity structure with low resistivity inclusion in the substratum for 2D MT model calculation and the response Bostick resistivity pseudosection of the modelled structure for TE mode (log-resistivity values are given as scale of the contour lines)

channel tensor MT arrangement applying remote reference technique for field data acquisition with cable communication and in-field data processing. A remote reference MT system including in-field data processing has been applied for GKV's field activity since 1985 (Ádám et al. 1989). Figure 6 shows a typical example of the measured apparent resistivity and phase curves. As a result of utilization of the methodical conclusions from the model study, an extensive system of tectonic zones has been discovered by areal survey exhibiting significant change of the resistivity in the basement rocks.

Figure 7 shows a characteristic vertical resistivity pseudosection by the profile MT 9 running through the borehole Nsz-3 in perpendicular direction to the strike of the tectonic zone. The Bostick resistivity contours compiled by TE mode data clearly reflect the significant "notch" over the conductive volume of the tectonized reservoir zone.

Figure 8 presents a resistivity *vs.* depth vertical section for the same profile MT9 evaluated on the basis of 1D Marquardt inversion of TM mode data. In spite of the fact that the evaluated depths are considered as approximation based on 1D inversion, the TM mode data provide realistic information on the lateral extent of the tectonic zone. The high resistivity layer in the basement rocks having resistivity values higher than 200 ohmm are considered as the "non-fractured" substratum. The decreased resistivities in the basement ranging between 20–100 ohmm represent the fractured and hydrothermally altered reservoir.

Contour depths of the top of the "non-fractured" substratum are mapped in

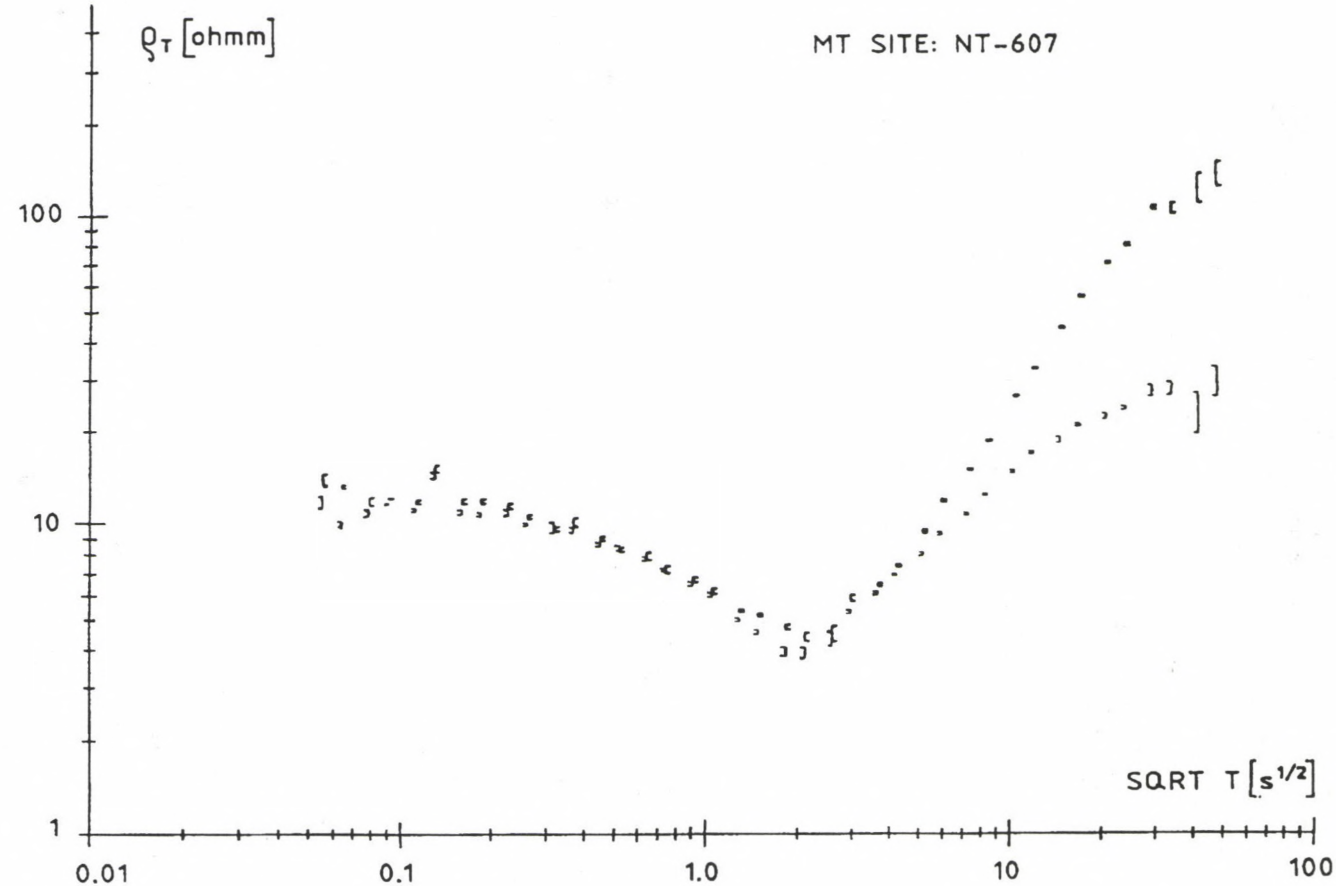


Fig. 6. a. Typical example of the MT apparent resistivity sounding curves measured in the Nagyszénás area; MT site: NT 607

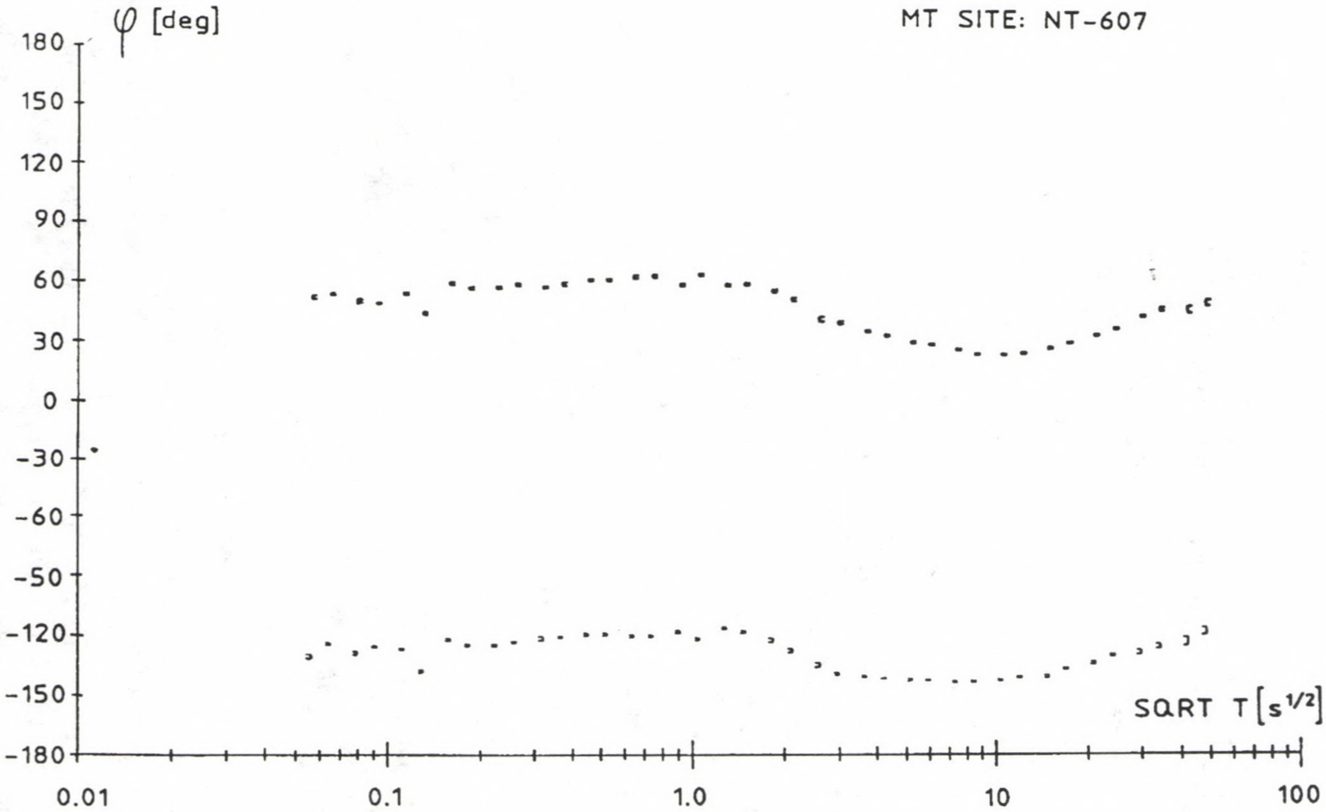


Fig. 6. b. Typical example of the MT phase sounding curves measured at same site, NT 607

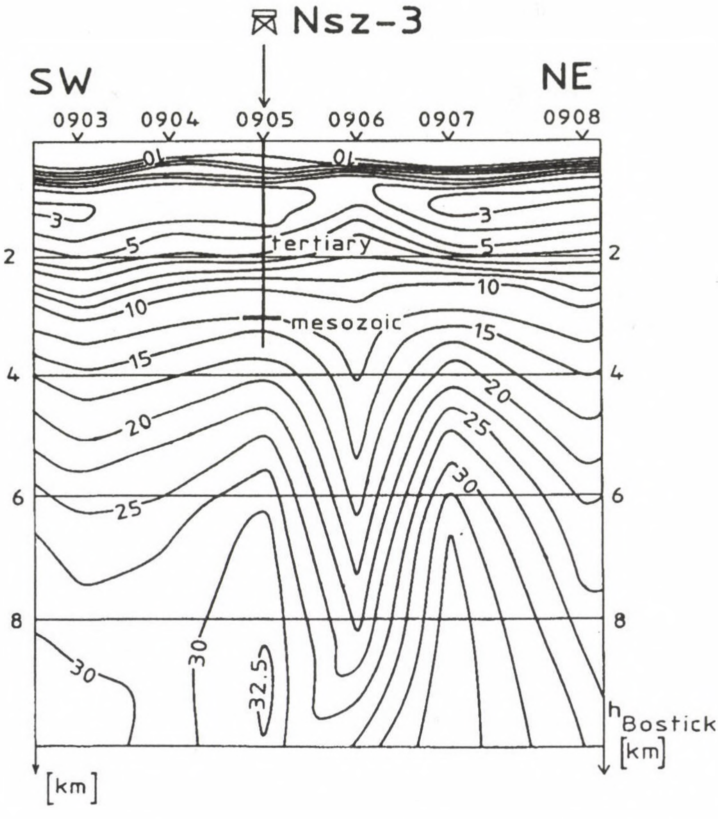


Fig. 7. Vertical MT pseudosection of the profile MT9 through the Nsz-3 borehole, with Bostick resistivity contours, compiled by TE mode data

Fig. 9. Hachuring in the depth contours denotes the area of the fractured and tectonized volume of reservoir rocks. Reference points A, B and C on Fig. 9 refer to the border of the higher conductivity altered reservoir rock volume as compared with seismic data presented in Fig. 11.

Joint interpretation of MT and seismic results

An integrated cross section has been compiled incorporating the results of the test MT survey and seismic reflection data for studying presumable interrelation between them.

Figure 10 shows the vertical cross section of seismic profile Fá-18 running through the Fáb-4 borehole. Seismic reflection horizons representing interfaces between Pliocene, Miocene and Mesozoic formations after time to depth conversion have been incorporated in the cross section together with the MT Bostick resistivity pseudosection. MT conductivity anomalies discussed in connection with Fig. 2 can

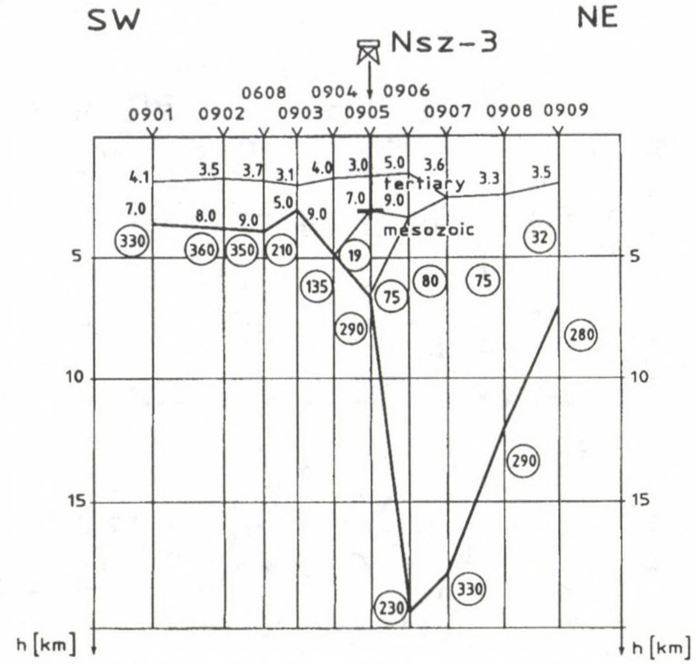


Fig. 8. Vertical resisitivity-depth section of the survey line MT-9 through the Nsz-3 borehole, constructed on the basis of 1D Marquardt-inversion results of TM mode data

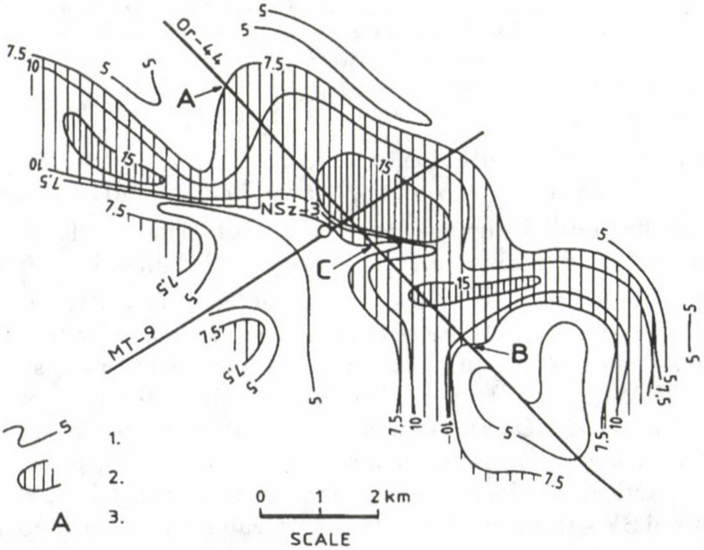


Fig. 9. Nagyszénás geothermal field. Map of the potential geothermal reservoir on the basis of MT survey. Legend: 1. Depth contours (km) of the top horizon of high resistivity (non-fractured) substratum. 2. Zone of fractured rocks (potential reservoir) with increased electrical conductivity, located in a considerable wide depth range in the Preneogene basement. 3. Reference points to compare with seismic results presented in Fig. 11

be identified with the fracture zones revealed by the seismic interpretation.

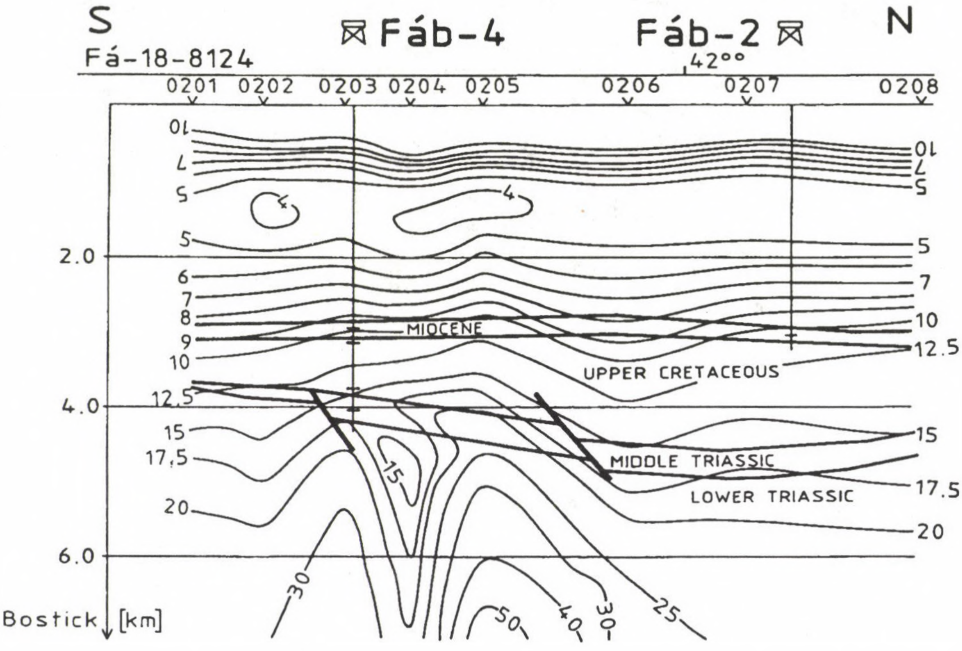


Fig. 10. Integrated results of MT and seismic survey at the Fáb-4 steam well. Seismic reflection horizons representing interfaces between the Pliocene, Miocene and Mesozoic formations are incorporated in the MT Bostick-resistivity *vs.* depth (H_{Bos}) pseudosection after time to depth conversion of seismic data

Interpretation of seismic reflection time sections surveyed in the Nagyszénás area provided evidence of a similar accordance between seismic and MT results.

Figure 11A illustrates the migrated time section of seismic profile Or-44 crossing the zone of higher conductivity reservoir rock in NW-SE direction (Fig. 9).

However, seismic signals recorded from beneath the surface of the Mesozoic basement (see horizon M in Fig. 11) are strongly attenuated, the widespread depression of the tectonized rock (F) is still recognizable without doubt in the seismic data in accordance with MT results. Figure 11B presents a "line-interpretation" version for the seismic profile Or-44. Question mark in Fig. 11B refers to the deep seated root of a fault zone caused by compressional strike-slip movements, presumable earlier than Neogene. Other seismic profiles crossing the area of strike slip movements in direction toward SW-NE do not have clearly evaluable information about the deep tectonic zone.

Figure 12 presents the modified structural map of the area incorporating the location of the deep tectonic zones revealed by MT survey including the geothermal reservoir.

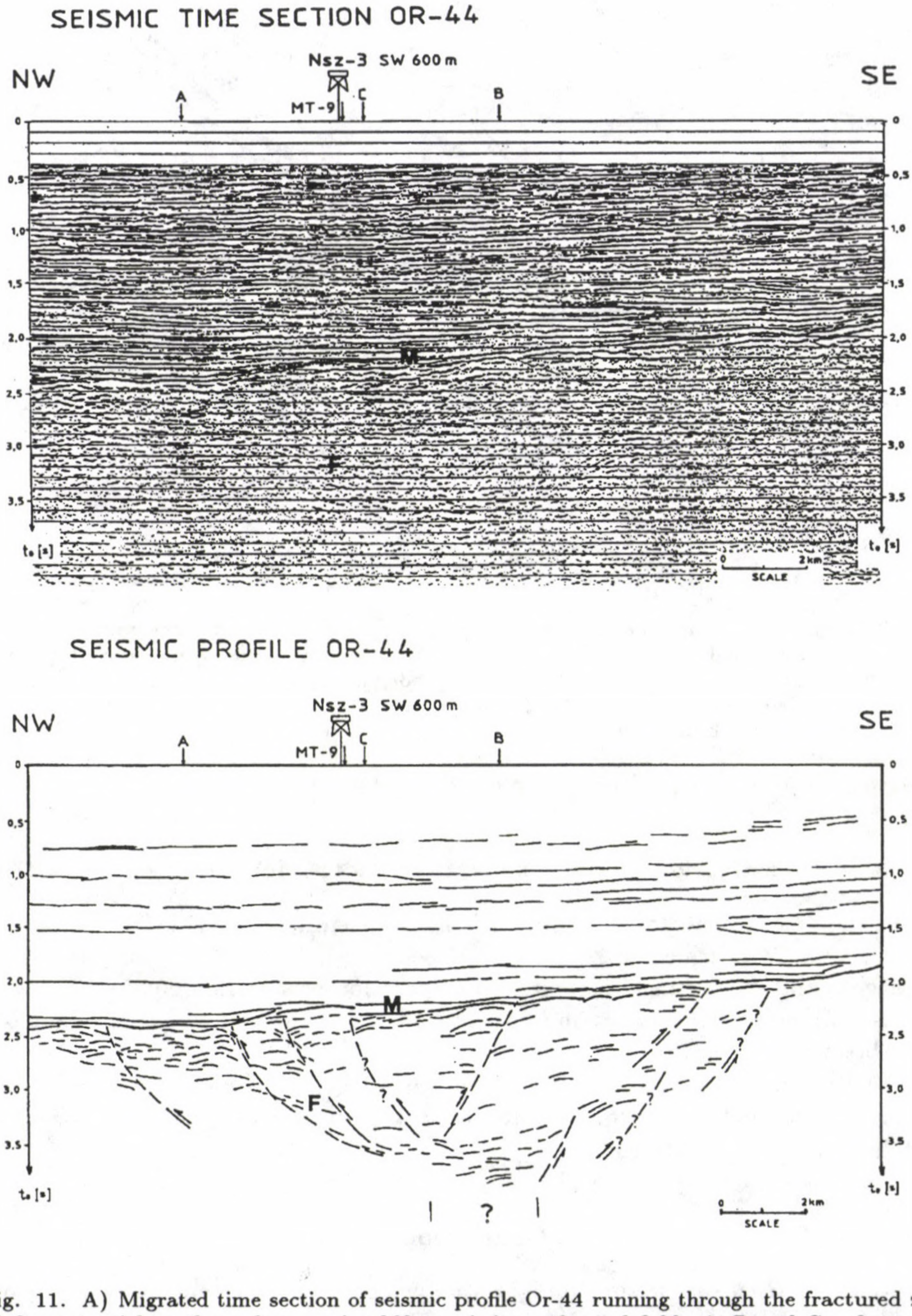


Fig. 11. A) Migrated time section of seismic profile Or-44 running through the fractured zone of the potential geothermal reservoir of Nagyszénás geothermal field. A, B and C refer to the limits of the higher electrical conductivity zone discovered by MT survey as marked on Fig. 9. B) "Line interpretation" version of the time section of Or-44 seismic profile with the same legend as Fig. 11A

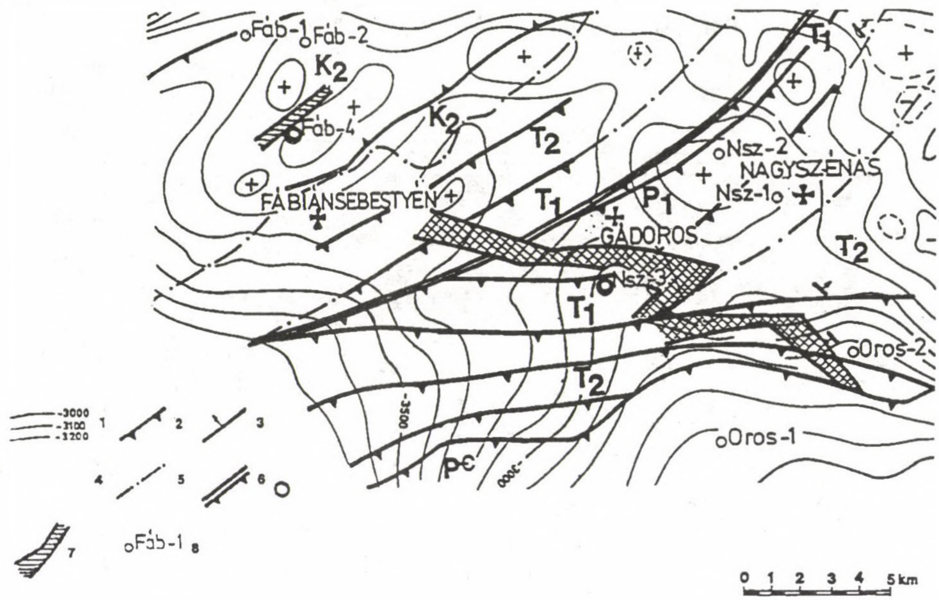


Fig. 12. Structural-tectonical map of Fábiánsebestyén and Nagyszénás areas incorporating both seismic and magnetotelluric survey's data. Legend: 1. Basin depth contour under sea level (m), 2. Overthrust zone, 3. Normal fault, 4. Lithofacies boundary, 5. Regional overthrust zone (the boundary of Villány-Bihar and Szeged-Békés zones), 6. Steam well, 7. Deep tectonic zones including geothermal reservoirs, 8. Exploratory well. K_2 - Upper Cretaceous fractured unit, P_1 - Lower Permian Quartzporfir, P^e - Precambrian-Paleozoic metamorphites, T_1 - Lower Triassic Sandstone, Conglomerate, Mudstone, T_2 - Middle Triassic Dolomite-breccia

Investigation of the Nagyszénás geothermal reservoir

On the basis of integrated results of MT and seismic methods MOL Hungarian Oil and Gas Co. decided to reopen the closed borehole Nsz-3 on the purpose to investigate the geothermal reservoir. This operation was carried out in 1991 with success. A long term investigation was accomplished in the well reopened in a depth range between 3038–3492 m under the surface.

The well produced steam with hot water with a maximum temperature of 185.47°C. Although, the salinity of the fluid is high, the well can be utilized for electrical energy production.

Conclusions

1. The existence of a presumed high enthalpy geothermal system has been proved true in the Pannonian Basin, in South-East Hungary.

2. Magnetotellurics has been found as a sensitive method even in detection of deep goethermal reservoirs.

3. Results of MT surveys confirmed the methodological expectations outlined on the basis of 2D numerical modelling.

4. Suitable integration of MT and seismic results makes an efficient technique utilizable in further prospecting for geothermal reservoirs in the Pannonian Basin or elsewhere.

References

Ádám A, Nagy Z, Varga G 1989: *Geophysics*, 54, 795–797.

Bostick F X 1977: A simple almost exact method of MT analysis. Workshop on Electrical Methods in Geothermal Exploration. USGS Contract No. 14080001-8-359

Goldberg S, Rotstein Y 1982: *Geophys. Prosp.*, 30, 211–216.

Haas J 1987: Outlines of structural geology of Hungary (in Hungarian). In: Structural evolution of the basement of the Great Hungarian Plain, 9–16, MTA SzAB Szeged, Hungary

Hughes L J, Maas J L 1987: Geothermal reservoir delineation and monitoring with CSAMT. Paper presented at 49th EAEG Meeting, Belgrade, 1987

Jones A G 1983: *Journal of Geophysics*, 53, 72–73.

Nagy Z, Beke B, Gulyásné-Formán Cs, Hajdu Gy, Landy I 1990: Results of 3D Electromagnetic Surveys with High Resolution in Hungary. Proceedings of 35th International Geophysical Symposium, Varna

Otten I, Mussmann G 1985: *Geothermics*, 14, 673–688.

Rumpler J ed. 1987: Research for deep-seated high enthalpy geothermal reservoirs in Hungary (in Hungarian). KFH Report, 201/86, Budapest

Stanley W D, Boehl J E, Bostick F X, Smith H W 1977: *J. Geoph. Res.*82, 2501–2514.

Stegena L 1985: In: Proceedings of 19th Petroleum Itinerary Conference, OMBKE, Budapest, 212–216.

Stegena L 1990: In: Proceedings of 21th Petroleum Itinerary Conference, OMBKE, Budapest, Vol. 2, 168–177.

Stegena L, Horváth F, Landy I, Nagy Z, Rumpler J 1991: High Enthalpy Geothermal Reservoirs in Hungary. Manuscript submitted to *Geothermics*

Thanassoulas C 1991: *Geoexploration*, 27, 321–350.

Wright P M, Stanley H W, Ross H P, West R C 1985: *Geophysics*, 50, 2666–2699.

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 103–110 (1992)

AMT STUDIES IN THE GEOTHERMAL AREA OF THE TUCUMAN PLAIN

A M OSELLA[1], M C POMPOSIELLO[2], A MAIDANA[3], E BORZOTTA[3]

Highly enhanced electrical conductivity profiles have been obtained from previous MT soundings in the geothermal area of the eastern border of the Andes at the Tucuman Province in the north of Argentina. These results are associated with geothermal gradients three or four times the normal ones, this characteristics being due to the presence of hot springs and aquiferous hyperthermal layers. In order to have a better description of the shallower layers, previous AMT soundings performed in July 1990 were completed during 1991. So, in the present work all the AMT data are analyzed and the resistivity profiles are estimated.

Keywords: aquifers; audiomagnetotellurics; geothermal area; Tucuman Plain

Introduction

The aim of our work was to study the geothermal area of the Tucuman plain (see Fig. 1). This plain is an extensive depression bounded to the west by the Aconquija range and to the east by Guasayan range. Both are mainly formed by rocks, possible of Precambrian age, which are unconformably overlaid by Tertiary rocks of continental origin (see Fig. 2).

There is an important thermal area in Tucuman and Santiago del Estero which is known through surface manifestations and drillings.

A great number of hot water bearing boreholes allowed to establish the limits of an important thermal district. Inside this area there are some hotter spots (e.g. site 3, see Fig. 2) where temperatures of about 51°C have been obtained in a 300 m deep borehole. In all the studied boreholes, thermal gradients around twice the expected values were obtained, indicating the presence of an important thermal anomaly (Mon and Vergara 1987). Since in this area there is no vulcanism, the presence of a heat source have to be related to deep tectonic and magmatic activity.

Deep MT soundings performed by Baldis et al. (1983), Vatin-Perignon et al. (1985) and Pomposiello et al. (1991) show the presence of a conducting layer at around 8 km depth which may indicate the presence of that heat source.

Taking into account these results together with the ones obtained by gravimetry, seismic and hydrogeology, the model shown in Fig. 3 was proposed (Mon and

[1]Dto. de Física, Fac. Cs. Exactas y Naturales, Univ. de Buenos Aires, Ciudad Universitaria, Pab. 1, 1428 Buenos Aires, Argentina

[2]Centro de Investigaciones en Recursos Geológicos, CONICET, Ramirez de Velazco 847, 1405 Buenos Aires, Argentina

[3]Instituto de Investigaciones Aplicadas a las Ciencias Espaciales, C.C. 131, 5500 Mendoza, Argentina

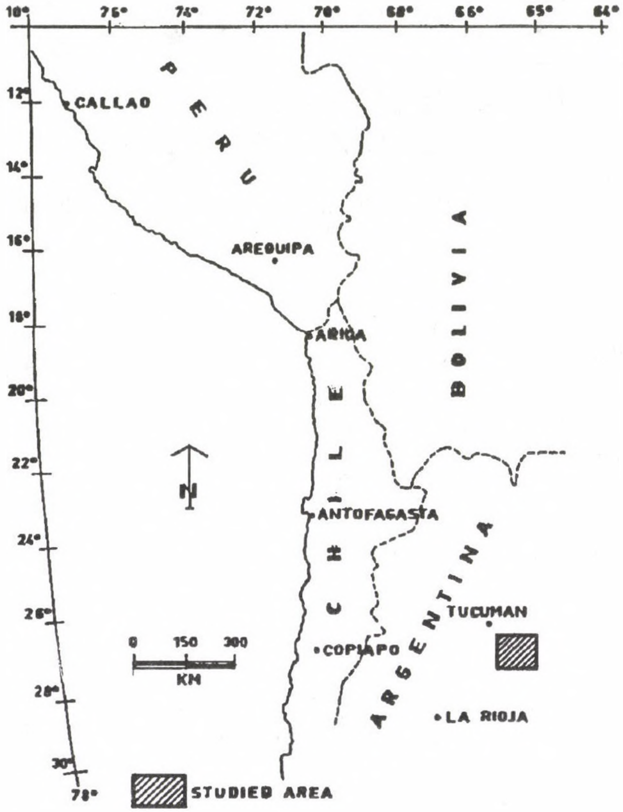


Fig. 1. Location map showing the region studied by AMT method

Vergara 1987). The sedimentary basin coincides with a tectonic depression and is limited to the east by Guasayan range and to the west by Aconquija range. The west border is constituted by faulted basement blocks. From gravimetric and magnetotelluric studies (Pomposiello et al. 1991a, b) a depth of approximately 3000 m has been obtained for the deepest part of the basin. Rains in the border of the Aconquija exceed approximately 1800 mm yearly. This excess water feeds the aquifers. The regional slope (500–600 m over sea level to the west and under 300 m to the east) makes water to flow to the east up to the meeting with the east border of the basin (Guasayan range) which acts as a barrier for the subsurface water. Due to the presence of the heat source, the temperature of the aquifers increases, giving rise to hot springs.

The analysis of water boreholes seems to indicate that the highest geothermometric values are aligned parallel to the regional structure near the east border of the basin. But aquifers in the surrounding area may not be underestimated, being this information of economical and also of social interest for the area.

Considering all the preceding facts, we made AMT studies in the area, beginning with the sites near known boreholes, to determine if the method could be applied to estimate the depth and thickness of the aquifers. From the measured electric and magnetic fields the apparent resistivity curves in both the north-south and east-

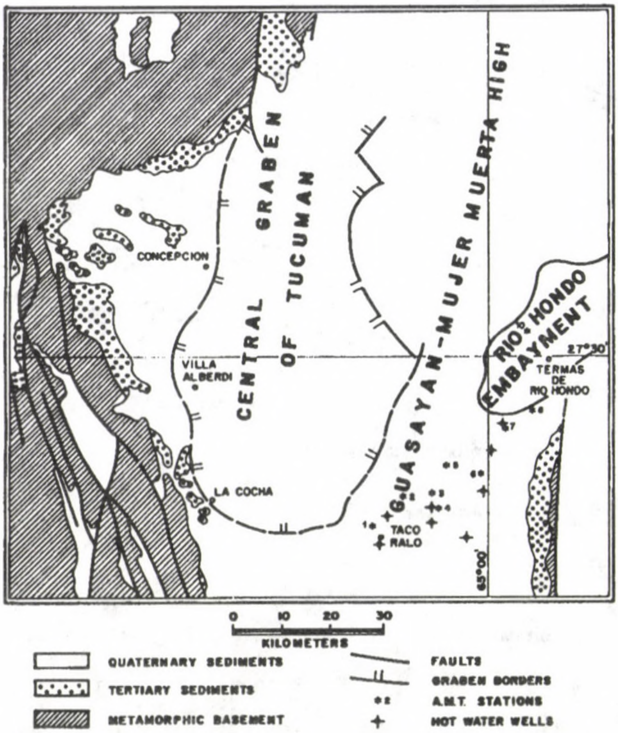


Fig. 2. Geological map of the Tucuman Basin and the distribution of AMT stations

west directions were obtained and the results were modelled to estimate the depth and thickness of the conducting layer. Finally these results were compared with the depth of the boreholes, to determine the efficiency of the method.

Data

As information is wanted about at most the first thousand meters, the higher range of the apparent resistivity spectra is to be interpreted.

We measured along a profile which includes the principal hot water boreholes (sites 1 to 8, Fig. 2) and we completed the work with two soundings in resistive areas (site 9 and 10, Fig. 2) to have an estimation of the electrical resistivity of the basement.

Data were acquired with an AMT equipment of variable frequency in the range 10–1000 Hz. The magnetic field was detected with an induction magnetometer of flat response in this range of frequency. The electric field is obtained by measuring the voltage between pairs of inoxidable steel electrodes.

Since only one horizontal component of the electric and magnetic fields can be measured simultaneously, no rotation to the principal axis can be performed. To

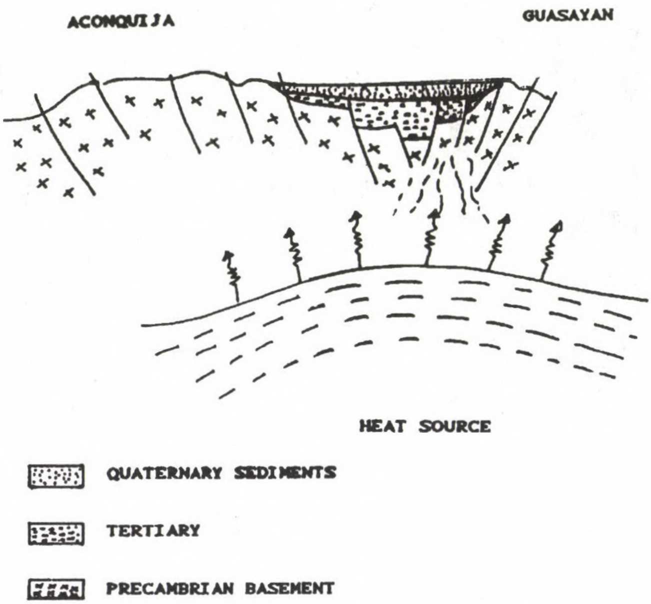


Fig. 3. Qualitative model of the south-east area of the Tucuman Basin

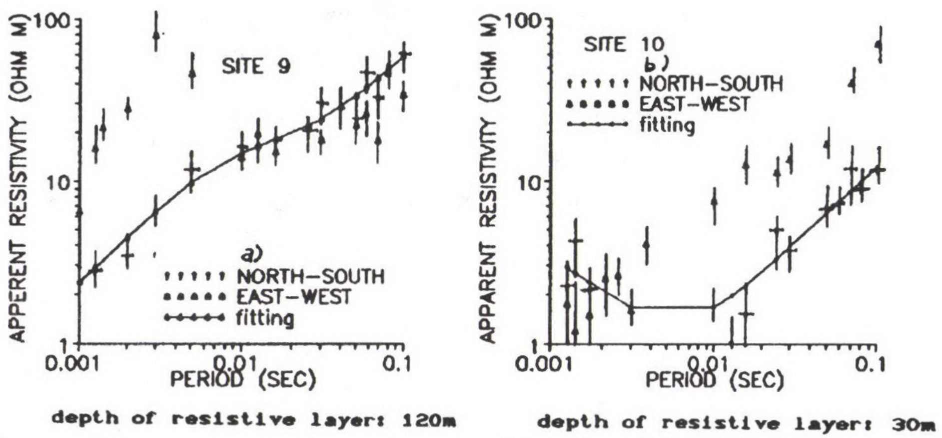


Fig. 4. a, b. Apparent resistivity curves together with the best theoretical fitting for sites 9 and 10, respectively. The estimated depth of the resistive layer is shown

overcome it, we used previous results to interpret the data. MT soundings were performed near sites 1 (Baldis et al. 1983), 10 (Vatin-Perignon et al. 1985), 3 and 9 (Pomposiello et al. 1991). The low values of the skew ($<.3$) and of the strike ($5°$) at the highest frequencies ($10\ \mathrm{sec}^{-1}$) obtained from the calculation of the transfer function indicate that, at least for the first layers, the structure is quite 2D and north-south, the approximate direction is the strike. These results were expected since topographic and shallow layer behavior would be influenced by the presence of the Aconquija and Guasayan ranges.

It is well-known that, when dealing with 2D problems, conducting layers can be

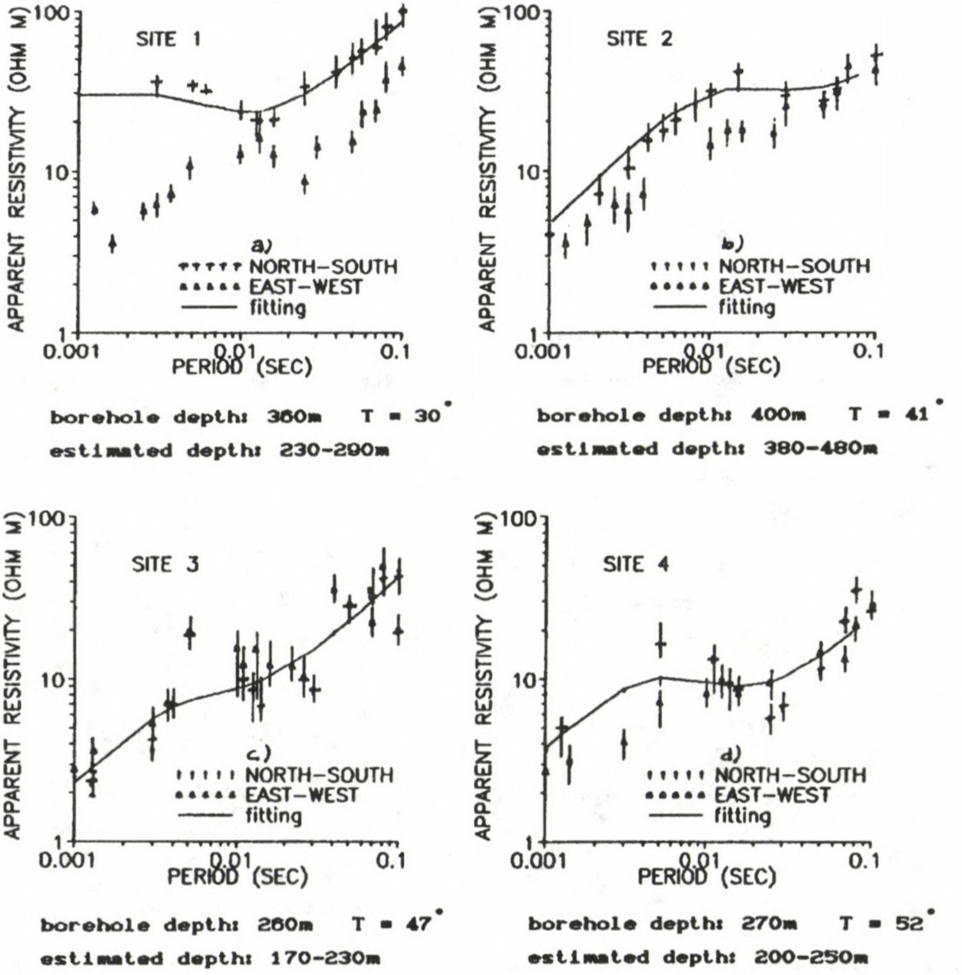


Fig. 5. a. Apparent resistivity curves together with the best theoretical fitting for sites 1 to 4, respectively. Borehole depths and temperatures together with the estimated depth and thickness of the conducting layer are also shown

obtained from the inversion of TE curves. So, we use the north-south component of the impedance tensor and the conducting layer will be estimated by applying 1D inversion method (Jupp and Vozoff 1975) to the apparent resistivity curves. A multilayered isotropic structure was assumed and the electromagnetic response due to the presence of an external uniform field was calculated as a function of the resistivity and thickness of each layer, these parameters being varied in order to reproduce the apparent resistivity and phase curves simultaneously.

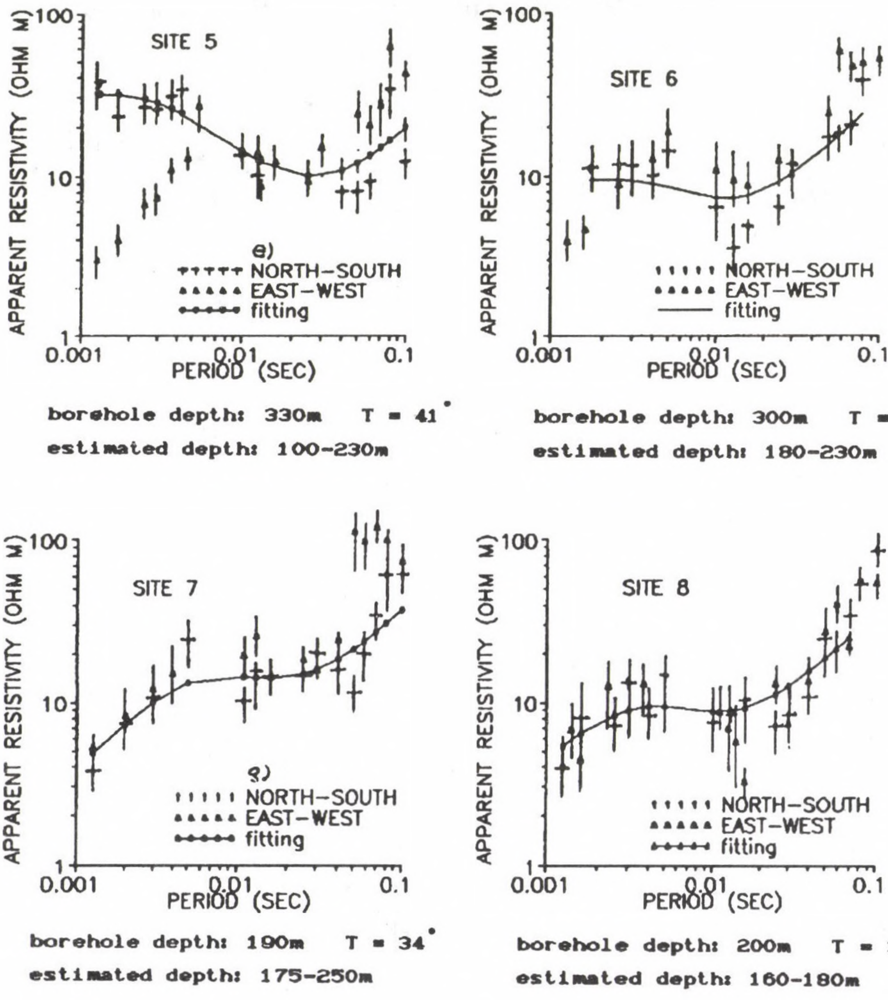


Fig. 5. b. Apparent resistivity curves together with the best theoretical fitting for sites 5 to 8, respectively. Borehole depths and temperatures together with the estimated depth and thickness of the conducting layer are also shown

Results

Figure 4 (a, b) shows the apparent resistivity curves for sites 9 and 10, respectively. These data show a very resistive subsurface, the value of the resistivity being approximately 300 Ωm at 100 m at site 9 and at 30 m at site 10. This resistivity value was assumed for the basement as starting point for the inversion of the data corresponding to sites 1 to 8.

Figure 5 (a to h) shows the apparent resistivity curves for sites 1 to 8, respectively, together with the best fitting. In this figure the depth and thickness of the conductive layer together with the depth and temperature of the water borehole (Vergara, personal communication 1991) are also shown.

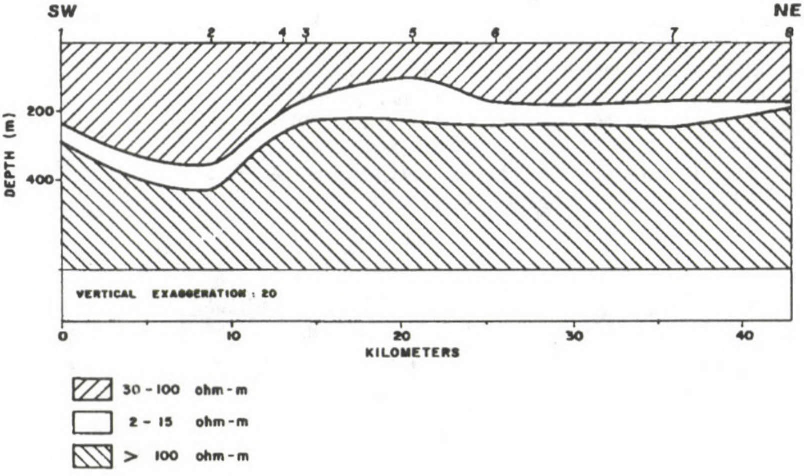


Fig. 6. Resulting resistivity model for sites 1 to 8

Figure 6 shows the resulting resistivity model.

Almost all of the curves have quite a 1D behavior at least in this range, and corresponding to this fact, close estimations of the depth of the aquifers are obtained. In sites 1 and 5, and less at site 6, the east-west component go apart from the north-south one for the lowest periods; in these cases the depth of the conducting layer obtained from 1D inversion of the north-south component (TE mode) is underestimated (see e.g. Osella and Martinelli 1991). This fact would explain why the depths of the boreholes are slightly greater than the results obtained from AMT data.

Conclusions

The geothermal area has great importance both from economical and social aspects. Though many aquifers are known, we think there are great possibilities of finding hot water in the neighbouring zones.

The AMT method has proved to give good estimates of the depth and thickness of the conducting layer associated with these aquifers. So, it could be used to detect new springs and to plan new drillings.

Acknowledgements

This work was supported by CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas).

We thank Dr G Vergara for providing us with the data about water boreholes.

References

Baldis B, Demicheli J, Febrer J, Fournier H, García E, Gasco J, Mamaní M, Pomposiello M C 1983: *J. Geomag. Geoelectr.*, 35, 609–621.

Jupp D L, Vozoff K 1975: *Geophys. J. R. astr. Soc.*, 42, 957–976.

Mon R, Vergara G 1987: *Bulletin Int. As. Eng. Geolog.*, 35, 87–92.

Osella A M, Martinelli P 1991: *Rev. Geof. Bras.* (in press)

Pomposiello C, Mon R, Diaz M T 1991a: *Geodynamique* (submitted)

Pomposiello M C, Osella A M, Maidana A, Borzotta E 1991b: *J. South Amer. Earth Sc.* (submitted)

Vatin-Perignon N, Pomposiello M C, Mamaní M, Maidana A, Keller M, Gasco J, Fournier H, Febrer J 1985: *Acta Geod. Geoph. Month. Hung.*, 20, 159–163.

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 111–119 (1992)

UPPER CRUSTAL HIGH CONDUCTIVITY LAYER IN THE INTRAPLATE SEISMIC PROVINCE OF NE-BRAZIL REVEALED BY MAGNETOTELLURICS

A L PADILHA[1], Í V J M DA COSTA[1], N B TRIVEDI[1]

A magnetotelluric survey consisting of fifteen high-frequency AMT and six low-frequency MT soundings was carried out in the intraplate seismic province of north-east Brazil. A NW-SE profile with nine AMT and four MT stations was positioned to cross the seismic zone of João Câmara (5.5°S, 35.7°W) which is the easternmost swarm of the NE province. Two conductive layers were derived from 2D modelling of the TM-mode of electromagnetic wave propagation: 1. a shallower one (depth of 10 km) in the NW-end of the profile with conductance of 60 S, and 2. a deeper one (at 30 km) in the SE-end. Preliminary interpretation of these results suggests that the crustal conductivity structure of the region could be controlled by unexpected temperature differences between conterminous crustal blocks. In the western block are concentrated the principal upper crust geophysical and geological inhomogeneities and the most important recent shallow-depth seismic activity. The 10 km deep conductive layer in this block could be possibly associated with a ductile layer which limits the maximum depht of the earthquakes.

Keywords: Brazil; geotectonics; geothermal anomaly; high conductive layer; intraplate seismicity; magnetotellurics

Introduction

Northeast Brazil has been historically marked as a region of recurrent seismic events of low intensity (Berrocal et al. 1984). Studies of recent activity indicate that earthquakes normally occur in the form of long swarms in well-defined localities, mostly about the borders of the Potiguar Basin. They exhibit no apparent relationship with surface features (Takeya et al. 1989). Common seismic characteristics include shallow hypocenter distribution depths and subvertical focal mechanisms that indicate strike-slip motions with a minor normal component (Assumpção et al. 1989). Site to site variations in the P and T axes suggest that the regional stress field is probably not uniform. The proposed models to explain these intraplate earthquakes include either the reactivation of preexisting features in response to a regional stress field (Assumpção 1990) or the existence of localized stress amplifications around crustal inhomogeneities (Moreira et al. 1990). Lack of corroborative geophysical data have hindered further tectonic interpretations. As will be shown in this paper, magnetotelluric investigation may fill this gap by providing insight into prevailing crustal resistivity conditions.

[1]Instituto Nacional de Pesquisas Espaciais - INPE, C.P. 515, 12201 São José dos Campos, SP, Brazil

Geologically, the seismicity concentrates in the northern part of a reworked and sheared Precambrian terrain intensely affected by the Brazilian tectono-thermal event (Upper Proterozoic to Ordovician). The region was tectonically active during the Mesozoic and Tertiary. A geotectonic and geological sketch of the northern part of the Brazilian northeastern region, together with the distribution of recent major earthquakes, is displayed in Fig. 1.

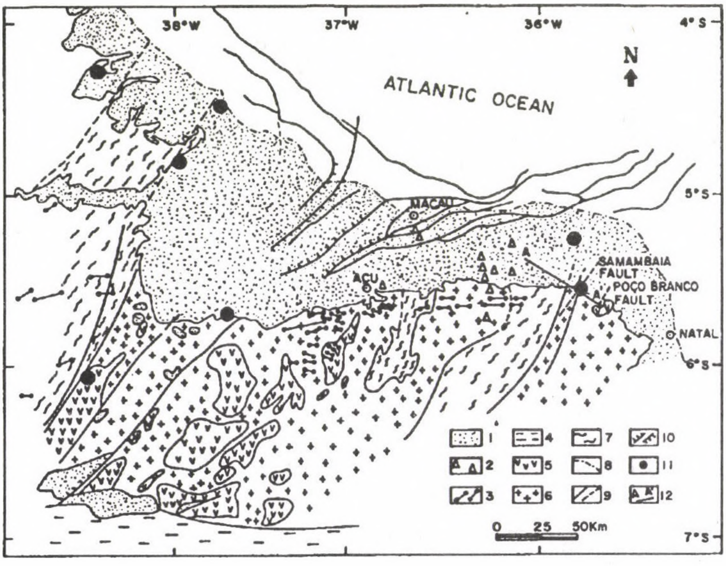


Fig. 1. Geotectonic and geological sketch of the Potiguar Basin in Northeastern Brazil (partially modified from Cordani et al. 1984). 1 – Mesozoic-Cenozoic sedimentary cover; 2 – Tertiary alkaline rocks; 3 – Mesozoic diabase dikes; 4 – Precambrian gneisses and migmatites; 5 – granitic massifs of Brazilian age; 6 – Archean gneisses and migmatites reworked during Brazilian cycle; 7 – Brazilian folded belt; 8 – coastal line; 9 – faults, fractures and lineaments (dashed when inferred); 10 – normal faults; 11 – earthquakes with magnitudes greater than 4.0; 12 – MT profile

The present study was concentrated in a small area around the town of João Câmara (5.5° S, 35.7° W) on the southeastern border of the Potiguar Basin, where seismicity has been regularly recorded by local stations since 1986 (Ferreira et al. 1987, Takeya et al. 1989). The activity occurs in a narrow N40E linear zone, east of João Câmara. This seismic zone, named Samambaia Fault, exhibits no correlation with any surface geological structure in spite of the shallow focal depths (1 to 8 km) and the absence of sedimentary cover over sections of the Precambrian basement cut by the seismically inferred fault. With additional seismographic stations, another small swarm was detected at about 12 km to the east and with the same direction as the Samambaia Fault and is presently referred to as the Poço Branco Fault.

In order to obtain a preliminary evaluation of the electrical conductivity structure of the João Câmara area, a short magnetotelluric campaign was carried out in August, 1989, which consisted of 15 high-frequency (8 to 4500 Hz) scalar audiomagnetotelluric (AMT) stations and 6 low-frequency (0.01 to 5 Hz) tensor magnetotelluric (MT) stations. From these soundings, 9 AMT and 4 MT stations compose a roughly NW-SE 52 km long profile traversing the seismically inferred faults (A-A' in Fig. 1).

Magnetotelluric results

Figure 2 shows the MT results for the four stations along the profile A-A'. Apparent resistivities and impedance phases rotated to the principal directions of impedance tensor are shown. Average geoelectric strike is N40E, nearly coincident to the major structural features in the area and perpendicular to the profile direction. Available Vertical Electrical Soundings (VES) results (Rebouças et al. 1967) were used to verify static shift effects or galvanic distortions on the MT impedance tensor. The scheme used is very similar to that proposed by Pellerin and Hohmann (1990), with a reference curve at high frequencies being derived from VES studies to correct MT data.

In the apparent resistivity graph, the AMT data obtained at the same sites are included as a check of the compatibility between the AMT and MT information. Error bars represent a confidence interval of 68% (1 standard deviation) for MT data and the dispersion in AMT readings. The sizes of these error bars when used together with the observed scattering of apparent resistivities and phases and multiple coherence results are good indicators of the data quality. Despite the small error bars, the pronounced scatter and relatively low coherences (less than 0.85 in the period range from 2 to 20 s) indicate the presence of local electromagnetic noise which probably originates from electrical transmission lines in the region.

The geological complexity of the crust beneath the sites may be estimated from skew values and separation between the TE and TM apparent resistivity curves. Our results show that these criteria usually become larger in the lower frequencies of the sampled spectra, indicating that the geological structures are complex in the area. From these analyses, we decided to employ a 2D modelling of the MT data. The construction of a 2D starting model for inversion began with a 1D inversion on invariant impedances (Berdichevsky and Dmitriev 1976) of each station. These 1D models were later placed together to construct the 2D starting model using additional information from VES and our AMT soundings. The 2D inversion was then performed on the TM-mode of the electromagnetic wave propagation using an inversion code developed by Jupp and Vozoff (1977). The TM data were chosen because it has been shown by Wannamaker et al. (1984) that 2D modelling of TM results can yield correct results in some 3D geometries.

The final 2D model is shown in Fig. 3, where the most striking features are two conductive layers: 1. a shallow one in a depth range of 7–10 km with a conductance of about 60 S in the NW-end of the profile and 2. a deeper one at depths greater than 30 km in the SE-end. Less conductive structure (conductance of 5 S) occurs in depths of less than 500 m in the seismic zone of João Câmara and the crystalline basement all over the region has high resistivity with values exceeding 10,000 ohmm.

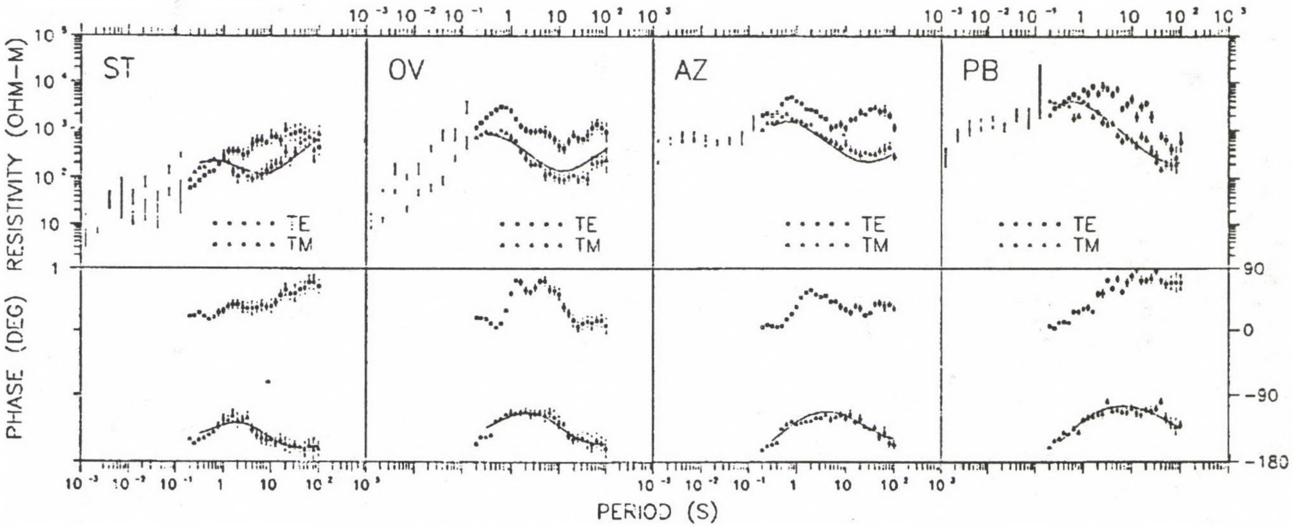


Fig. 2. AMT and MT results for the stations along the profile A-A'. Resistivities and impedance phases for the TE and TM modes are shown by the squares and triangles, respectively, and the continuous lines are the TM responses to the 2D model of Fig. 3. Vertical error bars without symbols are scalar AMT data

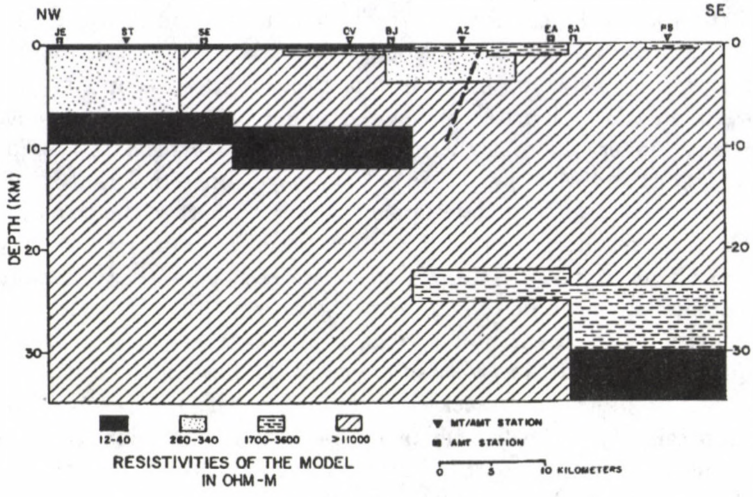


Fig. 3. Two-dimensional electrical model for the study section obtained from inversion of TM-mode. Location of the Samambaia Fault (heavy dashed line) is also included and the letters at the top are for station identification

Crustal conductive layers

There are many possible explanations for conductive layers in the earth's crust (e.g., Parkhomenko 1982). Fluids circulating between interconnected pores of rocks have been widely considered as one of the most probable sources of these layers. Superdeep drill holes in ancient crust (Kozlovsky 1984) give support to this hypothesis. However, there is some discomfort in accepting volatiles in the crust of the study region because of the past intense tectono-thermal activity that affected the area. The loss of volatile elements would probably have produced a drier mafic crust which would be resistive. Also, alternative hypotheses, such as thin graphite films (e.g., Frost et al. 1989), must be considered, specially to explain the easternmost deeper conductive layer.

Conductivity distribution in the area seems most probably controlled by major lateral temperature differences. This hypothesis is partially confirmed by available geothermal flux data for northeast Brazil (Carneiro et al. 1989). In spite of the small number and unfavorable distribution of measurements (most concentrated in sedimentary basins), they indicate very high geothermal flux values in the region, specially to the west of João Câmara.

MT data can then be used to give some insight on this temperature distribution. They point out the possible presence of higher isotherms to the west and lower isotherms to the east of João Câmara, characterizing two very distinct crustal blocks. The western block is dominated by rocks from the Brazilian fold belt and, in addition to the intense seismic activity, is characterized by geological and geophysical inhomogeneities. These include Tertiary volcanics and geophysical anomalies with gravity variations and the near-surface conductive layer detected by our MT survey. In the eastern block, though granitoids outcrop on Archean gneisses and migmatites, there are no significant geophysical variations nor seismic activity.

Considering the constraints of these results on the local seismicity, the relationships between ductile and brittle behaviour and maximum stresses sustained by the upper crust under different thermal conditions have been discussed by Chen and Molnar (1983) and Meissner (1986). They have concluded that high heat flow reduces the thickness of the brittle upper crust with a resulting concentration of stresses. Our MT results may then be indicative of a thinner brittle layer to the west of João Câmara, exactly the region where seismic activity is concentrated. Westernmost shallower conductive layer in Fig. 3, which has the same depth range as the maximum depth of earthquake foci, could therefore be associated with parameters linked to a ductile, low viscosity layer at the base of the brittle upper crust as in the model proposed by Jiracek et al. (1983) for the Rio Grande Rift.

Another possibility to be further investigated is the relationship between this shallower conductive layer and the Tertiary alkali-basaltic magmatism. Mapping the horizontal extension of this conductive layer may confirm a spatial coincidence with the Tertiary magmatism which is found mostly concentrated within 36° and 36°30'W longitudes. Unfortunately, the present study covers only the eastern border of the magmatism.

Referring to the easternmost deeper conductive layer, it should be mentioned the coincidence of its depth and that of the Moho discontinuity in the Brazilian northeast continental margin, obtained from detailed seismic surveys (Houtz et al. 1977, Cloetingh et al. 1979). This coincidence is very intriguing since the Moho normally can not be mapped with electrical methods. On the other hand, the absence of this conductive layer to the west of João Câmara can be attributed to screening effects generated by both the highly conductive surficial sediments of Potiguar basin and the shallower western conductive layer.

Geotectonic implications

Conductive crustal layers are usually present in tectonically active regions (e.g., Garland 1981), although they also have been detected in stable crustal environments (e.g., Van Zijl 1977). In northeast Brazil, available geophysical information reveals the possibility of recent tectonic activity in the region. In addition to shallow seismicity and a high conductivity layer within the upper crust, there is also evidence of anomalous geothermal flux. On the other hand, analysis of data recorded by a 9-station local seismic network detected a simple velocity structure normally associated with shield areas (Takeya et al., 1989). Gathering additional geophysical data will be necessary to settle the actual geotectonic situation.

In reference to the proposed models to explain the regional seismicity, one of them (Assumpção 1990) suggests a superposition of a plate-wide regional E-W compression with local extensional stresses (due to density contrasts across continental margins and sediment loading at the marginal basins). However, as pointed out by the author, this model would imply that the seismicity would occur along the whole northern coast and not only at Potiguar basin borders. The model can be considered consistent only if the reduction of the thickness of the brittle crust (with consequent amplification of stresses), as derived from our MT traverse, is concentrated just beneath the Potiguar basin.

The other model was based on the analysis of gravity data. Figure 4 shows the gravity anomaly residual map for the northeastern Brazilian region presented by Moreira et al. (1989). A small positive gravity anomaly centered on the Samambaia fault was interpreted by the authors as being caused by a shallow but not outcropping intrusive mafic body. Local earthquakes foci are concentrated on the western side of this heterogeneity. Consequently, Moreira et al. (1990) proposed that regional seismic activity would be generated by localized stress amplifications around crustal density inhomogeneities. Again, a thinner brittle crust would represent an additional stress concentration making this model possible. However, to conclude for its general validity it will be necessary to verify if the same geophysical characteristics (local gravity anomalies) are present at the other areas where recurrent seismic activity is observed.

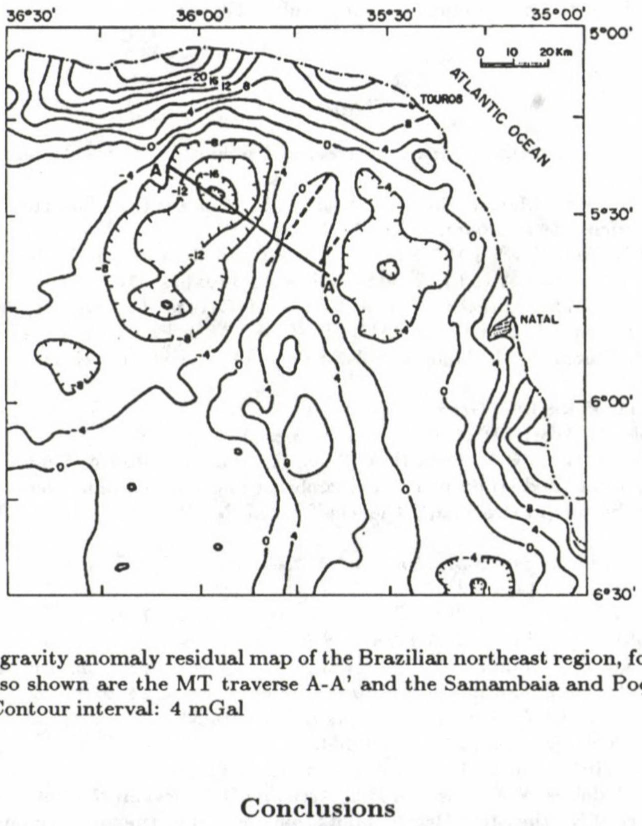


Fig. 4. Bouguer gravity anomaly residual map of the Brazilian northeast region, following Moreira et al. (1989). Also shown are the MT traverse A-A' and the Samambaia and Poço Branco faults (dashed lines). Contour interval: 4 mGal

Conclusions

The presence of an upper crustal conductive layer possibly associated with a ductile layer which limits the maximum depth of the earthquakes and the identification of two crustal sections with different geological and geophysical characteristics are the major findings of this work. In particular, the site of the more important recent seismic activity (the Samambaia Fault) may mark the boundary between the two blocks.

The geophysical results in the area are important enough to justify further study. Continuity of the profile to the west, to determine the extension of the shallower conductive layer, and densification of soundings in the João Câmara region, to acquire a better understanding of what occur in the transition zone between the two crustal blocks, are necessary additional work. Such studies might permit the evaluation of the different hypotheses for the origin of the geoelectrical layers.

Acknowledgements

We are grateful to Adauto G Motta and the personnel of the INPE-Natal for their collaboration during the field work, to Domingos Fonseca for his efforts in making feasible the acquisition of the MT data, to the members of the Physics and Geology departments of the UFRN (Universidade Federal do Rio Grande do Norte) for the previous discussions and access to geological and geophysical information of the study area, and to Naomi Ussami and Marcelo Assumpção for the discussions about the interpretation of the results. This reasearch was partially supported by CNPq (Proc. 400.429/89-6).

References

Assumpção M 1990: The regional intraplate stress field in South America. Unpublished Manuscript. Univ. of São Paulo

Assumpção M, Ferreira J M, Carvalho J M, Blum M L, Menezes E A, Fontenele D, Aires A 1989: *Revista Brasileira de Geofísica*, 7 11–17.

Berdichevsky M N, Dmitriev V I 1976: In: Geoelectric and Geothermal Studies, KAPG Geophysical Monograph, Ádám A (ed.), Akadémiai Kiadó, Budapest, 165–221.

Berrocal J, Assumpção M, Antezana R, Dias Neto C M, Ortega R, França H, Veloso J A V 1984: Seismicity of Brazil (in Brazilian). IAG/USP-CNEN, São Paulo

Carneiro C D R, Hamza V M, Almeida F F M de 1989: *Revista Brasileira de Geociências*, 19, 310–322.

Chen W P, Molnar P 1983: *J. Geoph. Res.*, 88, 4183–4214.

Cloetingh S, Nolet G, Wortel R 1979: *Tectonophysics*, 59, 335–346.

Cordani U G, Brito Neves B B, Fuck R A, Porto R, Thomaz Filho A, Cunha F M B da 1984: Preliminary integrated study of the Precambrian and the tectonic events in the Braizilian sedimentary basins (in Brazilian). Ciência-Técnica-Petróleo, Publ. 15, PETROBÁS, Rio de Janeiro

Ferreira J M, Takeya M, Costa J M, Moreira J A, Assumpção M, Veloso J A V, Pearce R G 1987: *Geophys. Res. Lett.*, 14, 1042–1045.

Frost B R, Fyfe W S, Tazaki K, Chan T 1989: *Nature*, 340, 134–136.

Garland G D 1981: *Ann. Rev. Earth Planet. Sci.*, 9, 147–174.

Houtz R E, Ludwig W J, Milliman J D, Grow J A 1977: *Geol. Soc. Am. Bul.*, 88, 711–719.

Jiracek G R, Gustafson E P, Mitchell P S 1983: *Tectonophysics*, 94, 299–326.

Jupp D L B, Vozoff K 1977: *Geophys. J. R. astr. Soc.*, 50, 333–352.

Kozlovsky Ye A 1984: *Sci. Am.*, 251(6), 98–104.

Meissner R 1986: The continental crust: A geophysical approach. Academic Press, Orlando

Moreira J A M, Medeiros W E, Lins F A P L, Archanjo C J, Ussami N 1989: Gravity map of the northern part of Northeastern Brazil and its contribution to the local tectonics (in Brazilian). 1°. Congresso da Sciedade Brasileira de Geofísica (Abstracts), Rio de Janeiro

Moreira J A M, Baraud R, Lins F A P L, Macedo J W P 1990: Gravity contribution to the understanding of João Câmara seismicity (in Brazilian). 36°. Congresso Brasileiro de Geologia, Natal

Parkhomenko E I 1982: *Rev. Geoph. Space Phys.*, 20, 193–218.

Pellerin L, Hohmann G W 1990: *Geophysics*, 55, 1242–1250.

Rebouças A C, Manoel Filho J, Benoit H 1967: *Brasil/SUDENE/Hidrogeologia*, 15, 53–59.

Takeya M, Ferreira J M, Pearce R G, Assumpção M, Costa J M, Sophia C M 1989: *Tectonophysics*167, 117–131.

Van Zijl J S V 1977: In: The Earth's Crust, Geophys. Monogr. 20. Am. Geophys. Union, Washington, D.C., 470–500.

Wannamaker P E, Hohmann G H, Ward S H 1984: *Geophysics*, 49, 1517–1533.

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 121–129 (1992)

AIRBORNE ELECTROMAGNETIC MEASUREMENTS ON THE WESTERN RIM OF THE VIENNESE BASIN (AUSTRIA)

A Römer[1]

The aim of these measurements has been the mapping of faults and other tectonic features in the thin covered basement of the most western part of the Viennese Basin. A multifrequency and multicoil aeroelectromagnetic-system (AEM) has been used. Forward modelling for two conductive layers has been made comparing them with borehole-data and DC-sounding data (Schlumberger). With the electromagnetic method only, young fault systems, which are not visible on geological maps, have been detected.

Keywords: airborne electromagnetic method; conductive layer; DC-sounding; recent fault system; Viennese Basin

Introduction

Since 1980 airborne geophysical measurements, using a helicopter, have been carried out for the exploration of minerals in Austria (Heinz and Seiberl 1990). Large parts of the Austrian territory are explored with electromagnetic and magnetic methods and measurements of the energy of gamma rays. Apparent resistivity, apparent depth of a conductor, anomalies of total intensity of gamma rays, potassium, uranium, thorium and their ratios and anomalies of earth's magnetic field are interpreted.

In this paper results of an aerogeophysical survey near "Herzogenburg", a town approximately 50 km west of Vienna, are presented.

The area of the survey is located at the western rim of the Viennese Basin in Lower Austria (Fig. 1). The dimension of the area is 15 km × 20 km.

The survey area has been covered with 64 parallel flight profiles (E–W) and 4 control profiles. The distance between the profiles was 200 m. The flight altitude was 80 m above ground.

Equipment

The measuring equipments consists of the following parts:

1. Electromagnetic multifrequency and multicoil system (coaxial (900 Hz), coplanar (3600 Hz)). The 10 m long bird is carried 30 m below the helicopter.

[1]Institute of Meteorology and Geophysics, A-1090 Vienna, Austria

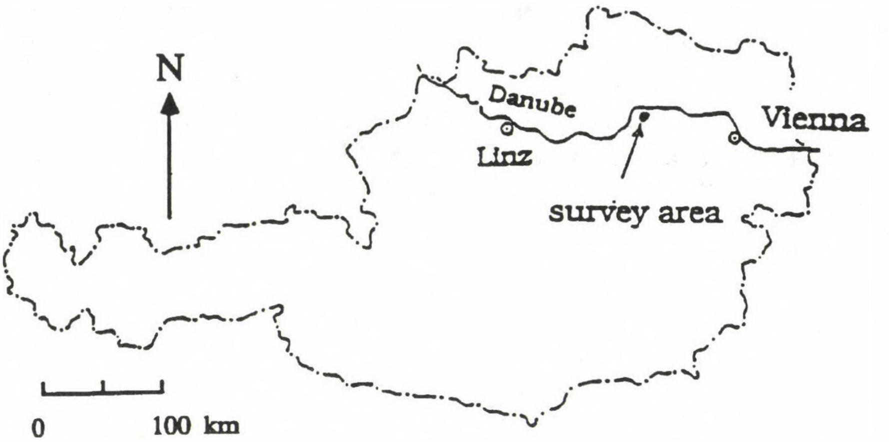


Fig. 1. Location of survey area

2. Gamma ray-spectrometer with two Sodium-Jodid crystals with a volume of 33.6 liters. It measures the intensity of gamma rays between 0.4 and 3.0 MeV in 256 channels.

3. Proton magnetometer with a sensitivity of 0.5 nT. The sensor is carried 20 m below the helicopter.

4. Auxiliary equipment: radar altimeter, Doppler navigation system, barometric altimeter, thermometer *etc.*

Description of the AEM-system and data processing

Airborne electromagnetic measurements are used to explore conductors at the uppermost regions of the earth crust ($\sim$ 200 m). The AEM-system used in this survey is a frequency domain system which is equivalent to the Slingram method. Signals are recorded in mV 10 times a second which are converted afterwards into parts per million (ppm). Accuracy under normal conditions is 1–2 ppm. Because ppm data cannot be interpreted in terms of the geology of the surveyed area, they have to be converted during data processing into parameters which can be correlated with rock types *etc.*

In this investigation the apparent resistivity and apparent depth are calculated from the real — and quadrature — components using the "homogeneous half-space model". To calculate the former parameters, resistivity and depth, it is necessary to make a few simplifying assumptions, e.g. the separation of the coils has to be small in relation to the flight altitude of the bird which is usually guaranteed. Three unknown parameters ρ, apparent resistivity, μ, magnetic permeability and h for the homogeneous half space have to be determined. h is the distance between EM-bird and surface of the homogeneous half space.

From the response function Θ of the homogeneous half-space

$$\Theta = e^{k_1 \ln(\frac{Q}{R}+k_2)} \tag{1}$$

h is calculated from Eq. (2)

$$h = c\left(\frac{A^*}{\sqrt{R^2+Q^2}}\right)^{\frac{1}{3}} \tag{2}$$

using the auxiliary function A^*

$$A^* = e^{k_3 \ln(\Theta+k_4)} \tag{3}$$

Q — quadrature component
R — real component
c — coil separation
k_1, k_2 — empirical constants
k_3, k_4 — empirical constants.

Knowing h it is easy to determine the resistivity (ρ) from Eq. (4)

$$\rho = \left(\frac{1}{\mu\omega}\right)\left(\frac{\Theta}{h}\right)^2 \tag{4}$$

ω — angular frequency of the inducing field.

Generally many rocks show almost no magnetic effects and therefore the magnetic permeability is set to 1.

Finally the flight altitude is subtracted from h to get the apparent depth of the half space below the surface. The buried half space model can be interpreted as a two layer model. The first layer is a layer with an almost infinite resistivity (air). The second layer represents the homogeneous half space. A more detailed description of the calculation of the resistivities and depths can be found at Seiberl and Koehazy (1988).

Geology

The survey area is located at the western part of the Viennese Basin resp. in the "Bohemian Massif". The "Bohemian Massif" is characterised by a few large fault systems striking NE-SW, some being perpendicular to them. West of the survey area the well known and most important fault, the "Diendorfer" fault system is located. It separates gneisses from granulites of the "Dunkelsteiner Wald" (Fig. 2) (Grill 1957, Fuchs and Grill 1984, Matura 1983).

The western part of the survey area is dominated by the granulites of the "Dunkelsteiner Wald" with ultrabasitic occurrences. There are also a few isolated granulitic outcrops surrounded by Tertiary sediments. Two rivers (Fladnitz, Traisen) are crossing the survey area from north to south. Both river basins are filled with clay, sand and gravel. The range of hills at the central part, lying between the two rivers, consists of sands in the south and of marl and gravel in the north. A simplified geological cross section (Fig. 3) shows a succesion of dip-slip faults from west to east. That is also confirmed by the DC-sounding results (Schlumberger).

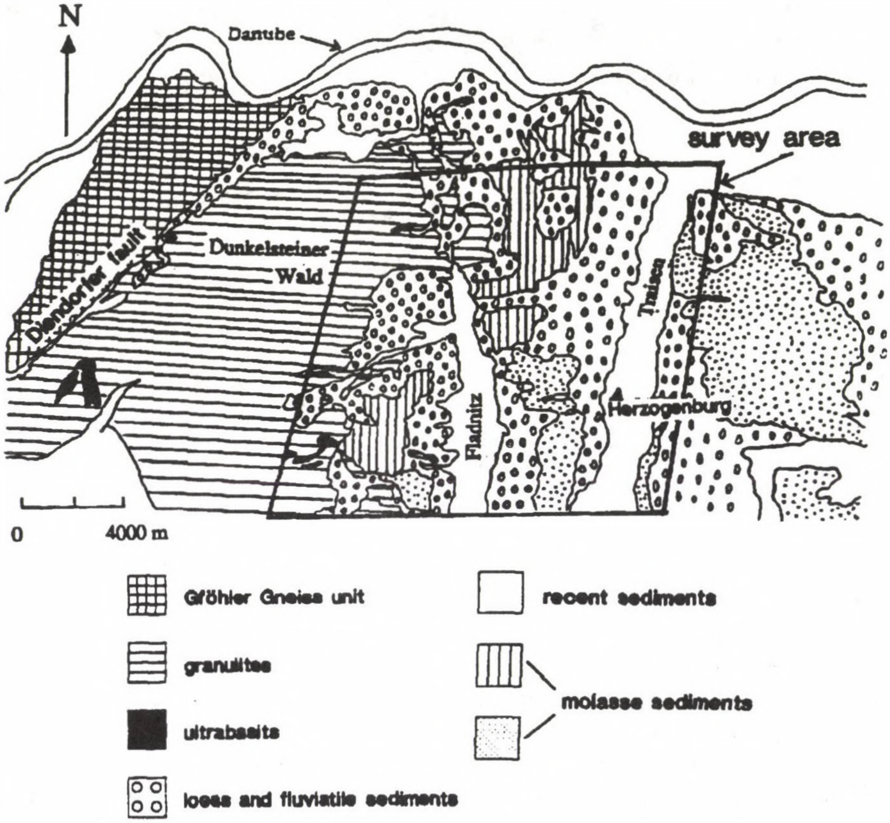


Fig. 2. Geology

Results

Electromagnetics

A remarkable contrast in resistivities between the Tertiary/Quarternary layers of the Traisen-valley and the granulites of the "Dunkelsteiner Wald" can be observed (Fig. 4). In the granulite more than 500 Ωm is measured locally (frequency 3 600 Hz). In the granulites a few relative minima are due to local zones of clay. The young cover of Tertiary sediments shows no large thickness what is also confirmed by the apparent depths and the geosounding results. In the southwestern part of the survey area there area a few granulitic outcrops partly covered by very thin sediments. In Fig. 4, the values of $\rho \sim 100$ Ωm show that the granulitic body is striking NE-SW. Additionally two major ρ-minima, striking also NE/SW, can be observed in Fig. 4. One is crossing the Fladnitz-valley, the other one starts at the upper part of the Traisen-valley. The first one separates the just mentioned granulitic outcrops from the main granulitic massif of the "Dunkelsteiner Wald" in the W. The second

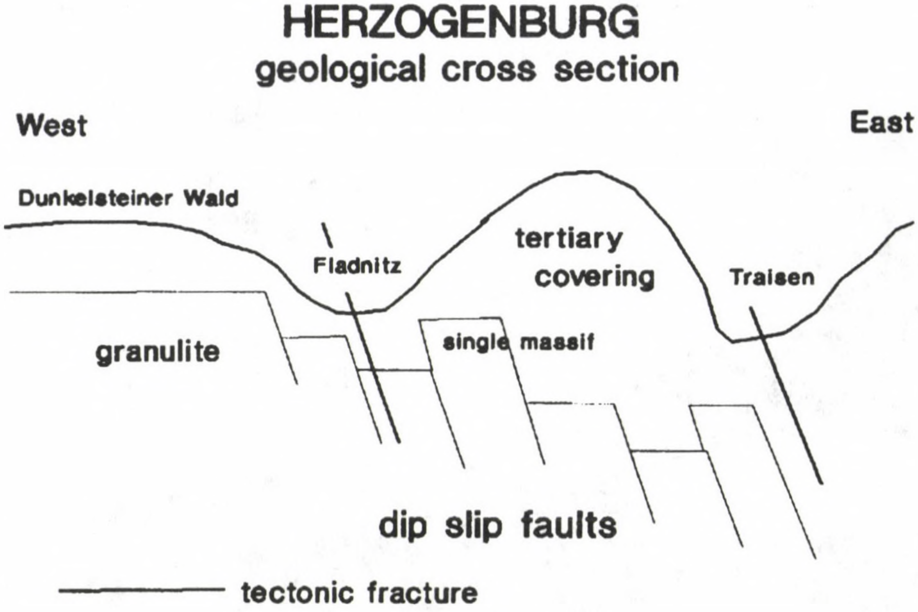


Fig. 3. Geological cross section

one separates the former rocks from the thicker Tertiary/Quarternary sediments in the E. These distinctive ρ-minima are interpreted as young faults lying parallel to the "Diendorfer-fault". Perpendicular to them there are also a few graben systems filled with young sediments. These fault systems causes the dip-slip faults of the granulites from west to east. In a previous survey (Seiberl and Heinz 1985) the "Diendorfer-fault" is also accompanied by zones of lower resistivities. Moreover, Grill (1957) described a so-called "St. Pöltner" fault — a tectonic line — being probably the continuation of the above described NE/SW structures.

In Figs 4 and 5 the two electromagnetic structures, striking NE-SW and the structures perpendicular to them are marked.

In addition the results have been compared with borehole data (Vetters 1925) and geoelectric soundings (Schlumberger method). High correlation has been found. The sounding data have also been inverted with a two layer modell and is in correspondence with the electromagnetic results of the half space model.

Magnetics

There are large anomalies — from west to east — caused by ultrabasites and pyroxenamphibolite accompanying the granulites (Seiberl and Heinz 1985). They are covered by Tertiary sediments. The thickness of the covering increases to the East. A reduction to the magnetic pole has been made (Fig. 6) (earth magnetic field: Inclination I = 63°, Declination: D = 1°, remanent magnetic field I = 63°,

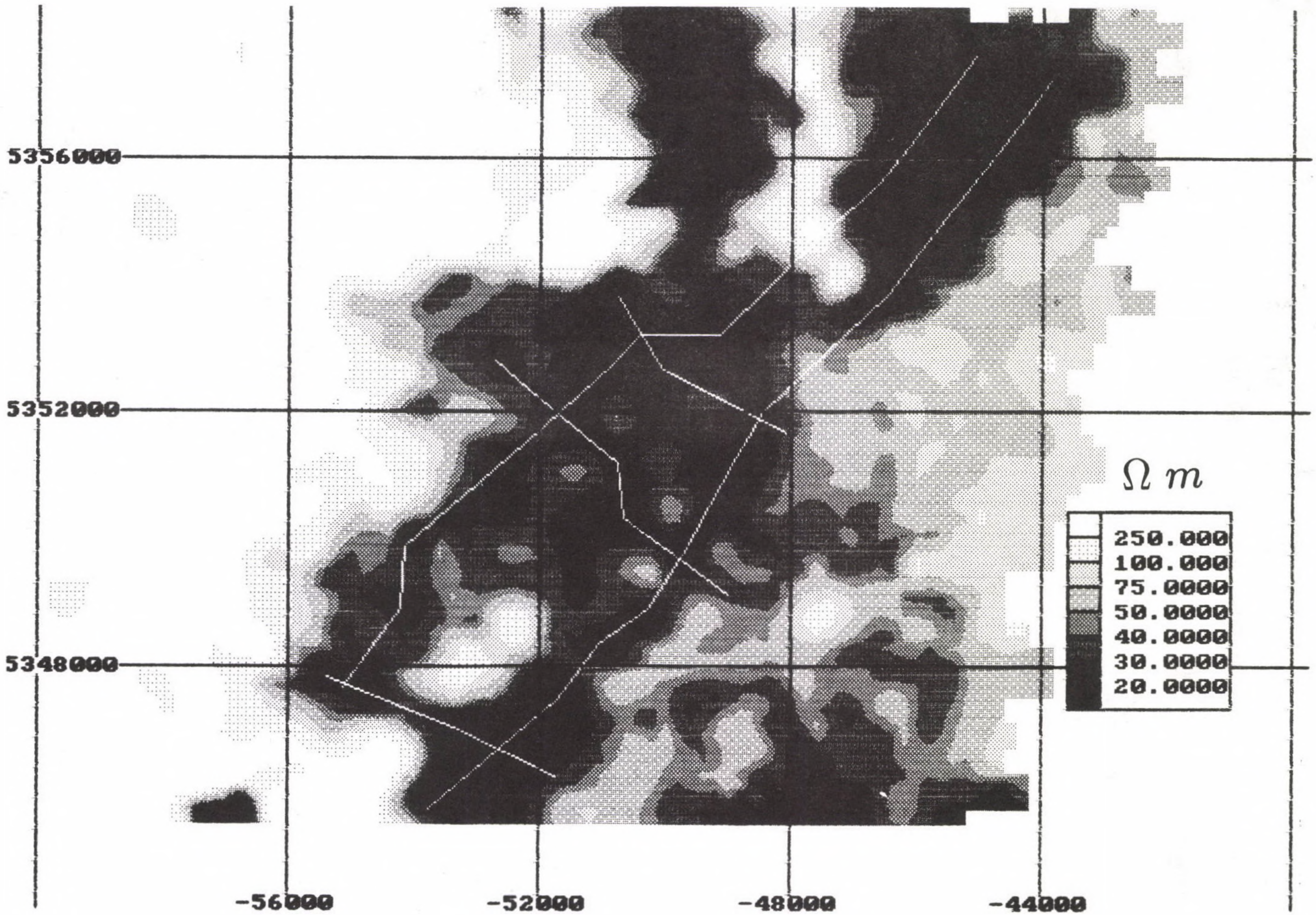


Fig. 4. Apparent resistivity (3 600 Hz) in Ωm

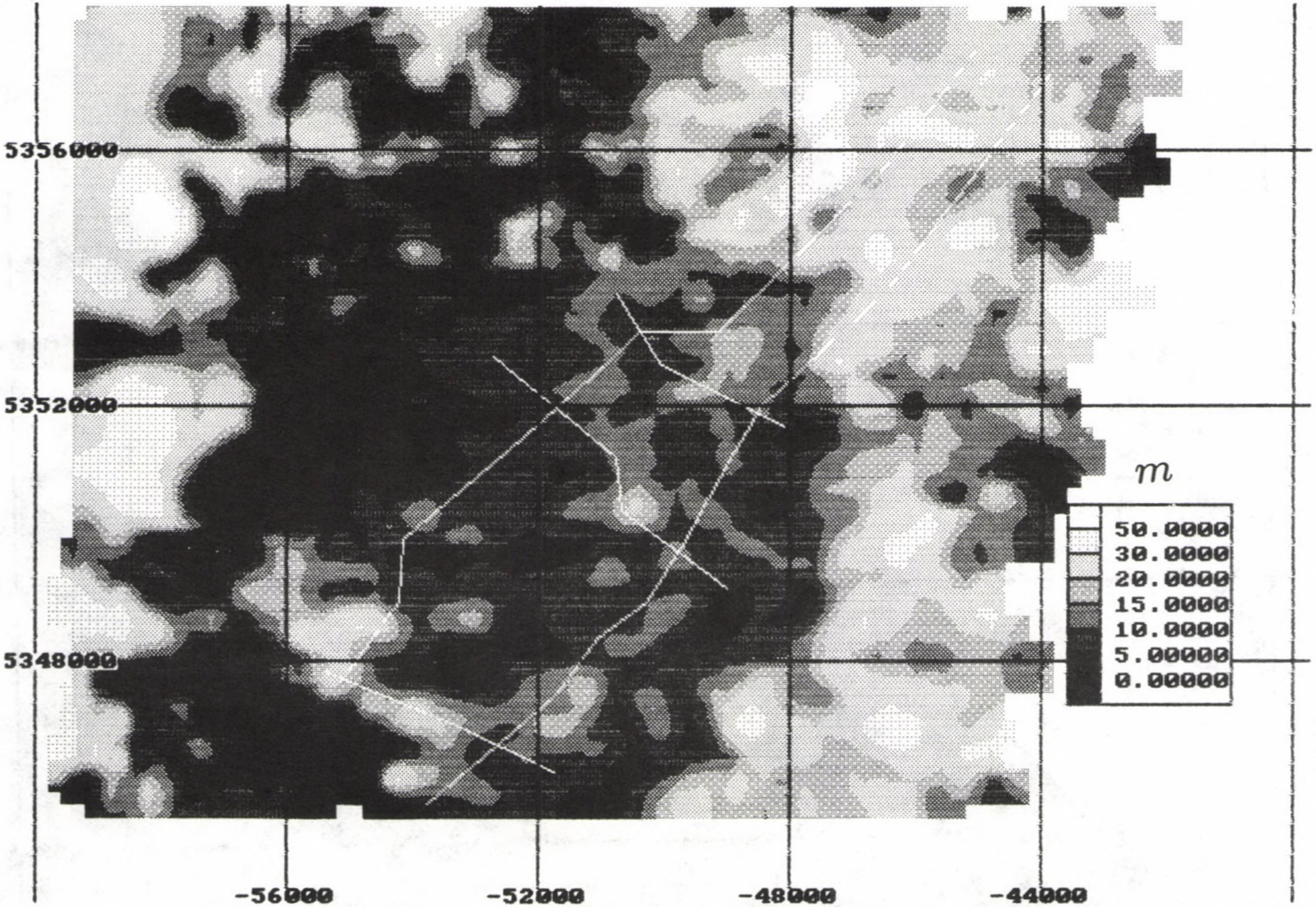


Fig. 5. Apparent depth (3 600 Hz) in m

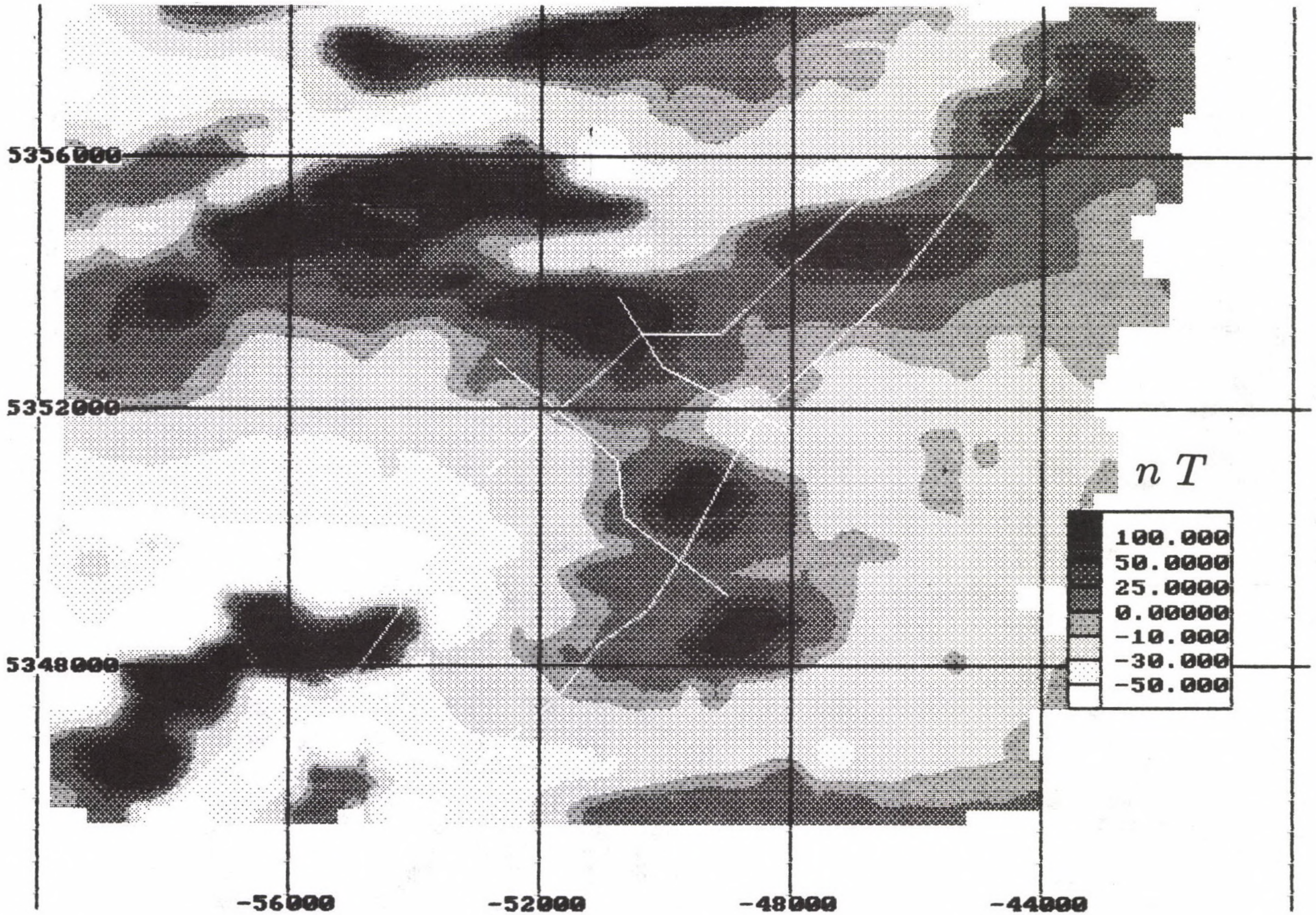


Fig. 6. Reduction to the pole of the magnetic field in nT

$D = 1°$). Comparing the magnetic data with the electromagnetic results the two above mentioned tectonic features can be partly observed in the magnetic trend lines. The map depicts that even the molasse sediments (cf. Fig. 3) have been involved into the activity of the fault systems.

Conclusions

With aerogeophysical measurements, especially with the electromagnetic method, it was possible to detect young recent fault systems in the area of "Herzogenburg". This faults are not visible on geological maps. They are detected even within the very young sediments and therefore are still active. This has to be taken into consideration when estimating existing or planned waste deposits.

References

Fuchs W 1986: Bericht 1966 über Aufnahmen auf den Blätterm Obergraphendorf (55), St.Pölten (56), Spitz (37) und Krems (38) Verh. Geol. B.-A., 1967, A 24–A 26.

Fuchs W, Grill R 1984: Geologische Karte der Republik Österreich 1:50 000. Blatt 38 Krems. Geol.B.-A.

Grill R 1957: Aufnahmen 1956 auf den Blättern Krems an der Donau (38), Obergrafendorf (55) und St.Pölten (56). Verh. Geol.B.-A., 1957, 29–32.

Heinz H, Seiberl W 1990: Bewertung und Problematik aerogeophysikalischer Anomalien im österreichischen Bundesgebiet. Stand: Mitte 1990. Abh. Geol.B.-A., 44, 1–244.

Matura A 1983: Geologische Karte der Republik Österreich 1:500 000, Blatt 37 Mautern. Geol.B.-A.

Seiberl W, Heinz H 1985: Aerogeophysikalische Vermessung im Raum Dunkelsteiner Wald. Proj. Berl., Proj. NC6c/81, 32, ÖAW.

Seiberl W, Koehazy R 1988: Die quantitative Auswertung aeroelektromagnetischer Messdaten (DIGHEM–II). Ber. Geol.B.-A., No. 13, Wien.

Vetters H 1925: Bohrungen im Raum von Satzendorf. Unveröff. Schichtverzeichnisse, Lagerstättenarchiv, Geol.B.-A.

Acta Geod. Geoph. Mont. Hung., Vol. 27(1), pp. 131–138 (1992)

A POSSIBILITY OF AN ELECTROMAGNETIC TECHNIQUE TO LOCATE OIL RESERVOIR BOUNDARIES ON BASIS OF ANALOGUE MODELING EXPERIMENTS

L SZARKA[1], Z NAGY[2]

According to the results of analogue modeling experiments, the boundary of circular oil reservoir models (which usually have higher resistivity than the host rock) can be detected by surface electromagnetic measurements. The method needs one surface- and one lowered current electrode placed in an off-centre borehole. From an areal distribution of the individual electromagnetic field components by a simple normalization procedure, information concerning the boundary of the oil reservoir models can be obtained. In the paper the most promising E_r and H_z analogue modeling anomalies are presented in detail.

Keywords: analogue modeling; electromagnetic methods; oil reservoir; reservoir boundary

Introduction

Boundaries of discovered oil reservoirs, usually remain unknown in spite of they are penetrated by several boreholes. Their identification by electromagnetic methods has been an unsolved problem. In this paper possibilities of a special controlled source electromagnetic method, somewhat similar to the VSP are described on basis of analogue modeling experiments: one current electrode is put to the mouth of a borehole penetrating the oil reservoir, and the second current electrode is lowered into the same borehole below the reservoir, and on the surface electric and magnetic components due to the vertical electric dipole are measured. The original idea of this measuring arrangement comes from DC field experiments by Nagy (1970). Assuming simple circular models with higher resistivity than that of the host rock, a close connection was found according to the analogue modeling results between the areal distribution of different electromagnetic components. This technique can be directly applied in the field.

It was also found that a routine-like application of the CSAMT method must not be successful.

[1]Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H–9401 Sopron, POB 5, Hungary
[2]Geophysical Exploration Company, H–1068 Budapest, Gorkij fasor 42, Hungary

Description of the modeled problem

In Fig. 1 a typical cross section including a circular oil reservoir model is shown. The oil reservoir model was made of high resistivity plexiglass and the host rock was represented by NaCl-solution. Small copper rings were used as current electrodes. Three different borehole positions (one central and two off-centre ones) and three different buried electrode depths were studied.

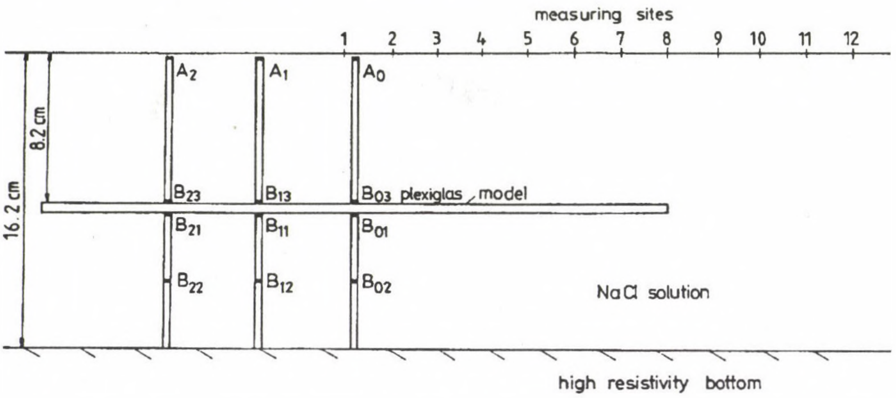


Fig. 1. Cross section of the analogue modeling problem. Nine different current electrode positions (three borehole positions and for each borehole three different B electrode depths) were applied. Distance between measuring sites is 2.5 cm

The question was how the effect of the reservoir model is reflected in different electromagnetic components in the near and intermediate zones (in the range $0.3 < k \cdot R < 10$, that is in the frequency range 0.125–4 MHz. k is the wavelength in the hostrock and R is the distance between the borehole and the measuring sites).

Anomalies over an oil reservoir model having a certain depth and size are influenced by numerous factors, like: depth of the buried electrode, mutual position of borehole and model, frequency of the electromagnetic field, well-casing, near-surface geological inhomogeneities, *etc.*

The present results refer to a single model radius/sediment thickness value (this ratio is 1) and the model depth/borehole depth ratio was in the range 0.6–1.0. Two different plexiglass models: a solid and a perforated one were used. The difference between them proved to be insignificant (smaller than 2–3 p.c.) for all components. Effects of well casing and near surface geology were neglected. Nevertheless, the model measurements give general information about exploration possibilities of such structures.

The technical details of the model laboratory are summarized by Märcz et al. (1986).

Theoretical and practical considerations

It is well known that the *surface electric field* due to a vertical electric dipole submerged in a layered half space has radial and vertical components. In presence

Table I. Radial electric and the tangential magnetic field normalized to those measured in absence of the oil reservoir model as a function of the borehole position and the electrode-depth (Numbers are according to Fig. 1)

a) Maximum values of the normalized electric field E_r along the measuring profile shown in Fig. 1

depth	Borehole position		
	2	1	0
3	0.4–1.2	0.4–0.6	0.5
1	40	17	7.2
2	17	8	4.8

b) Maximum value of the normalized magnetic field H_t along the measuring profile shown in Fig. 1

depth	Borehole position		
	2	1	0
3	1.1	1.04	1
1	1.7	1.20	1
2	3.5	1.6	1

of lateral inhomogeneities these two components are modified, and at the same time — on condition that the cylindric symmetry is distorted — a tangential field component appears, too.

The *magnetic field* around this transmitter is due to currents flowing in the earth and currents flowing in the cable section connecting the electrodes. In case of cylindrical symmetry neither the cable currents nor the earth currents have vertical and radial magnetic field components on the surface. In addition the tangential magnetic field component on the surface must be zero, too, since the earth- and the cable currents have just opposite and equal tangential magnetic fields (Veitch et al. 1990).

Consequently all magnetic components are in theory due to any deviation from the cylindrical symmetry. Circular oil reservoir models penetrated by central boreholes (like A_0B_0 in Fig. 1) cannot be detected by superficial magnetic components.

Geometric inaccuracies because of non-vertical boreholes may appear in all components, but they affect more significantly E_t, H_t, H_r and H_z than E_r and E_z.

In the following analogue modeling results will be disscussed. Two different techniques were used:

- at first, electromagnetic field components in presence and in absence of oil reservoir models were compared along a characteristic profile (in the case of E_r and E_z it was a normalization to the reservoir-free primary field; in the case of other components not having primary field, it was a simple normalization just to the analogue modeling noise);
- then the areal distribution of the electromagnetic field components were studied in details.

Effect of borehole positions and electrode depths

The effect of electrode depths and borehole positions is summarized in Table I for the E_r and H_t components. (Both components were normalized to their reservoir-free values.) If electrode B was below the model (electrode depths 1 and 2), both the magnetic and electric field values significantly increased compared to the corresponding reservoir-free values. The eccentricity also has a favourable signal-enhancing effect for both components. Therefore in the followings results are shown for off-centre boreholes and for the deepest B_2 electrode-position.

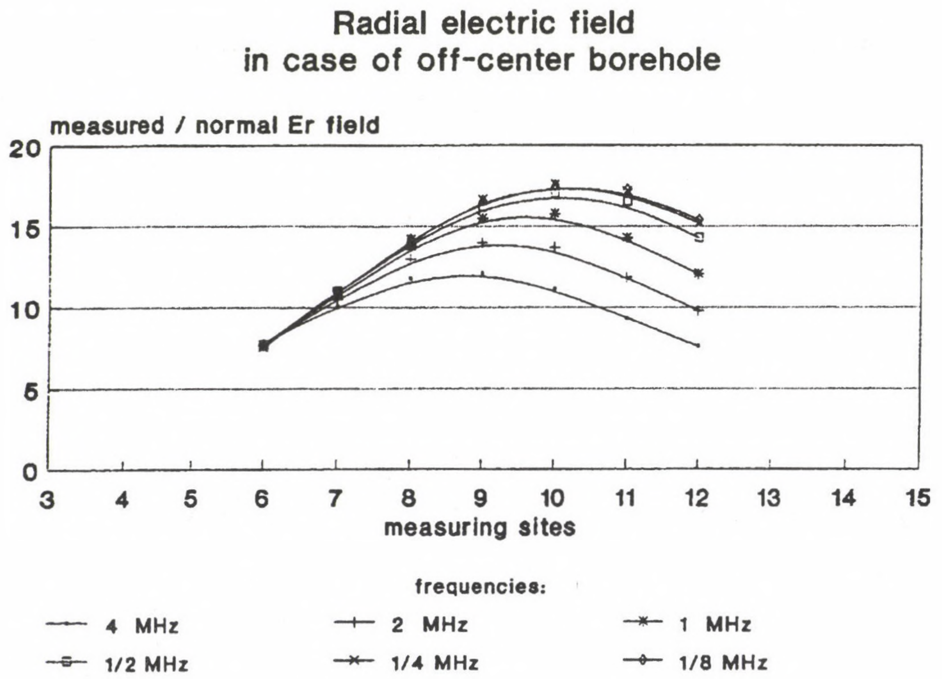


Fig. 2. Normalized (reservoir/reservoir-free) E_r profiles over the model boundary at different frequencies

In Fig. 2 normalized E_r profiles at six different frequencies are shown for current electrodes A_2–B_{21}. At lower and lower frequencies the maxima appear somewhat farther and farther away from the model boundary. (The corresponding H_t profiles have similar maxima over the DC maximum site of that of the electric component, but they do not have any frequency dependence.) From the two similar maxima in E_r and H_t it follows that the conventional impedance E_r/H_t is less sensitive to the model boundary than the electric and magnetic components separately.

Since the reservoir-free values in field conditions are not known, the normalization shown in Fig. 2 cannot be carried out in the practice. The problem is how to find a convenient normalizing function instead of the reservoir-free electromagnetic profiles.

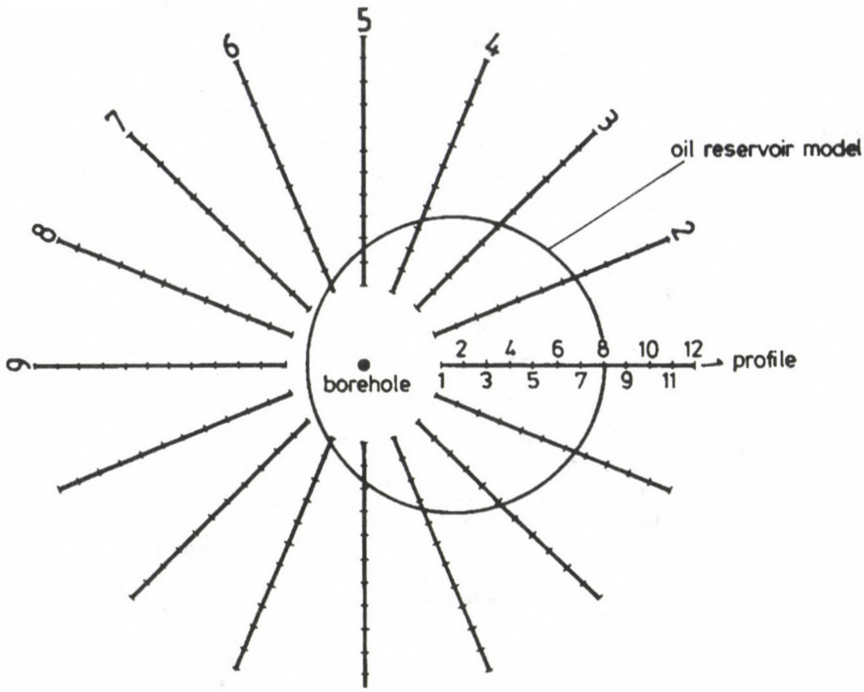


Fig. 3. Plane view of the modeled problem, showing 16 measuring profiles and the measuring sites over a circular oil reservoir model penetrated by an off-centre borehole

The proposed method

In Fig. 3, 16 measuring profiles having radial directions around the borehole were selected. Each profile had equidistant measuring sites in the vicinity of the expected boundary.

In Fig. 4a and in Fig. 5a, resp. measured E_r and H_z profiles are shown along the 1–8 profiles. The difference between them is not significant. In Figs 4b and 5b profiles 1–8 after normalization by the mean curve of the 16 measured ones are shown: in the normalized E_r profiles the maximum sites depend on the horizontal position of the boundary along the actual profile; in the corresponding H_z profiles the boundary is connected by a sharp minimum zone. In Figs 4c and 5c normalized E_r and H_z maps are shown. The effect of the boundary in E_r is not everywhere expressed: far from the borehole it is more emphasized than in its close vicinity. At the same time (see Fig. 5c) the oil reservoir model is unambiguously surrounded by a H_z minimum zone.

It must not be forgotten that this normalizing procedure does not work in the case of a central borehole. Fortunately in the practice the probability of a perfect cylindrical symmetry is near to zero.

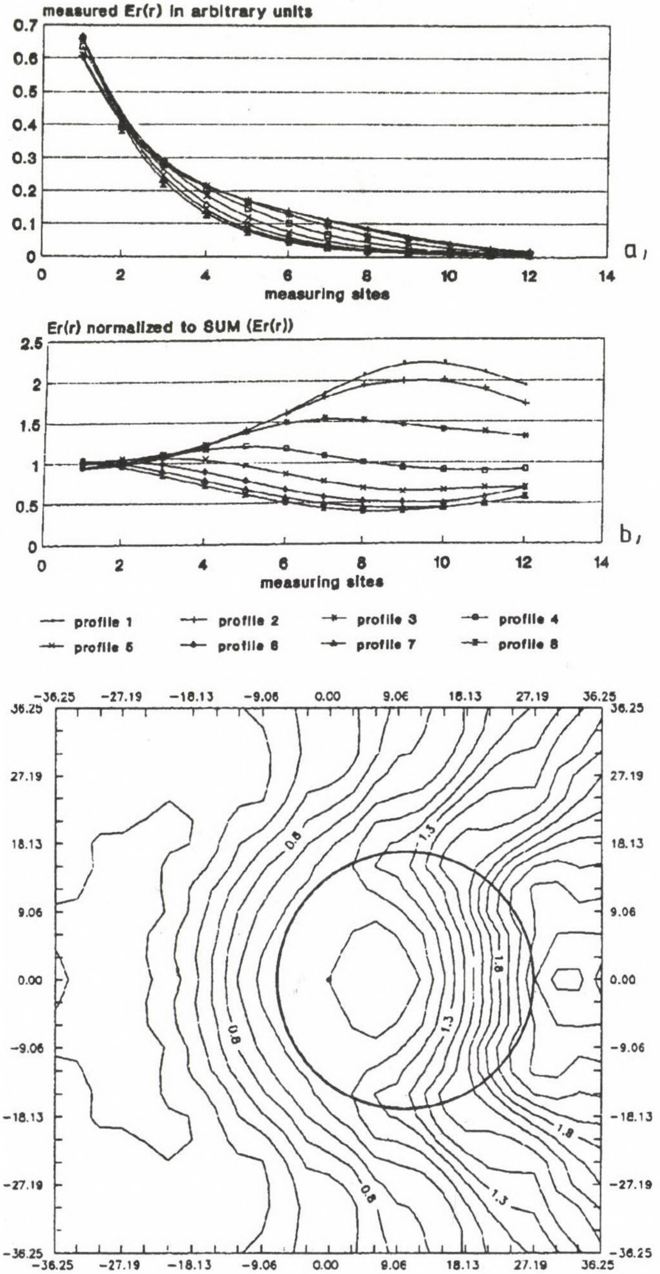


Fig. 4. E_r anomalies over a circular oil reservoir model at a characteristic intermediate-zone (f = 1 MHz). a) Measured E_r profiles along profiles 1–8 (in relative units); b) Normalized E_r profiles along profiles 1–8 (see the maxima closely connected to the model boundary); c) Normalized E_r map over the measuring area, showing local maximum zones far from the borehole

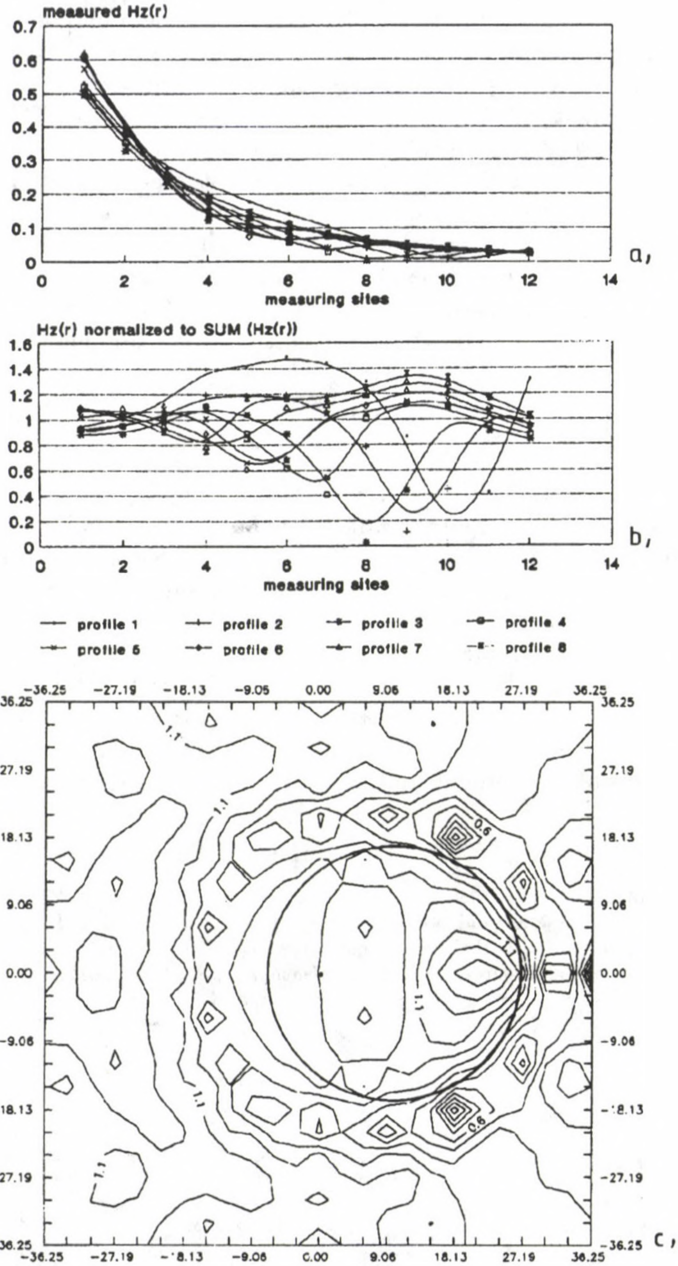


Fig. 5. H_z anomalies over a circular oil reservoir model at a characteristic intermediate zone frequency (f = 1 MHz). a) Measured H_z profiles along profiles 1–8 (in relative units); b) Normalized H_z profiles along profiles 1–8 (see the sharp minima closely connected to the model boundary); c) Normalized H_z map over the measuring area, showing a sharp minimum zone surrounding the model

In spite of the apparently better resolution of H_z, E_r proved to be much more reliable from technical points of view, because the H_z anomaly is much more sensitive to geometry inaccuracies which might even destroy the H_z anomaly image. Anyway, borehole-geometry based corrections are strongly proposed to be carried out for all components.

Behaviour of other components (both amplitude and phase) was also studied. The phase curves along different profiles had a systematic variation (up to 10 degrees), but an appropriate normalizing procedure for the phase was not found.

Finally it must be remarked that in the case of using the mean curve for the normalization, the frequency dependence of the normalized profiles practically disappeared (even for E_r) in contrast with that shown in Fig. 2. It means that in such a situation the quasi-stationary current deflection is the dominant effect.

Conclusions

The boundary of oil reservoir models having higher resistivity than the host rock can be detected by a simple surface electromagnetic technique. The proposed method needs a reservoir-penetrating off-centre borehole. (Boreholes in the centre of a circular structure are to be avoided.) Having one current electrode on the surface, and the second one in the borehole below the reservoir, the current deflection due to the reservoir results in typical electromagnetic field distortions on the surface.

From individual electromagnetic field components measured along different profiles radiating from the borehole, a mean curve for each component can be derived. Normalizing the individual measuring profiles to the mean curve, boundary-dependent anomalies can be determined. The E_r component has lower resolution, but it is less sensible to borehole inaccuracies than the other components.

References

Märcz Gy, Pongrácz J, Szarka L 1986: *Scientific Instrumentation*, 1, 119–133.

Nagy Z 1970: DC profiling around boreholes. Report of the Geophysical Exploration Company, Budapest (in Hungarian with English abstract)

Veitch D, Asten M W, van Leeuwen E H 1990: *Geophysics*, 55, 1605–1612.

Book review

R EMMERMANN and J WOHLENBERG eds: *The German Continental Deep Drilling Program (KTB). Site-selection studies in the Oberpfalz and Schwarzwald.* Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, 1989, 553 pages, 259 figs

This volume has been published as a valuable contribution to the series 'Exploration of the Deep Continental Crust' edited by H J Behr, Göttingen and C B Raleigh, Hawaii. It contains 21 papers by different scientific groups on the results of geological and geophysical pre-site investigations in the target areas Oberpfalz and Schwarzwald in Germany where the great challenge of the German geoscientists: the German Continental Deep Drilling Program (KTB) has been planned. The drilling started in late September 1987 near Windischeschenbach/Oberpfalz.

After the short Preface of the editors, one can read an interesting story on the development of a great and very costly project.: KTB — How It All Began by E Althaus. It is worthy to note the reasons mentioned by the Geocommission in its arguments at the very beginning of the project (1978) *"why a deep drilling should be performed: both scientific and economic goals like energy and mineral resources, drilling techniques, borehole measurements and acquisition of data on rocks of the deeper parts of the continental crust".*

The book is well balanced between the two perspective KTB drill sites in Oberpfalz and in Schwarzwald and the reader is consequently conducted from the basic geological, geo-chronological, geochemical, petrological *etc.* knowledge of these areas to the main geophysical results obtained by deep seismic, geoelectrical (electromagnetic), gravitational, magnetic and geothermal investigations. One can learn that the position of the KTB drilling site in Oberpfalz is representative of Variscan Belt in Europe and it probes into an Early Paleozoic rift basin. The other possible site of a deep drilling has been chosen in the metamorphic massif of the Central Schwarzwald which comprises Variscan and pre-Variscan rock sequences of magmatic and sedimentary origin. There are characteristic differences in the seismic structure (e.g. depth of the Moho discontinuity), in the heat flow values of the two areas being the latter greater in Schwarzwald than in Oberpfalz. It is interesting to note that highly conducting graphitic metasediments appear at different levels in the upper crust, but in both places at about 10 km depth.

As concerns the geophysical and borehole techniques, a detailed description of the Processing of Reflection-Seismic Data in the DEKORP Processing Center Clausthal is given together with the first result of the KTB Pilot Hole.

The book is rich in references.

Just to mention that a few papers did not have abstracts and in some cases the German text has not been changed by English in some figures (e.g. p. 227). Of course, these small remarques do not decrease the value of this excellent book.

A Ádám

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 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.* Fifty *offprints* are supplied free of charge.

CONTENTS

Index: 26.031

Acta Geodaetica, Geophysica et Montanistica Hungarica

VOLUME 27, NUMBERS 2–4, 1992

1828

Akadémiai Kiadó, Budapest

AGGM 27 (2–4) 141–424 (1992) 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

AKADÉMIAI KIADÓ
H-1519 Budapest, P.O. Box 245

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

CONTENTS

Book reviews

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 141–152 (1992)

AN INTERFEROMETRIC LENGTH COMPARATOR FOR CONTROLLING INVAR LEVELLING RODS

K KRAUSZ[1] and L ÖLBEI[1]

[Manuscript received June 11, 1991]

The authors have developed and built an interferometric length comparator. It can measure distances up to 10 m with a resolution 0.1 μm. The comparator has been equipped with photo-electric microscope and its function is controlled by a computer. The first goal was to measure the scale and graduation errors of invar levelling rods.

Several rods were measured. The system provides on rods a measuring accuracy better than ± 2 μm.

The development of the measuring system is to be continued.

Keywords: geodynamics; interferometric length comparator; laser interferometer; levelling rod

Many errors of precise levelling (e.g. refraction, adjustment of the instrument, base point error of rods *etc.*) can be eliminated totally or reduced considerably by the measuring technology. The scale of the rod and mainly the errors of graduation lines have not been taken into account, because it is too difficult to measure them. In the Geodetic and Geophysical Research Institute (GGRI) many methods have been worked out for measuring of levelling rods. A few field comparators with fixed halfmeter base were built which are suitable for the determination of the scale and its change in the field, caused by temperature, humidity or mechanical force (Halmai 1978). A mathematical procedure was published (Somogyi and Závoti 1981) which reduced the number of comparing observations. Despite the advantage of this measuring technology it has not become a widespread practice.

The most suitable equipment for measuring the errors of the graduation lines is a laser interferometer (Schlemmer 1975, Takolo 1974). The first experiment were made together with Geodetical and Mapping Company in 1983 (Tikász 1983). The laser provided an accuracy better than 1 μm, but the slow manual measuring process and the applied devices did not make it possible to bring the standard deviation below ± 3 μm. Disturbances of the laser made impossible to check the results. The experiment showed clearly that the comparison is to be faster, independent from individual error, the date flow is to be computerised thus the reliability of comparison is to be increased.

[1]Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H–9400 Sopron, Csatkai E. u. 6–8, Hungary

1. Measuring system

A He-Ne laser provides the physical basis of the length comparator. The interferometrical length measurement detects displacement. The travelled distance is obtained as the multiples of one-eighth wavelength. Several measuring set-ups can be realized (Heister 1988, Weise 1989, Gerwert 1989, Takolo 1985, Peterson 1982). When planning the comparator we strove to keep rigorously the Abbe prin-

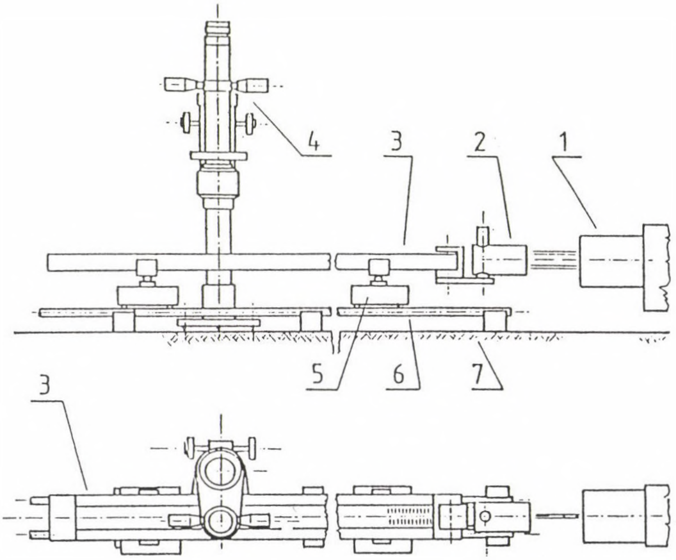


Fig. 1. The comparator. 1 — Laser and interferometer; 2 — Cube corner; 3 — Levelling rod; 4 — PE microscope; 5 — Carriage, 6 — Rail; 7 — Concrete base

ciple. Therefore the laser beam, corner cube and the tested side of the rod are in a straight line and the optical axis of the microscope is perpendicular to it (Fig. 1). The arrangement of the laser interferometric length comparator has not changed since the experiment in 1983. An automatic system was developed according to previous experience (Fig. 2). One side of the rod is measured continuously. The photoelectric microscope senses the centre of each graduation line and controls the automatic storage process of the measured lengths in the memory of the computer. Data can be used for diverse programmes to calculate the results.

1.1 Mechanical structure

The length comparator was built in a laboratory which had been established in the cellar of the GGRI. The base is a 22 m long concrete block. It lies on sand, which works as a vibration isolation. The daily fluctuation of temperature is within 4°C, in shorter periods it is constant to some tenth of degree. The carriages transporting

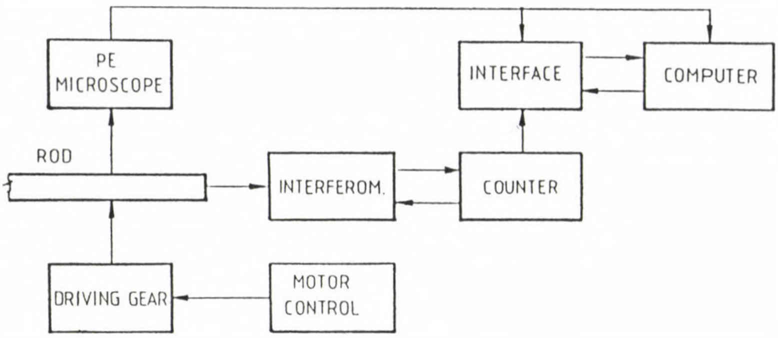


Fig. 2. Measuring and control system

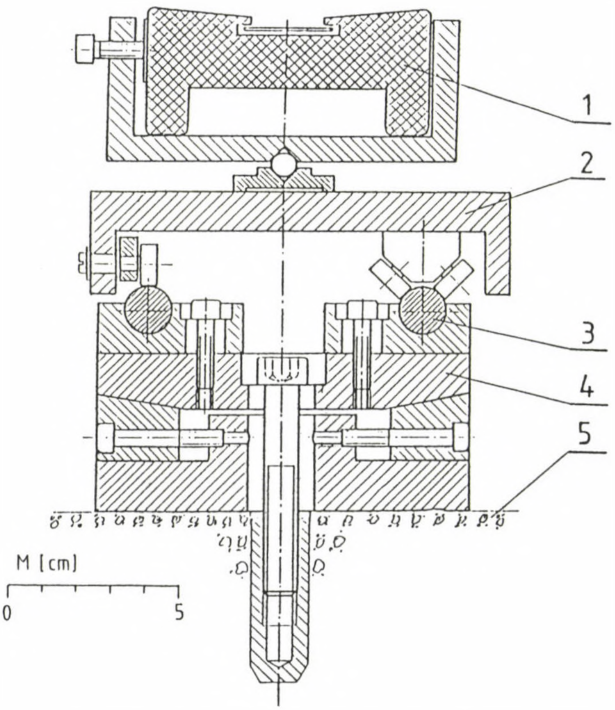


Fig. 3. Guidance. 1 — Levelling rod; 2 — Carriage, 3 — Rail; 4 — Support, 5 — Base

the rod move on the 10 m long guide rail. Its material is grinded and finished round steel. The supports are mounted in every half meter (Fig. 3) and provide the adjustment in two directions. The constant load and the measuring arrangement does not need a wider gauge. The rail material was out of true. After truing up, the cylindric shape was better than ±0.02 mm/500 mm.

The carriages roll on ball-bearings. Their positions are definite, their rigidity perpendicular to the measuring direction is suitable (0.02 mm/10N). The movement is made by a spliced, iron wire cable. This is sufficient for a continuous moving and

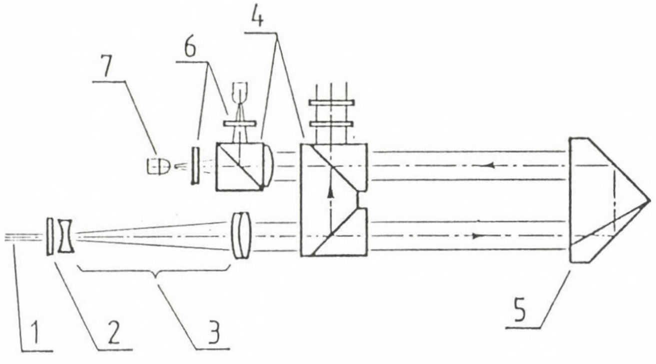


Fig. 4. Optical set up. 1 — Laserbeam; 2 — Retardation plate; 3 — Collimator; 4 — Beamsplitter; 5 — Cube corner; 6 — Polarizer; 7 — Sensor

fits to the length of the guide rail. A stepper motor drives the wire cable with a speed reducer. Various speeds can be switched on, the highest is 5.5 mm/sec. Failure of the scroll makes the driving speed fluctuate.

The levelling rod is transported by two carriages. The cube corner of the interferometer is mounted on the bottom of the rod. The invar strip is adjusted on the level attitude. The space distribution of the graduation lines is controlled by a stereo microscope which has been equipped with measuring marks. The rigidity and weight distribution is not constant along the rod, so the two supporting points are empirically found. Every observed graduation line was in a horizontal position within 1 mm. The deviation was even less, about 0.7 mm along 90 percent of the rod without the ends of the rod.

The guide rail was set and adjusted by a fixed telescope and level. Its difference from the optical beam was better than ±0.1 mm. The mutual position of the rails was checked by bubble, their divergence was better than ±0.02 mm.

1.2 Optical construction

The applied laser did not work regularly. It was substituted by a UNIPHASE He-He laser Typ 1410. Its frequency stability is ±25 MHz (UNIPHASE 1988 preliminary). The measuring uncertainty caused by laser is ±0.2 μm along the 3 m rod, adequate for the purpose. The new laser works well what compensates the difficulty of the rebuilding.

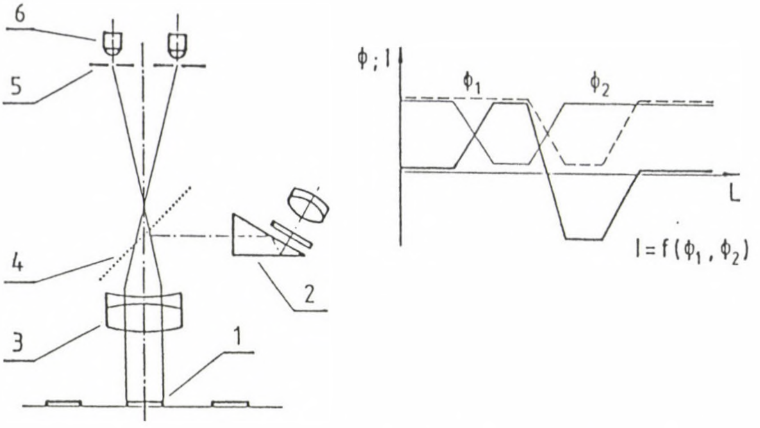


Fig. 5. Photo-electric microscope. 1 — Graduation line; 2 — Visual observing (it is turned by 90°), 3 — Objective; 4 — Beamsplitter (with 90/10 T/R ratio); 5 — Slit; 6 — Sensor

The disadvantage of the Michelson interferometer (Fig. 4) used is the dead path. The distance between the beam splitter and corner cube cannot be eliminated. It causes a zero drift depending on meteorology and on measuring period.

The electro-optical microscope (Fig. 5) is used for simultaneous observation. The objective projects the picture with magnification on two slits. The slits with the sensors are movable and can be turned around the centre line. The slits modify the sensitivity, so the electric signal has a shorter rising time.

Visual observation is only used for focussing and checking the spatial arrangement between the graduation line and optical centre.

The vertical optical axis is checked with a bubble. The bubble was adjusted by a horizontal reference plain. A quick checking method uses the rod itself. One graduation line is to be observed from different heights the deviation can be calculated in arcs of minute. This is allowed by the large depth of focus (4 mm).

The graduation lines are illuminated with a 100 W quartz halogen lamp using a 1.8 m long fiber optic light guide. Even the decrease of the irradiance by more than 50 percent does not influence the accuracy.

1.3 Electronic construction

The measuring system is seen on Fig. 2. The fringes produced by the optics of the interferometer are electronically sensed and amplified. The counter is able to count a quarter of the interferometric period (so its resolution is the eighth part of the wavelength) and determines the direction of the motion. The counter converts the result in metric system. The actual wavelength depending on meteorology is set manually. Nine digits are displayed which appear in BCD code on the back front of the equipment. If the mark is in an adequate position the photoelectric microscope gives a control signal which makes the interface read the BCD signals and send the stored values to the computer as well. It forbids a new control signal during the input.

1.31 The photoelectric microscope

The image of a graduation line is projected by the photoelectric microscope through two slits on phototransistors which are in a differential circuit (Fig. 6) (Flügge 1967). The output signal O_1–O_2 is independent of any distortion which appears with common sign in both signals (e.g. change of the main, alteration

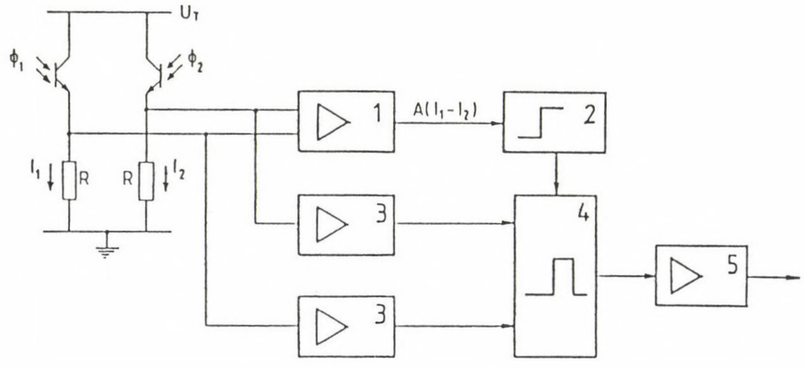


Fig. 6. Sensing unit. 1 — Difference amplifier; 2 — Comparator; 3 — Amplifier and filter, 4 — Impulse forming logic unit; 5 — Driver

of the reflexion). Further the difference output provides a faster transition in the centre of a graduation line, so it enables a more accurate comparison. The electronic comparator indicates the O_1=O_2 status by changing the output signal. Then the impulse forming logic unit gives a signal if it is allowed by the amplifier and filter. It is necessary to forbid by the high and low level of the O_1, O_2 signals, the O_1=O_2 status may appear several times between two graduation lines at a bright stretch what would result in a series of false signals. The driver unit forces the signal to be sent to the line transmission. This signal detected the centre of a mark controls the interface and the computer.

1.32 Interface

The interface reads the displayed value at the very moment as the centre of a mark passes by (the read-out time < 1 μsec), stores it and makes it suitable for transfer to the computer. The signal means a non-maskable interrupt for the computer. At this time the computer begins to read and to store the data in the memory. The computer forbids the interface input during the process, so a false signal cannot disturb the input. After the storage the computer again lets the interface, prepares it to the reading of the next measured value. The programme carrying out this process was written in assembler language to be able to compute fast.

2. Data flow

Data collection and computation is made with a Commodore 64, although the evaluation caused some difficulties in programming because of its small capacity.

The computation begins with the conversion of the data from BCD to decimal and correcting them according to the refraction. It is followed by a numerical filter, because the microscope may send false signals or may miss it on account of abraded rod marks or false adjustment of the equipment. The filter is based on the fact that a false impulse cannot appear within ± 0.5 mm around the centre of a graduation line. The computer stores the data on floppy disc, together with identity and meteorology data.

A lot of programmes help to determine the accuracy and reliability of the interferometric comparator.

3. The accuracy of the system

The invar levelling rod can be correctly modelled by a complicated mechanical model. The change of its form in the space depends on acting forces, temperature, humidity and time. Therefore the rod can extend, slew and bend in. Thus between the periodical comparisons, a larger error may appear than the accuracy of the comparator.

The development of the interferometric comparator has not been finished yet. So this report characterizes an intermediate state (Table I).

The accuracy of the interferometer is limited by the uncertainty of laser frequency, the digital display and the dead path. The last one can be reduced at the measurement because the rods have the same length and so the stretch between the interferometer and cube corner can be set small.

The accuracy of single line detection or the repeatibility by the photoelectric microscope (at least 50 repetition from one direction) is far better than the visual observing. The tilt of the vertical axis or its inaccurate adjustment is the main error source of the comparator. Its adjustment needs an exact horizontal reflecting plain in which to adjust. The decrease of the bend of the invar tape may also improve the accuracy. The electronics of the microscope is made fast enough to avoid being sensitive for the fluctuation of the rod velocity caused by mechanical deficiencies.

The straightness of the rails was checked with an autocollimator, the mirror was mounted on the rod itself. The measured values were within 30 sec of arc, so the error caused was less than 0.1 μm, if the Abbe principle is kept precisely.

The errors affected by the rod, like abrasion, humidity of the wooden construction need additional investigation.

Table I. The errors of the laser comparator

Instability of the laser (±25 MHz) (UNIPHASE 1988: Preliminary)	±0.053 μm/m
Meteorology (Kohlrausch 1968)	
Temperature (±0.2°C)	±0.214 μm/m
Pressure (±0.2 hPa)	±0.043 μm/m
Humidity (±3 %)	±0.027 μm/m
Uncertainity of the start point (by 0.8 m long dead path)	±0.18 μm
Display	±0.1 μm
PE microscope (the extreme is < 0.3)	±0.1 μm
Vertical position of the PE microscope (by 0.5 mm bend-in of the rod, 5 arc of min tilt)	±0.7 μm
Response to the velocity (the altering of the velocity is 1 mm/sec)	±0.16 μm
Errors caused by the rail	
Cosinus error	0
Sinus error (deviation is 5 mm)	±0.7 μm
Turning of the cube corner	0

4. Results

We began the first experiments on a rod type Rost as used in GGRI. The errors of each graduation line were counted. The errors of the rod sides showed an error from a former production technology namely a 300 mm long mold were used (Fig. 7). The interval of the marks was 5 mm and the graduation line numbered 597 was considered as 0. The largest error was 114 μm.

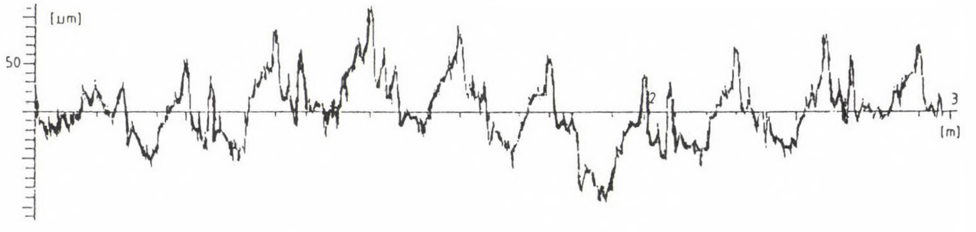


Fig. 7. Measuring outputs

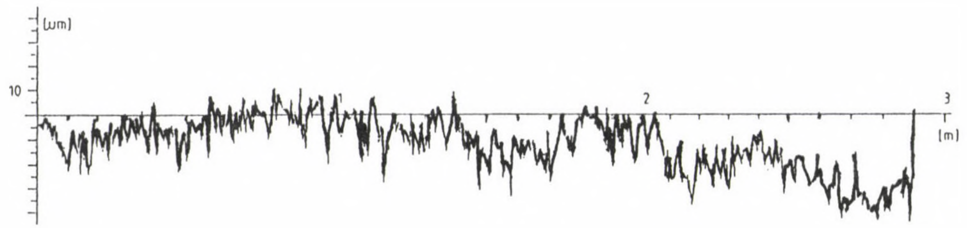


Fig. 8. Measuring outputs

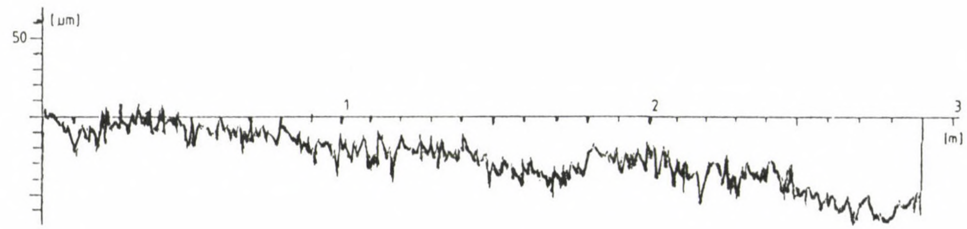


Fig. 9. Measuring outputs

The first practical application was done with a pair of rods type Zeiss Jena. The goal was to compare the rods before and after a four weeks field levelling campaign and to get some information about reliability of the laser comparator. The left and right sides of the rods were measured at least five times and averaged the errors of each graduation line. The dimension of both the rods were nearly the same. The largest scale deviation was -45 μm (Fig. 8) on the left side of the rod No. 53815 and -69 μm on the right side (Fig. 9).

After the field levelling the rods were measured again, their new scales were calculated. The change of the error belonging to every single mark in four weeks showed an increase of the rod dimension. Its value on the left side was 38 μm on 3 m length (Fig. 10) and 34 μm on the right side. These values were 34 μm and 25 μm on the rod No. 53816. The differences of the errors of repeated measurements were between extremes ± 3 μm disregarding the dimension change.

What could have made the rods extend? We made an effort to provide identical conditions by comparing. The system could not get deadjusted, because it could not have worked well. The position of the rods were the same as previously. We were not able to measure the temperature of the invarband and the humidity of the wooden case.

The measurement was carried out between 20.9°C and 22°C air temperatures. We supposed that the invar band followed the temperature with a large delay. This was verified by the compartment of left side of rod No. 53816 had been made in 2

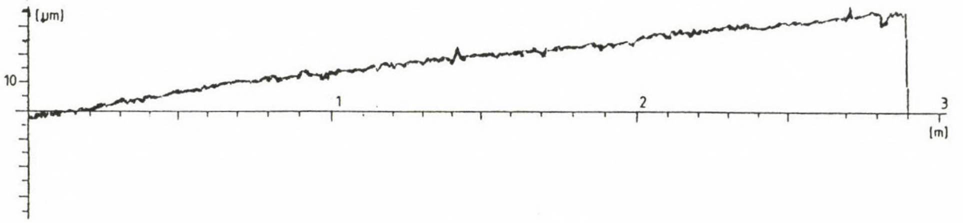


Fig. 10. Measuring outputs

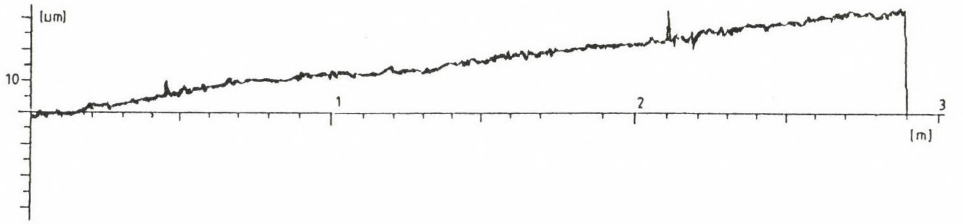


Fig. 11. Measuring outputs

days. The temperature of the laboratory was stable up to 0.2°C in the meantime. Even the repeated switch on did not cause any change in the measured results. The error of each graduation line was the same within 1 μm, the standard deviation was ± 0.4 μm. (Every measurement of one side is an independent process, the change of temperature is only disturbing during the process). The difference of errors which were measured from two directions continuously was zero in practice, because the values up to ± 0.9 μm may be considered as noise (Fig. 13). The measurements made in altering temperature conditions brought worse results.

An other error source of the unknown expansion may be the wooden structure

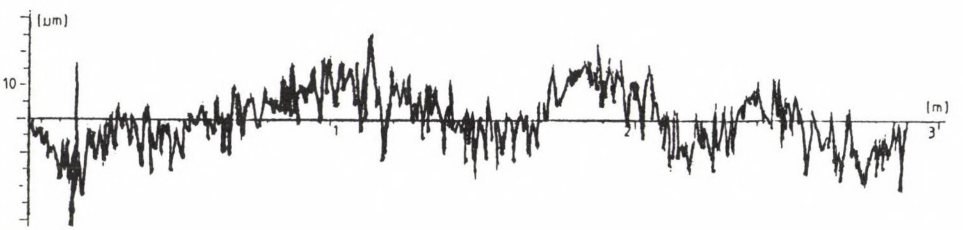


Fig. 12. Measuring outputs

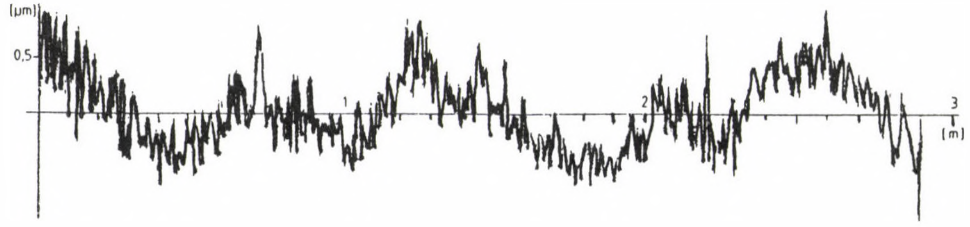


Fig. 13. Measuring outputs

of a rod. Wood has a hygroscopic property so the humidity influences its dimension thus the scale of a rod can change. This symptom needs a long term investigation.

We made a test with the left side of rod No. 53816. When it was measured first, its bottom faced the laser, then the top. The differences of errors were not zero (Fig. 14), what has several causes. The graduation lines of a rod are not in a

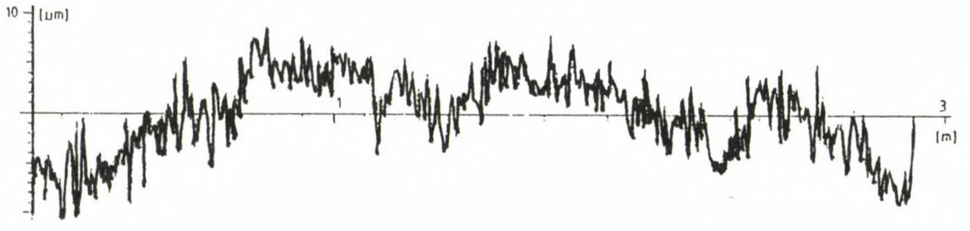


Fig. 14. Measuring outputs

straight line therefore they are measured at another section turning over the rod. The tilted axis of the microscope doubles the error casued by the tilt. The error increases if the horizontal position of the rod changes, the graduation lines abrade and the illumination is not symmetrical.

5. Conclusion

The measuring system has achieved an accuracy level which will serve the requirement of levelling. The repetition of the measuring condition is considered important. Therefore the carriages are to be modified to guarantee more rigidity for the rod. The temperature of the invar band will be measured with an infrasensor thus the coefficient of the termal expansion will be. Two directional illuminations are planned. The data flow will be faster and more comfortable because the programmes will run on a PC.

Special thanks are due to Dr Aladár Orbán and Dr József Závoti for their contribution.

References

Flügge S ed. 1967: Optik, Handbuch der Physik (Band XXIX), 801. Springer Verlag

Gerwert O E 1989: *Das Markscheidewesen*, 96, 314–319.

Halmai E 1978: Length comparator with ceramic base. Paper on Conference Optika'78

Heister H 1988: Zur automatischen Kalibrierung geodätischer Längenmessinstrumente. Schriftenreihe der Ud.BW München, Heft 27

Kohlrausch F ed. 1968: Praktische Physik, 1. Band, Teubner Verlag, Stuttgart

Peterson I 1982: Om invarstanger och deras komparation. Nordiska Kamission för Geodesi. Det Nionde Nordiska Geodetmötet. Gävle, Band 1, 164–197.

Schlemmer H 1975: Laser-Interferenzkomparator zur Prüfung von Präzisionsnivellierlatten. DGK Reihe C, München, Heft 210.

Somogyi J, Závoti J 1981: *AVN*, 88, 213–219.

Takolo M 1974: The Laser Rod Comparator. Reports of the Finnish Geodetic Institute, 5, Helsinki

Takolo M 1985: Horizontal-Vertical Laser Rod Comparator. Reports of the Finnish Geodetic Institute, 2, Helsinki

Tikász E 1983: Technology and software development on unified high data system of the country. Report of BGTV

UNIPHASE 1988: 1400 series frequency stabilized heliumneon Lasers systems. Preliminary

Weise H 1989: *Vermessungstechnik*, 37, 7–11.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 153-166 (1992)

THE GEOID IN EUROPE: REQUIREMENTS AND MODELING STRATEGIES[1]

W TORGE[2]

[Manuscript received September 16, 1992]

Regional geoid resp. quasigeoid determinations are nowadays required with an accuracy of ±1 to 10 cm over distances from 100 to some 1000 km in order to meet the demans of geodesy, geophysics, oceanography and engineering. Especially the combination of GPS heighting with classical leveling is one of the primary drivers for precise geoid computations. As a consequence, the IAG International Geoid Commission recognized at its meeting in Milano, 1990, that there is an urgent need for a new European geoid computation. This solution should be significantly improved in spatial resolution and accuracy as compared to presently available models. This led to the decision to form a Subcommission for the Geoid in Europe, and the Institut für Erdmessung (IfE), University of Hannover, was asked to serve as a computing center in this project.

In the first part of this paper early geoid/quasigeoid computations for the area of Europe as well as more recent results obtained at IfE are summarized. The latter solutions include a gravimetric and astrogravimetric quasigeoid, which have a spatial resolution of about 20 km and a relative accuracy of some dm. Then the possibilities for an improved European quasigeoid calculation are outlined, considering the availability of new and better global and regional data sets. An overview is given on the procedures currently under study at IfE and on the work perfomed at IfE since 1990. This work includes the collection and screening of new point gravity and terrain data, some investigations on the use of topographic information available at present, and the calculation of a preliminary quasigeoid solution for central, northern and western Europe including a GPS/leveling control. The paper closes with a survey on future activities at IfE within the European geoid project.

Keywords: geoid; gravity field modeling; height reference surface; quasigeoid

1. Requirements at regional geoid determinations

The geoid, being an equipotential surface of the earth's gravity field, is a natural reference for describing the heights of the topography on land (continental topography) as well as on sea (sea surface topography). In practice, national height reference surfaces are defined by benchmark heights. They approximate the geoid within a few decimeters in the absolute sense, and with additional few centimeters deviation over 100 to 1000 km. These approximation errors stem from bias effects at the tide gauges used for fixing the zero height, from leveling and gravity reduction errors, and from density hypotheses. In order to avoid the errors of hypotheses, the

[1] Inaugural Lecture, Hungarian Academy of Sciences, Budapest, December 9, 1991

[2] Institut für Erdmessung, University Hannover, Nienburger Str. 6, D-3000 Hannover 1, FRG

quasigeoid has been introduced in many countries as a height reference surface. In the sequel, when discussing geoid determinations in Europe, we also prefer to deal with the quasigeoid. This has the advantage, that only gravity field data observed at the earth's surface and in the exterior of the earth enter into the calculations, and no assumptions about the gravity field in the earth's interior have to be made. A subsequent transformation to the geoid introducing a density model can easily be performed.

In a regional sense with dimensions of some 100 to some 1000 km, the accuracy demands on the quasigeoid may be described as follows (*e.g.* Torge 1987):

— $\pm 0.1 \ldots 1$ m for the reduction of distance measurements;

— $\pm 0.01 \ldots 0.05$ m for the mutual transformation and combination of heights derived from GPS (Global Positioning System) and geometric leveling;

— ± 0.0 describing the sea surface topography;

— ± 0.001 m for application in vertical crustal movement research.

While the first demand is practically fulfilled in Europe (see section 2), the "cm" accuracy level has not been reached. However, due to the rapidly increasing use of GPS heighting, which already now gives the same accuracy as leveling over some 10 to 100 km, there is an urgent need to provide the "cm" quasigeoid to geodesists and surveyors.

In order to estimate the spatial resolution and the accuracy of the gravity field data to be employed for the solutions, we may use available information about the statistical behaviour of the gravity field. If we take the global spectral decomposition, as provided *e.g.* by the anomaly degree variance model derived by Tscherning and Rapp (1974), we recognize that the omission errors resulting from neglected high frequency field parts at a resolution (corresponding to an average station distance) of a few km are less than 1 cm for height anomalies, a few 10 $\mu\mathrm{ms}^{-2}$ for gravity anomalies, and 1" for vertical deflections respectively. This also gives an idea about the necessary observation accuracy of the gravity field data which can be used. As the omission errors and the observation errors add in the calculations, the average distance of the data points should not exceed the limit given above.

In the following, we first mention some of the more important geoid (quasigeoid) calculations performed for Europe since the 1950's. We then describe the possibilities given by more recent global and regional data sets of high quality, as well as by more sophisticated modeling procedures and by improved computing facilities. The paper focuses on procedures for an improved quasigeoid calculation for Europe pursued at Institut für Erdmessung (IfE), being the computing center in this project as requested by the IAG Subcommission for the Geoid in Europe. First numerical results are presented for central and northern Europe using point gravity data, a digital terrain model and a global model. An independent control of this solution with different GPS/leveling data sets reveals the efficiency and quality of such new solutions.

2. Existing geoid determinations for Europe

At the end of the 1940's attempts for a geoid calculation in Europe were made. The first one uses astrogeodetic vertical deflections and the astronomic leveling technique in order to reduce the Central European Triangulation Net to the Hayford ellipsoid (Wolf 1949). The gravimetric solution of Tanni (1949) is based on $1^\circ \times 1^\circ$ and $5^\circ \times 5^\circ$ isostatic anomalies. Both determinations give relative geoid height with accuracies of a few meters only. Work then concentrated on the astrogeodetic solution with strong efforts undertaken within the framework of IAG. These attempts led since 1954 to the continuously improved Bomford geoid (Bomford 1972). The last version of this astrogeodetic geoid includes some 1000 vertical deflections, resulting in a relative accuracy at the one meter level in well surveyed areas like Finland, France, Federal Republic of Germany and others (Levallois and Monge 1978).

A major step forward then was the gravimetric quasigeoid calculation EGG1 perfomed at IfE, Hannover, in the beginning of the 1980's (Torge et al. 1982). This solution is characterized by the following features:

— establishment of a high resolution gravity data base ($1^\circ \times 1^\circ$ and $6' \times 10'$ mean gravity anomalies) with careful analysis and screening of the data;

— statistical analysis of the data and derivation of signal and error covariance functions, needed for anomaly prediction and for error propagation in the transformation process;

— least squares spectral combination with closed integral formulas, using a global satellite derived model (GEM 9), 12000 $1^\circ \times 1^\circ$ anomalies (integration radius 20°), and 104000 $6' \times 10'$ anomalies (integration radius 3°), yielding quasigeoid values in a $12' \times 20'$ grid;

— straightforward error calculation with *a priori* error variances and covariances (for the anomalies), resulting in an absolute error estimate of ± 0.9 m, and relative errors of $\pm 0.3 \dots 1.1$ m/$100 \dots 1000$ km, for the quasigeoid derived, see Fig. 1.

By including about 5000 vertical deflections, this solution was extended to the EAGG1 quasi geoid (Brennecke et al. 1983), where the astrogeodetic data improved the gravimetric solution mainly in areas with lacking or insufficient gravity data.

An independent accuracy control of the quasigeoid is provided by the central and northern part of the European GPS traverse, established by IfE in cooperation with many other institutions in 1986 and 1987 (Torge et al. 1989). This traverse follows first order leveling lines of the United European Leveling Network (UELN), thus providing about 70 geometrically (GPS/leveling) derived quasigeoid heights, covering a range of 3000 km with an average station distance of 50 km (see Fig. 2). From loop misclosures of the GPS traverse and UELN, and from the results of comparisons with more recent quasigeoid calculations (see section 4), we estimate the relative accuracy of the quasigeoid heights thus derived to be a few centimers to one or two decimeters for distances between 50 and 3000 km.

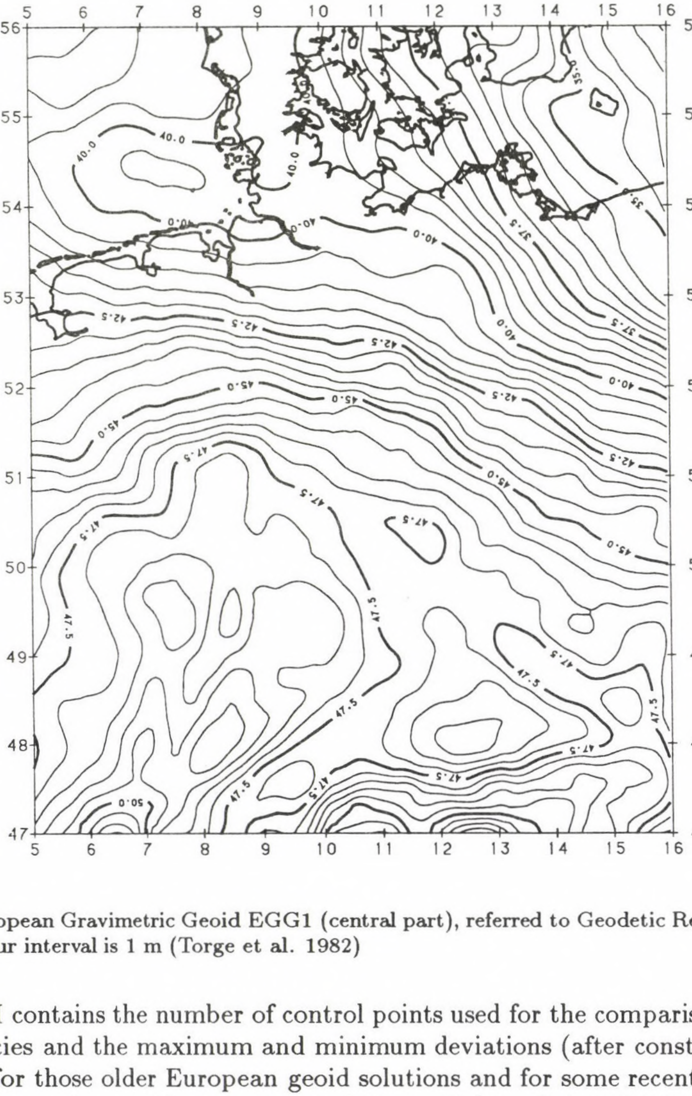


Fig. 1. European Gravimetric Geoid EGG1 (central part), referred to Geodetic Reference System 1980, contour interval is 1 m (Torge et al. 1982)

Table I contains the number of control points used for the comparison, the RMS discrepancies and the maximum and minimum deviations (after constant bias subtraction) for those older European geoid solutions and for some recent global models. The comparisons with the astrogeodetic geoid (Levallois and Monge 1978), the gravimetric quasigeoid EGG1 (Torge et al. 1982), the astro-gravimetric quasigeoid EAGG1 (Brennecke et al. 1983), the tailored spherical harmonic model IFE88E2 (Bašić et al. 1989) and the global model OSU89B (Rapp and Pavlis 1990) are given in Fig. 2. The large scatter of the pure astrogeodetic geoid reveals the deficiencies of this solution, with large station distances (especially in Norway) and not homogeneously distributed data, as well as systematic errors of the astronomic positions. The gravimetric geoid EGG1 is distorted by a strong slope of 1.5 m in Denmark, Sweden and Norway, which is due to the bad quality of the data used in this region (Bašić et al. 1989). This slope has completely disappeared in the more

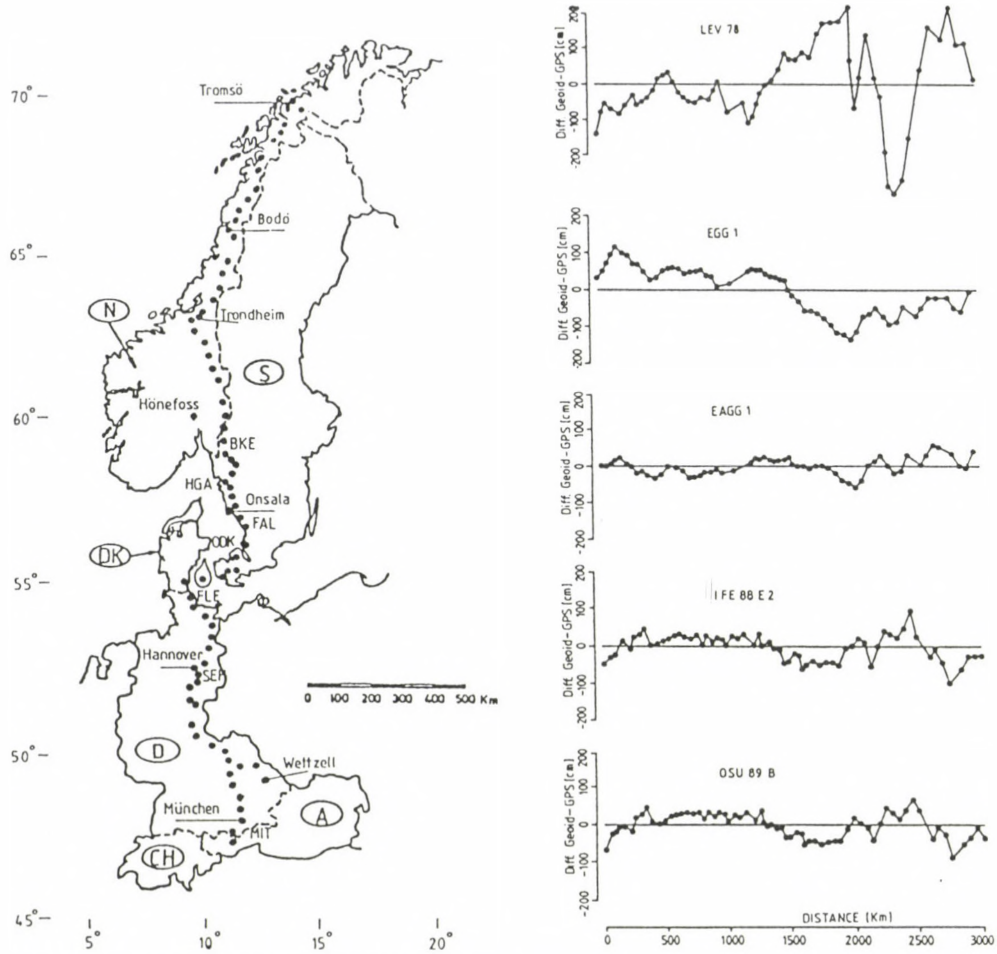


Fig. 2. European north-south GPS traverse (left part) and comparison of results from gravity field modeling for Europe with GPS/leveling quasigeoid heights along this traverse (right part)

recent solutions for Scandinavia (see Forsberg 1990 and section 4) as well as in the astrogeodetic solution EAGG1, demonstrating how even scarce vertical deflection data may improve a regional calculation.

3. Data sources and modeling strategies for improved geoid determinations

In the 1980's, new global, regional and local gravity field data sets became available, as well as digital terrain models (DTM's). In connection with progress in gravity field modeling procedures and computing facilities, the possibility of proceeding from regional dm ... m accuracies to cm ... dm accuracies became visible.

Table I. Statistics for the comparison of different geoid/quasiogeoid solutions with the European GPS traverse results

Solution	Authors	No. of Stations	RMS [m]	Min. [m]	Max [m]
Astrogeodetic	Levallois/Monge 1978	60	±1.17	−2.73	+2.38
Gravimetric (EGG1)	Torge et al. 1982	60	±0.63	−1.33	+1.13
Astrogravimetric (EAGG1)	Brennecke et al. 1983	60	±0.24	−0.59	+0.60
Global Model (GPM2)	Wenzel 1985	60	±0.70	−1.45	+1.26
Global Model (OSU86F)	Rapp and Cruz 1986	60	±0.77	−2.04	+1.09
Tailored Model (IFE88E2)	Bašić et al. 1989	60	±0.32	−0.89	+0.94
Global Model (OSU89B)	Rapp and Pavlis 1990	60	±0.31	−0.81	+0.69

The progress in data collection may be summarized as follows:

— availability of improved satellite-only models including the associated covariance matrices, *e.g.* GEM-T1 (Marsh et al. 1987) and GEM-T2 (Marsh et al. 1989);

— availability of new satellite altimeter data from the Geosat Exact Repeat Mission including precise orbits;

— availability of global high resolution geopotential models, as GPM-2 (Wenzel 1985), with a spherical harmonic development to degree and order 200, corresponding to a spatial resolution of 1°, and OSU86E/F (Rapp and Cruz 1986) or OSU89A/B (Rapp and Pavlis 1990) complete to degree and order 360, corresponding to a resolution of 30'; the overall accuracy of these models is estimated to be ±0.3 to 0.7 m (see also the comparison with the GPS traverse in Table I) including the omission error of ±0.4 m and ±0.2 m respectively;

— calculation of regional geopotential models, by tailoring global models to the gravity field in Europe. We mention the 360 models IFE88E1 (Bašić 1989) and IFE88E2 (Bašić et al. 1989). Both use OSU86F, and tailor the harmonic coefficients in the medium spectral range either to a $12' \times 20'$ or to an updated $0.5° \times 0.5°$ mean gravity anomaly data set for Europe; the estimated accuracy is ±0.3 m, as verified by comparison with the GPS traverse (see Table I);

— availability of point gravity anomalies, at least for larger parts of Europe and its surroundings, with $\pm 2 \ldots 20\ \mu\text{ms}^{-2}$ accuracy and station distances of 2 to 5 km (see Fig. 3);

— availability of high resolution digital terrain models, as the global $5' \times 5'$ model, and regional models with a block size of 1 km × 1 km or smaller;

— increasing availability of point geoid/quasigeoid heights from Doppler/leveling (±0.2 ... 0.5 m/100 ... 1000 km) and GPS/leveling (±0.01 ... 0.1 m/1 ... 1000 km) control points.

This led to the development of a strategy for improving the existing European geoid solutions by applying the "remove-restore technique" (Denker et al. 1986). Here the gravity field information is splitted up into three different parts:

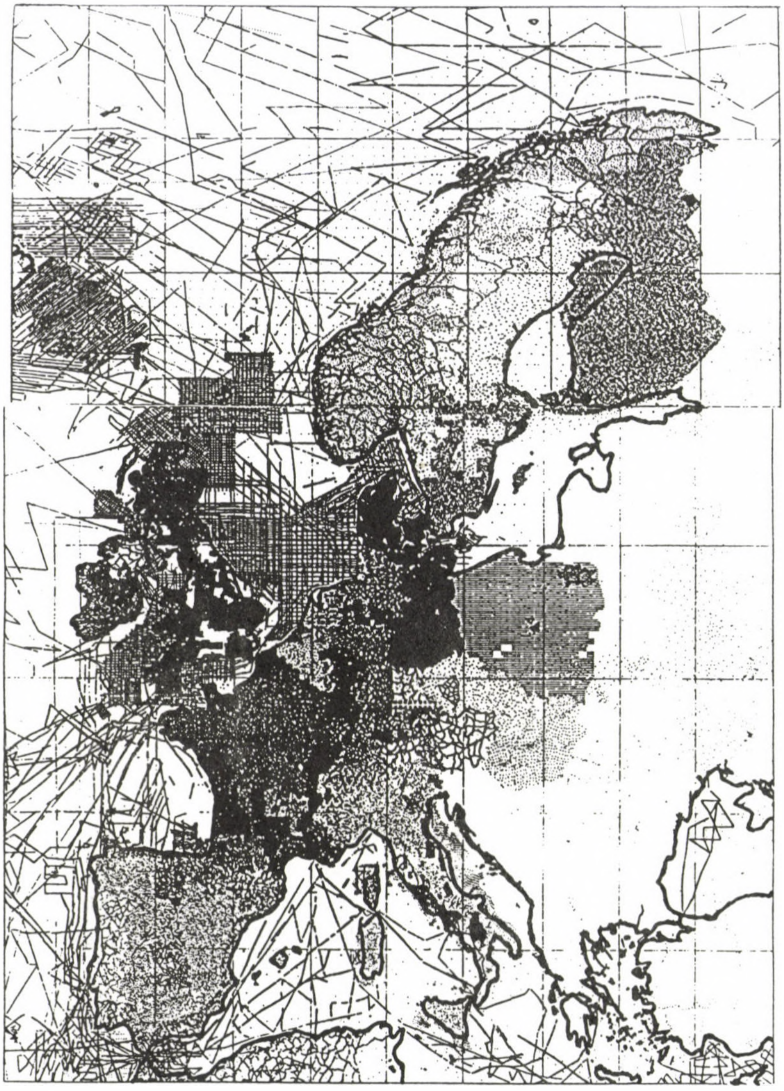


Fig. 3. Locations of point gravity data stored in the IfE data base (status of July 1991)

— the long wave part up to about 200 km wavelength is taken from a global or regionally tailored geopotential (spherical harmonic) model;

— the medium wave part (200 km to 5 ... 20 km wavelength) is taken from terrestrial point gravity field data, as gravity anomalies and astrogeodetic vertical deflections;

— the short wave part is derived from a high resolution digital terrain model.

Removing the long wave and the short wave part from the point data (which corresponds to a spectral filtering) leads to a residual gravity field data set, on which gravity field transformation algorithms for deriving the quasigeoid are applied. As shown, among others, by Denker (1988) and Bašić (1989), integral formulas and least squares collocation give comparable results at this transformation. By restitution of the long and short wavelength gravity field part to this residual quasigeoid solution, the final quasigeoid is derived. Simulation studies revealed that by applying this technique, the data collection area in integral formulas and collocation may be strongly reduced, leading to more economic calculations.

Theoretical and numerical investigations proved the great potential of updated quasigeoid calculations based on the concepts described above, with a few cm accuracy over distances of some 10 to some 100 km. As two remarkable examples, we mention here the new quasigeoid for the western states of the Federal Republic of Germany (Denker 1989) and the new Scandinavian quasigeoid (Forsberg 1990). Both calculations were done using Stokes formula in connection with the remove-restore technique, where the numerical evaluation was perfomed by means of FFT. For the 1989 solution for the Federal Republic of Germany, the tailored spherical harmonic model IFE88E1, 46000 point gravity anomalies (2 ...5 km, $\pm 3\ \mu\mathrm{ms}^{-2}$), $6' \times 10'$ mean anomalies in the adjacent area (distance $2^\circ \ldots 3^\circ, \pm 10 \ldots 50\ \mu\mathrm{ms}^{-2}$), and a $30'' \times 50''$ DTM ($\pm 5 \ldots 50$ m) were used. While the long wave model part of the quasigeoid (calculated in a $60'' \times 100''$ grid) provides values between 39.2 and 50.5 m, the medium wave part adds only a signal of ± 0.18 m (maximum 1.2 m), while the terrain model part is ± 0.02 m (maximum 0.17 m). GPS/leveling quasigeoid heights again provide a reliable control for the gravimetric quasigeoid solutions. This was possible in local areas (Hannover test area), where a ± 0.02 m RMS discrepancy was found as well as for the entire computation area, where the comparison with the European GPS traverse gave a RMS discrepancy of ± 0.05 m, while for the DÖNAV campaign (preliminary results obtained at the computing center IfE, Hannover; Prof. Seeber, personal communication) a RMS discrepancy of ± 0.06 m was obtained (Denker 1990). Forsberg (1990) applied very similar procedures for Scandinavia and obtained a RMS difference of ± 0.10 m versus GPS/leveling results from the European GPS traverse over a range of about 2000 km.

4. The IAG/IfE European geoid project

From the discussion in the previous section it is clear, that new gravity field data sets of high quality are available today and will be extented in the near future. This includes new global models, new satellite altimeter data from the ERS-1 and TOPEX missions, further gravity data not yet released, more GPS/leveling quasigeoid heights, as well as regional high resolution digital terrain models. Taking these possiblities into account, and with the experiences already obtained, IfE, Hannover, has proposed at the First International Geoid Commission Symposium in Milano, 1990, to attack the determination of an improved quasigeoid for Europe and has offered to do the calculations (Torge and Denker 1990). The Geoid Commission, recognizing the urgent need for such an improved solution, decided to establish a Subcommission for the Geoid in Europe in order to support the geoid project and elected Dr. Vermeer as its chairman. IfE was then officially requested by the chairman of the subcommission to serve as a computing center in this project. The procedure for the computation of an improved quasigeoid for Europe to be followed at IfE, will be — at the present state of discussion — the following one:

— development of improved global geopotential models, with inclusion of improved "satellite-only" models, and straightforward error calculation, with eventual inclusion of global topographic-isostatic models;

— extension of the existing IfE gravity data base by including point gravity field data (or high resolution mean anomalies), regional DTM's, satellite altimetry and sea surface topography models, as well as GPS/leveling quasigeoid heights;

— careful screening of these data with respect to gross and systematic errors, transformation the same reference (height system, tidal reductions), and error assessment of the data;

— improvement of the existing software for quasigeoid calculations, including error propagation;

— development and testing of new gravity field solutions for Europe.

The strategies for the calculation may be characterized by the following items:

— quasigeoid determination using improved global models, observed high resolution (eventually gridded) gravity field data and digital terrain models;

— quasigeoid determination as above with inclusion of satellite altimeter data and sea surface topography models;

— quasigeoid determination as above with inclusion of GPS/leveling point data;

— quasigeoid determination using locally derived gravity field data, especially national quasigeoid solutions, as a pragmatic solution, to be followed if high resolution (point) gravity anomalies are not provided for large areas. This strategy requires more investigations about the combination of national solutions with special emphasis on the transformation to a common reference, and the control of long wavelength error propagation.

Since the initiation of the European geoid project in 1990, the subcommission and IfE have approached a number of national agencies and individuals to release high resolution terrain data (block size 250 ... 1000 m) as well as additional gravity data not stored at Bureau Gravimétrique International (BGI). This led to the release of a significant amount of new data. The status of the IfE gravity field data base may be characterized by the following figures. In the mid of 1991 the data base contained about 1.5 million point gravity data, with the majority of data coming from BGI, and about 15 million terrain data. About 800000 gravity values for central, northern and western Europe have been validated so far using batch and interactive procedures developed at IfE. Figure 3 shows the distribution of the available gravity data for central and northern Europe.

At the present state of evaluation, IfE is concentrating on a pure gravimetric solution, using residual gravity anomalies reduced for the effect of a global spherical harmonic model as well as for the effect of the topography using the residual terrain reduction (RTM) technique (see Forsberg and Tscherning 1981). Due to the large amounts of data which have to be employed, FFT is used for the field transformation from residual gravity anomalies to corresponding height anomalies. The effect of the global model and of the topography are added back subsequently. As at present, high resolution terrain data are lacking for large parts of Europe, test calculations were performed for Scandinavia using the $5' \times 5'$ terrain model. However, the internal comparisons as well as an independent control with GPS/leveling results from the European GPS traverse showed clearly, that the global $5' \times 5'$ DTM can not be used for this purpose due to the poor resolution and quality of this model. Large aliasing errors were found, which disturb the solution in a systematic manner. This can be explained by the fact that the classical terrain corrections, entering directly or indirectly in the computations, are always too small when terrain models with an inadequate resolution are used. The test solution using the $5' \times 5'$ DTM did not show any improvement as compared to the existing solution EGG1. Thus it is clear that high resolution terrain data plays a crucial role for the development of an improved quasigeoid for Europe.

Therefore, in a first iteration, a new quasigeoid solution has been computed for central and northern Europe, as only here we have reliable terrain and gravity data of sufficient resolution as well as GPS/leveling data for an independent control available. For this new calculation the spherical harmonic model OSU89B complete to degree and order 360 (Rapp and Pavlis 1990) was used. Terrain reductions were computed using the RTM technique in connection with a $15' \times 15'$ moving average filter for the computation of the reference topography. The residual gravity data were gridded on a $3' \times 5'$ grid for northern Europe and a $60" \times 100"$ grid for central Europe, depending on the density of the available terrain and gravity data. The field transformation was carried out by FFT. Finally the effect of the global model and of topography were added back. The result is given in Fig. 4.

The quality of the computed solution was evaluated by GPS/leveling data from four different campaings and is summarized in Table II. Here it has to be noted that strict normal heights were only available for the European GPS traverse (Torge et al. 1989) and for the 3D traverse in Switzerland (Wirth 1990), while so-called normal-orthometric heights (approximation to normal heights) and orthometric heights had to be used for the DÖNAV (Seeber 1988) and the ALGESTAR campaign (Marti 1990) respectively. The comparisons were always done using a bias as well as a bias and tilt fit to acount for possibly existing long wavelength gravity model error and inaccuracies in the GPS/leveling results. Using the bias fit only, the comparison of the 1991 quasigeoid with preliminary results for the DÖNAV campaign (computing center IfE, Hannover; Prof. Seeber, personal communication) yielded differences between −0.21 m and +0.14 m. Altogether, 35 stations located in the Federal

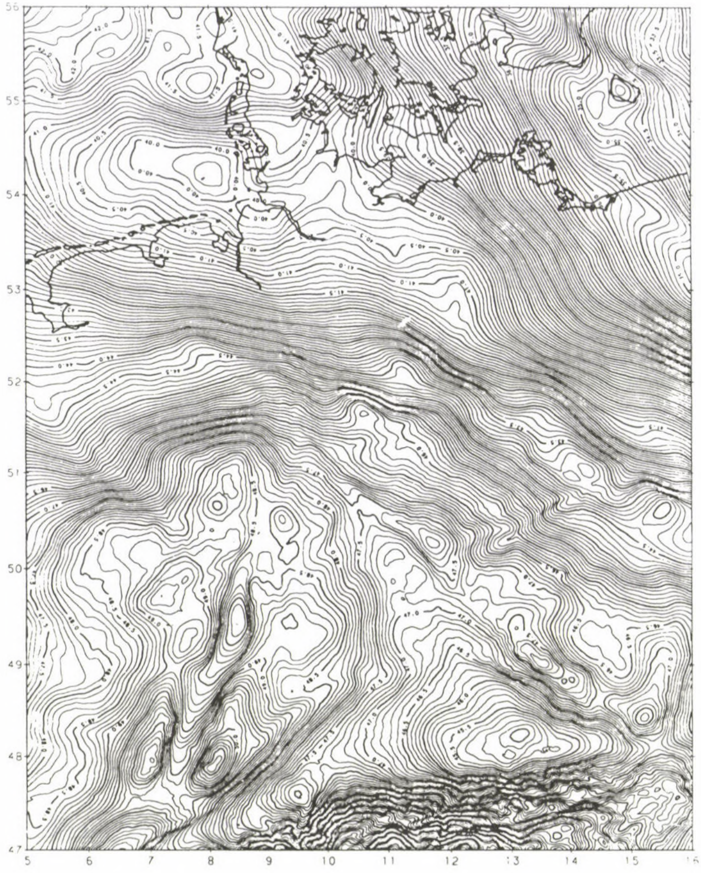


Fig. 4. 1991 quasigeoid solution for Central Europe, contour interval is 0.1 m (Denker and Torge 1992)

Republic of Germany have been used, as at present leveled heights are not available for the other DÖNAV stations. As a long wavelength discrepancy in north-south direction exists between the 3 data sets involved, this may be an indication for long wavelength gravity model errors. The RMS discrepancy then reduces significantly ± 0.086 m for the bias fit to ± 0.048 m for the bias and tilt fit. A second comparison of the 1991 quasigeoid was possible using GPS/leveling results from the ALGESTAR campaign (Marti 1990) covering entire Switzerland. Again we observe a significant reduction of the RMS difference from ± 0.116 m to ± 0.084 m for the bias versus the bias and tilt fit. However, here it must also be considered that strict normal heights were not available, and that the gravimetric quasigeoid may suffer from lacking high resolution terrain data in Italy, which were not at our disposal when the solution was done. For the 30 km long 3D traverse in Switzerland (Wirth 1990), located in an extremely rugged area of the Alps with elevations above 4000 m, we get a RMS

Table II. Statistics for the comparison of the 1991 quasigeoid solution for central and northern Europe with different GPS/leveling results

Description	No. of Stat.	Bias Fit			Bias + Tilt Fit		
		Rms [m]	Min. [m]	Max. [m]	Rms [m]	Min. [m]	Max. [m]
DÖNAV	35	0.086	−0.214	+0.144	0.048	−0.074	+0.096
ALGESTAR (CH)	37	0.116	−0.262	+0.260	0.084	−0.162	+0.186
3D Traverse (CH)	28	0.082	−0.177	+0.133	0.043	−0.110	+0.088
GPS Traverse	67	0.169	−0.364	+0.274	0.158	−0.333	+0.277
	32	0.136	−0.343	+0.167	0.071	−0.147	+0.126
	27	0.063	−0.131	+0.115	0.049	−0.093	+0.102

difference of ±0.082 m and ±0.043 m for the bias and bias and tilt fit respectively. The comparison of the 1991 quasigeoid with the European GPS traverse is shown in Fig. 5. Here it has to be mentioned that high resolution terrain and gravity data were only available for the southern part, while $3' \times 5'$ mean gravity data had to be used for Scandinavia. Therefore, as was to be expected, we get the best results in the southern traverse section with RMS differences for the bias fit of ±0.136 m (32 stations) resp. ±0.063 m (27 stations; 5 southernmost stations eliminated due to possibly existing problems in this traverse part). For the entire traverse with a length of about 3000 km the RMS difference is ±0.169 m. This is a significant improvement as compared to the existing solutions EGG1 and EAGG1 (see Table I). From Fig. 5 we can also observe a very long wavelength discrepancy, especially in Scandinavia. This may come from long wavelength gravity model error, but also from possibly existing small but systematic errors in the $3' \times 5'$ mean gravity data for Scandinavia. A further improvement in the RMS discrepancies is expected by using more high resolution terrain and gravity data in and around the computation area, and by converting the leveling heights to strict normal heights before doing the comparisons.

5. Conclusions

Since the development of the most recent European quasigeoid solutions EGG1 and EAGG1 in 1982 and 1983 respectively, significant new gravity field data sets of high quality have become available. This includes new satellite-only models, new satellite altimeter data from GEOSAT and ERS1, point gravity data, digital terrain models, as well as an increasing number of GPS/leveling control points. Test computations in central and northern Europe promise a significant improvement of the existing European quasigeoid solutions by almost one order of magnitude using these new data sets. The RMS discrepancies versus different GPS/leveling data sets range from ±0.04 to 0.16 m over distances from a few 100 km to 3000 km. IfE will continue its work as the computing center within the European geoid project. The present solutions will be updated and extended to other areas as soon as the necessary high resolution data sets become available. In the future, more emphasis will be put on the establishment of a high resolution terrain model covering Europe,

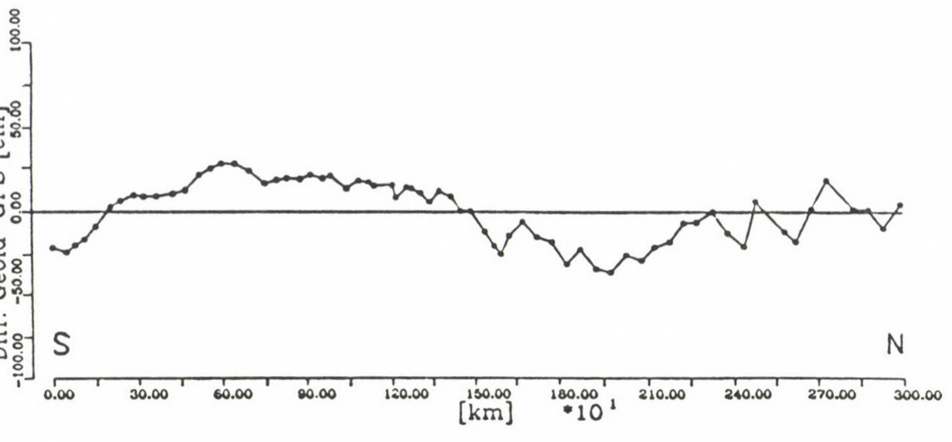


Fig. 5. Comparison of the 1991 quasigeoid solution for central and northern Europe with GPS/leveling results from the European GPS traverse (constant bias substracted)

as these data play a crucial role for the successfulness of the entire project.

Acknowledgments

Thanks are due to Bureau Gravimétrique International, Toulouse, Prof. R.H. Rapp, The Ohio State University, and NASA Goddard Space Flight Center for providing regional and global gravity field data sets as well as to the numerous agencies and individuals who released high resolution gravity and terrain data for the geoid project. Thanks go to Dr. Vermeer for his cooperation within the frame of the IAG Subcommission for the Geoid in Europe. Geoid calculations at IfE Hannover have been substantially supported by the German Research Society (DFG), through several research grants To 46 and within the framework of SFB 149.

References

Bašić T 1989: Untersuchungen zur regionalen Geoidbestimmung mit "dm" Genauigkeit. Wiss. Arb. d. Fachr. Verm.wesen, Univ. Hannover, Nr. 157.

Bašić T, Denker H, Knudsen P, Solheim D, Torge W 1989: In: IAG Symposia Proceed., Rummel R and Hipkin R G eds., Vol. 103, Gravity, Gradiometry and Gravimetry, Springer Verlag, Berlin – Heidelberg – New York, 109–118.

Bomford G 1972: The astrogeodetic geoid in Europe and connected areas 1971. Travaux de l'AIG24, Paris, 357-370.

Brennecke J, Lelgemann D, Reinhart E, Torge W, Weber G, Wenzel H G 1983: A European astro-gravimetric geoid, Deutsche Geod. Komm., Reihe B, Nr. 269, Frankfurt a. M.

Denker H 1988: Hochauflösende regionale Schwerefeldbestimmung mit gravimetrischen und topographischen Daten. Wiss. Arbeiten d. Fachr. Verm.wesen, Univ. Hannover, Nr. 156, Hannover

Denker H 1989: A new gravimetric quasigeoid for the Federal Republic of Germany. Deutsche Geod. Komm., Reihe B, Nr. 291. München

Denker H 1990: In: IAG Symposia Proceed., Rapp R H and Sansó F eds., Vol. 106, Determination of the Geoid — Present and Future, Springer Verlag, New York – Berlin – Heidelberg, 152–159.

Denker H, Torge W 1992: Recent developments of the European Geoid project. Proceed. "First Continental Workshop on the Geoid in European — Towards a Precise Pan-European Reference Geoid for the Nineties", Prague, 1992 (in press)

Denker H, Lelgemann D, Torge W, Weber G, Wenzel H G 1986: In: Proceed. Intern. Symp. on the Definition of the Geoid. Ist. Geogr. Mil. Ital., Florence, 1, 207–222.

Forsberg R 1990: In: Proceed. 11th General Meeting of the Nordic Geodetic Commission, 1990, Copenhagen, 75–89.

Forsberg R, Tscherning C C 1981: *J. Geoph. Res.*, 86, 7843–7854.

Levallois J J, Monge H 1978: In: Proceed. Intern. Symp. on the Geoid in Europe and Mediterranean Area, Ancona, 1978. Soc. Italiana di Fotogrammetria e Topografica, 153–164.

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 J W, 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, Washington D.C.

Marsh J G, Lerch F J, Putney B H, Felsentreger T L, Sanchez B V, Klosko S M, Patel G B, Robbins J W, Williamson R G, Engelis T E, Eddy W F, Chandler N L, Chinn D S, Kapoor S, Rachlin K E, Braatz L E, Pavlis E C 1989: The GEM-T2 gravitational model. NASA Technical Memorandum 100746, Greenbelt/Maryland

Marti U 1990: ALGESTAR — Satellitengestützte Geoidbestimmung in der Schweiz mit GPS. Geodätisch-geophysikalische Arbeiten in der Schweiz. Schweiz. Geod. Komm., Vol. 41, Zürich

Rapp R H, Cruz J Y 1986: Spherical harmonic expansions of the earth's gravitational potential to degree 360 using 30' mean anomalies. Dept. of Geod. Science and Surveying, The Ohio State Univ., Rep. No. 376, Columbus, Ohio

Rapp R H, Pavlis N K 1990: *J. Geoph. Res.*, 95, 21885–21911.

Seeber G 1988: In: Lecture Notes in Earth Sciences, Groten E and Strauss R eds., 19, 201–209.

Tanni L 1949: The regional rise of the geoid in Central Europe. Publ. Isost. Inst. of IAG, No. 22, Helsinki

Tscherning C C, Rapp R H 1974: Closed covariance expressions for gravity anomalies, geoid undulations, and deflections of the vertical implied by anomaly degree variance models. Dept. of Geod. Science, The Ohio State Univ., Rep. No. 208, Columbus, Ohio

Torge W 1987: In: Determination of Heights and Height Changes, Pelzer H, Niemeier W eds., F. Dümmler Verlag, Bonn, 69–82.

Torge W, Denker H 1990: In: IAG Symposia Proceed., Rapp R H and Sansó F eds., Vol. 106, Determination of the Geoid — Present and Future, Springer Verlag, New York – Berlin – Heidelberg, 287–295.

Torge W, Weber G, Wenczel H G, 1982: In: Proceed. 2nd Intern. Symp. on the Geoid in Europe and Mediterranean Area, Rome 1982. Ist. Geogr. Mil. Ital., Florence, 437–460.

Torge W, Bašić T, Denker H, Doliff J, Wenzel H G 1989: Long range geoid control through the European GPS traverse. Deutsche Geod. Komm., Reihe B, Nr. 290, München

Wenzel H G 1985: Hochauflösende Kugelfunktionsmodelle für das Gravitationspotential der Erde. Wiss. Arb. d. Fachr. Verm.wesen, Univ. Hannover, Nr. 137, Hannover

Wirth B 1990: Höhensysteme, Schwerepotentiale und Niveauflächen: Systematische Untersuchungen zur zukünftigen GPS-gestützten Höhenbestimmung in der Schweiz. Geodätisch-geophysikalische Arbeiten in der Schweiz. Schweiz. Geod. Komm. Vol. 42, Zürich

Wolf H 1949: Die angenäherte Bestimmung des Geoids mittels astronomischen Nivellements im Bereich des Zentraleuropäischen Netzes. Veröff. Inst. f. Erdmessung, Heft 6, Teil I, 57–71, Bamberg

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 167-170 (1992)

EVALUATION OF A GPS SINGLE FREQUENCY MODEL

K A ABDALLA[1]

[Manuscript received May 14, 1989]

A single frequency ionospheric model is evaluated. The investigation showed that the model has limitations in accurately quantifying the effect of ionospheric propagation delay on GPS observations.

The quality of the model prediction to the propagation delay effects on point positions was found to have a maximum of 30 % of dual frequency compensation. This was much less than that originally expected. The results showed a substantial lack of consistency in the prediction of the single frequency ionospheric model used.

Keywords: GPS; ionospheric model; positioning; single frequency model

1. Introduction

The study highlights the results, effectiveness and limitations of the GPS single frequency ionospheric model, described by Klobuchar (1982) and ICD-GPS-200 (1981).

Many investigators have indicated that the dual frequency availability of GPS should be used to accurately measure the ionospheric propagation delay effect on GPS observations. But sometimes noise due to the combination of the two L-band signals (L1 and L2) could cause large residuals as compared to single frequency observations (Abdalla 1987).

For single frequency GPS observations there is no method for automatically determining accurate information required about the ionosphere and its variations. The efforts are mainly concentrated in this regard on developing mathematical models to compensate for ionospheric refraction effect on satellite observations.

This study also focuses on Klobuchar's (1982) single frequency model, its use and limitation in determining the effects of ionospheric refraction on point positions. The model used takes advantage of the ionospheric parameters broad-casted in the GPS navigation message. The model also uses some numerical values and algorithms to estimate the obliquity factor, time and latitude variations.

TI 4100 GPS observations are provided by the Geophysical Service International (GSI) of Bedford, from meaurements taken at Liddington Castle, United Kingdom.

[1]Department of Surveying Engineering, Faculty of Engineering and Architecture, University of Khartoum, P.O.B. 321, Sudan

2. The model equations

The ionospheric propagation delay, Δt, is given by Klobuchar (1982) as:

$$\Delta t = F\left(5\cdot 10^{-9} + \sum_{m=0}^{3} \alpha_n \Phi_m^n \left(1 - \frac{x^2}{2} + \frac{x^4}{24}\right)\right), \quad \text{for } x < 1.57 \text{ sec}$$

or

$$\Delta t = F(5\cdot 10^{-9}), \quad x > 1.57 \text{ sec}$$

and

$$\begin{aligned}
x &= (t-50400)/\Big(\sum_{n=0}^{3} \beta_n \Phi_m^n\Big) \\
F &= 1 + 16(0.53 - E)^3 \\
t &= 4.32\cdot 10^4 \lambda_i + \text{GPS time} \quad \text{if } t > 86400, \text{ then } t = t - 86400 \text{ is used} \\
\Phi_m &= \Phi_i + 0.064\cos(\lambda_i - 1.617) \\
\lambda_i &= \lambda_u + \psi \sin A / \cos \Phi_i \\
\Phi_i &= \begin{cases} \Phi_u + \psi\cos A, & \text{if } \Phi_u < 0.416 \\ \Phi_u & \text{elsewhere} \end{cases}
\end{aligned}$$

and

$$\psi = -0.022 + 0.0137/(E + 0.11)$$

where

- α_n = four coefficients representing the vertical delay (obtained from GPS navigation message)
- β_n = four coefficients representing the normalized period (obtained from GPS navigation message)
- E = satellite elevation angles
- A = satellite azimuth
- Φ_u = user's geodetic latitude (semi-circles)
- λ_u = user's geodetic longitude (semi-circles)
- t = local time
- Φ_m = geomagnetic latitude of the ionospheric intersection point
- Φ_i = geodetic latitude of the ionospheric intersection point
- ψ = earth central angle between user's position and the ionospheric intersection point
- λ_i = geodetic longitude of the ionospheric intersection point.

3. Comparison with dual frequency

To examine the efficiency of the model, TI 4100 dual observations where used. Comparisons of the single frequency model results and dual frequency ionospheric compensation are shown in Table I.

Table I. Ionospheric corrections (in metre) to point positions determined by TI 4100 observations, using dual frequency algorithm and the single frequency ionospheric model

Geocentric coordinates	Corrections dual frequency	Corrections ionospheric model
X	2.280	0.731
Y	−2.672	−0.802
Z	5.402	0.982
X	1.722	0.601
Y	−2.769	0.957
Z	5.417	0.676
X	10.757	0.429
Y	−3.533	0.830
Z	19.872	0.289
X	2.264	0.434
Y	−2.707	0.830
Z	5.292	0.289
X	5.497	0.429
Y	7.259	0.526
Z	8.564	0.285
X	5.508	0.553
Y	−6.825	0.797
Z	7.694	0.828
X	6.502	0.521
Y	−5.881	0.783
Z	7.795	0.825
X	4.171	0.549
Y	−5.805	−0.794
Z	7.786	0.828

The comparison indicated that the model can predict a maximum of about 30 % of the effect of ionospheric propagation delay on pseudo-range measurements. This showed that the efficiency of the model is much less than that originally expected (*i.e.* 50 %) and reported in Klobuchar (1982). This could be due to the limitation of the model in accurately quantifying the ionospheric propagation effects on GPS satellite observations. According to Klobuchar (1982) the model has some limitations including that the algorithm does not accurately represent the behaviour of the ionosphere in and near the equatorial region of the earth, and to some extent in high latitude-regions, further, due to simplified approximations used in the model to compute earth centred angle, user geomagnetic latitude and local time.

4. The model use and limitation

In spite of the variability and irregularity of the characteristics of the ionosphere, the model is found to predict only about 30 % of the ionospheric propagation delay effects determined by using the basic dual frequency algorithm. So, the results call for further modifications to improve the obliquity factor and the user's geomagnetic latitude and local time.

In general, the results showed some of the problems related to the use of modelling technique in estimating ionospheric refraction effects such as:

(i) The model cannot estimate any sudden change or irregularity in the characteristics of the ionosphere. For example, the electron content of the ionosphere is not always smoothly distributed, but often irregularities may occur.

(ii) Using predictive methods over-modelling problems may appear.

(iii) The quality of the prediction of the ionospheric propagation delay obtained by using the model was much less than the expected values. This indicates the lack of consistency of the results obtained by using the model.

5. Conclusions

The results of the investigations into GPS dual frequency observations and the application of the single frequency ionospheric model described in Klobuchar (1982) showed some of the problems related to the use of modelling techniques.

One of the main disadvantage of the model is that sometimes the resulting range and consequently point positions are corrupted by using the ionospheric effects obtained from the model.

The best prediction quality of the ionospheric model was about 30 %, much less than expected, due to the fact that the model cannot predict irregularities and sudden changes in the characteristics of the ionosphere.

References

Abdalla K A 1987: An analysis of transit and GPS point positioning results. P.HD thesis, University of Newcastle upon Tyne, U.K.

ICD-GPS-200 1981: NAVSTAR GPS Space Segment Navigation User Interfaces. Interface Control Document 03953, Satellite System Division, Rockwell International Corporation

Klobuchar J A 1982: Ionospheric corrections for single frequency user of the global positioning system. Proc. National Telecommunications Symposium

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 171-175 (1992)

PULSATIONS AT THE GEOMAGNETIC EQUATOR

J VERŐ[1], LE MINH TRIET[2], J SZENDRŐI[1]

[Manuscript received August 5, 1991]

A sample of earth current records have been compared from an equatorial station in Vietnam with two mid-latitude stations ($L \sim 2$ to 2.6). It was found that similarly to results obtained at a station in India ($L \sim 1.1$), Pc 3 pulsations are practically absent, and a few Pc 4 events can be better correlated with the higher latitude station than with the lower latitude one. In the former case the period at the equatorial station is near to the field line resonance period at the higher latitude station.

Keywords: dynamic spectrum; Equatorial zone; geomagnetic equator; geomagnetic pulsations; Vietnam

In the framework of a cooperation between the Ho Chi Min City Branch of the Vietnamese Scientific Research Centre and the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, an earth current recording station has been established in Baclieu, in the Mekong delta, very near to the geomagnetic equator.

Baclieu is situated at $\varphi = 9°18'$N, $\lambda = 105°43'$E about 100 km South of the geomagnetic equator. The instrument has analogue recording with a speed of 6 mm/min. Due to local problems, a limited amount of records was available. Taking into account the 6 h difference in local time between Baclieu (BAC) and the observatories Nagycenk (NCK) and Niemegk (NGK) in Central Europe, only 0300 - 1000 UT can be used for comparisons at the three stations, as in this interval local times both in Central Europe (4 to 11 h LT) and in BAC (10-17 h LT) correspond to daytime hours.

Thus, the following intervals have been selected:

1990, August 4, 0344-0414 UT
0422-0452 UT
0555-0725 UT

The continuity of these records was interrupted by noise, mostly at BAC. No other records could be used in this comparison, due to noise or to lack of activity.

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

[2]National Centre for Scientific Research of Vietnam Ho Chi Minh City Branch, 1. Mac Dinh Chi, Ho Chi Minh City

Part 1, 0344-0414 UT, August 4, 1990

Part 1 includes the beginning of the daily pulsation activity at NCK, at BAC, it is around its midday maximum 11 h LT (no dynamic spectrum is shown here for this interval). In spite of this, there is more structure in the NCK records in the Pc 3 range, in BAC, the greatest part of the Pc 3-4 activity may be noise. Thus *e.g.* there are two activity peaks around 30 s periods at NCK at 0350 UT which are completely absent in the BAC records. Later on some small maxima in the range 30 to 40 s do not coincide neither in time nor in period at both stations. The rather strong maximum at NCK at 0412, period 30 s, is again absent at BAC.

Maximum amplitudes occurred at both stations during Part 1 at longer periods, at NCK at first around 90 s, towards the end at 70 and 110 s. At BAC, the first part is rather weak in this period range, whereas in the second part there is a splitting similarly to NCK with periods of 75 and 180 s. Thus, in the longer period range, there is some similarity between the two stations, even if it is rather obscure.

Part 2, 0422-0452 UT, August 4, 1990

During Part 2, the pulsation activity increased significantly at NCK (Fig. 1). Long period (65 s) activity was switched off at NCK at 0430 UT, the same switch-off is hardly seen at BAC. Longer period activity restarted soon at NCK, and the lower period limit of the activity shifted towards shorter periods till 0443 UT when it reached periods of about 22 s. A certain decrease of the period is also evident at BAC, but only from 65 to 50 s.

Without details it is only mentioned that the time evolution of the longer period activity is again somewhat more similar.

Part 3, 0555-0725 UT, August 4, 1990

This Part 3 is the most significant in this comparison (Fig. 2), as during this interval pulsation activity may be rather strong at all sites (BAC 12-13 h LT, NCK and NGK 7-8 h LT). Generally the lower 'limit' of the pulsation activity was at BAC at 35-40 s, at NCK at 16-18 s, at NGK at 20 s, with many impulses at even shorter periods (which may be noise). Thus, the absence of Pc 3 pulsations at BAC is evident.

Some details from Part 3: at 0605, in BAC there is an impulse with periods down to 40 s. No simultaneous impulse is seen in the Pc 3 range in the NCK record only at 60 s is some increase; at NGK, a quasi-switch-on can be supposed at the same time, at periods from 20 s (or even shorter) to 60 s.

At 0622 UT, an impulsive event can be seen at BAC with periods around 60 s; both at NCK and NGK, simultaneous events are visible, too, but with periods 30-45 s. Another strong event at 0627 UT (both NCK and NGK) is absent at BAC; around 0640 UT, only a slight trace of an impulse is seen at BAC, while both at NCK and NGK, impulsive pulsation events are evident.

A very characteristic long period event at 0700 UT appeared at each of the three stations, but both at NCK and NGK, it included shorter period amplifications, too,

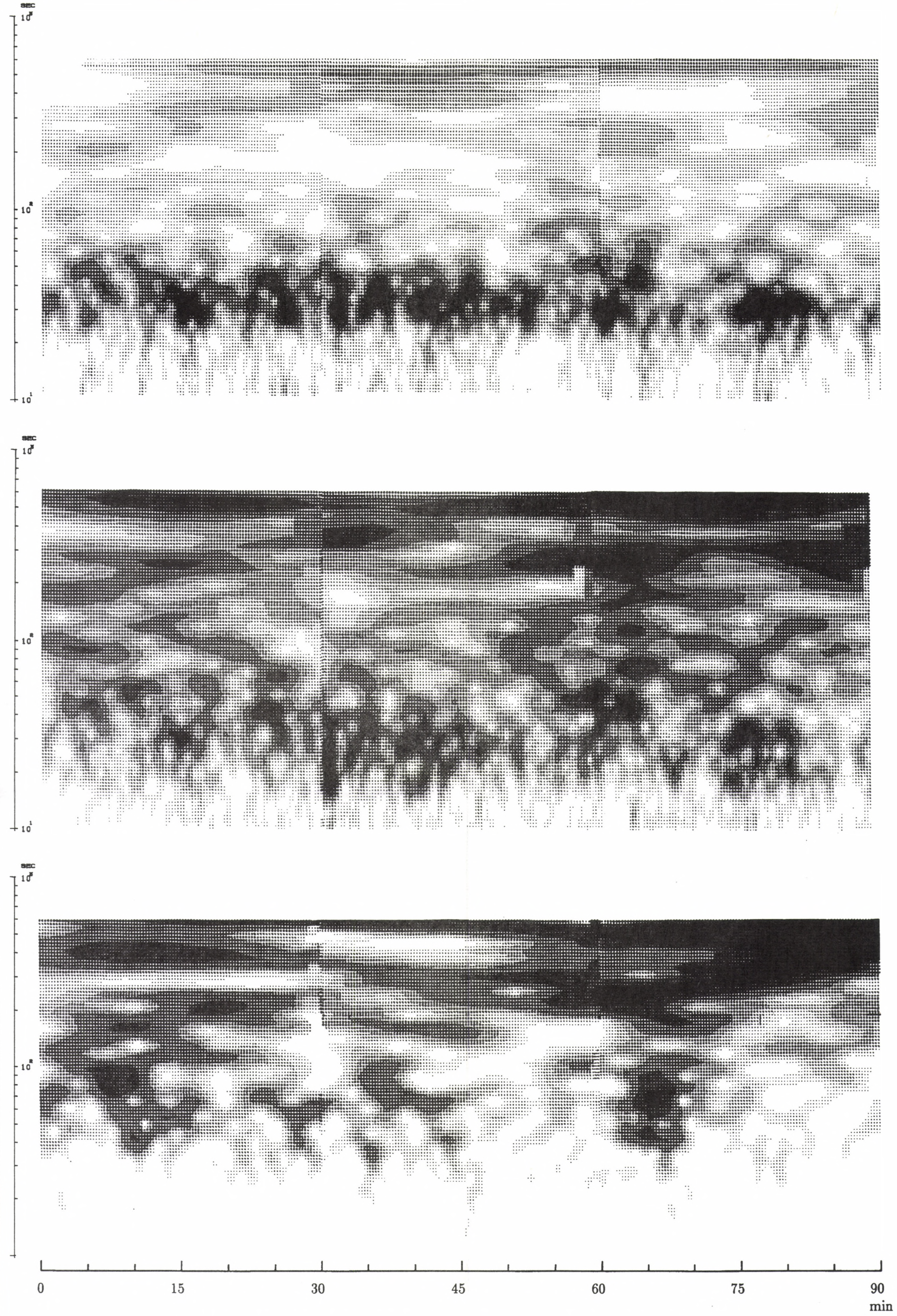


Fig. 2. Dynamic spectra for Part 3, August 4, 1990, 0555–0725 UT, similarly to Fig. 1, but for stations NGK, NCK and BAC (from top to bottom)

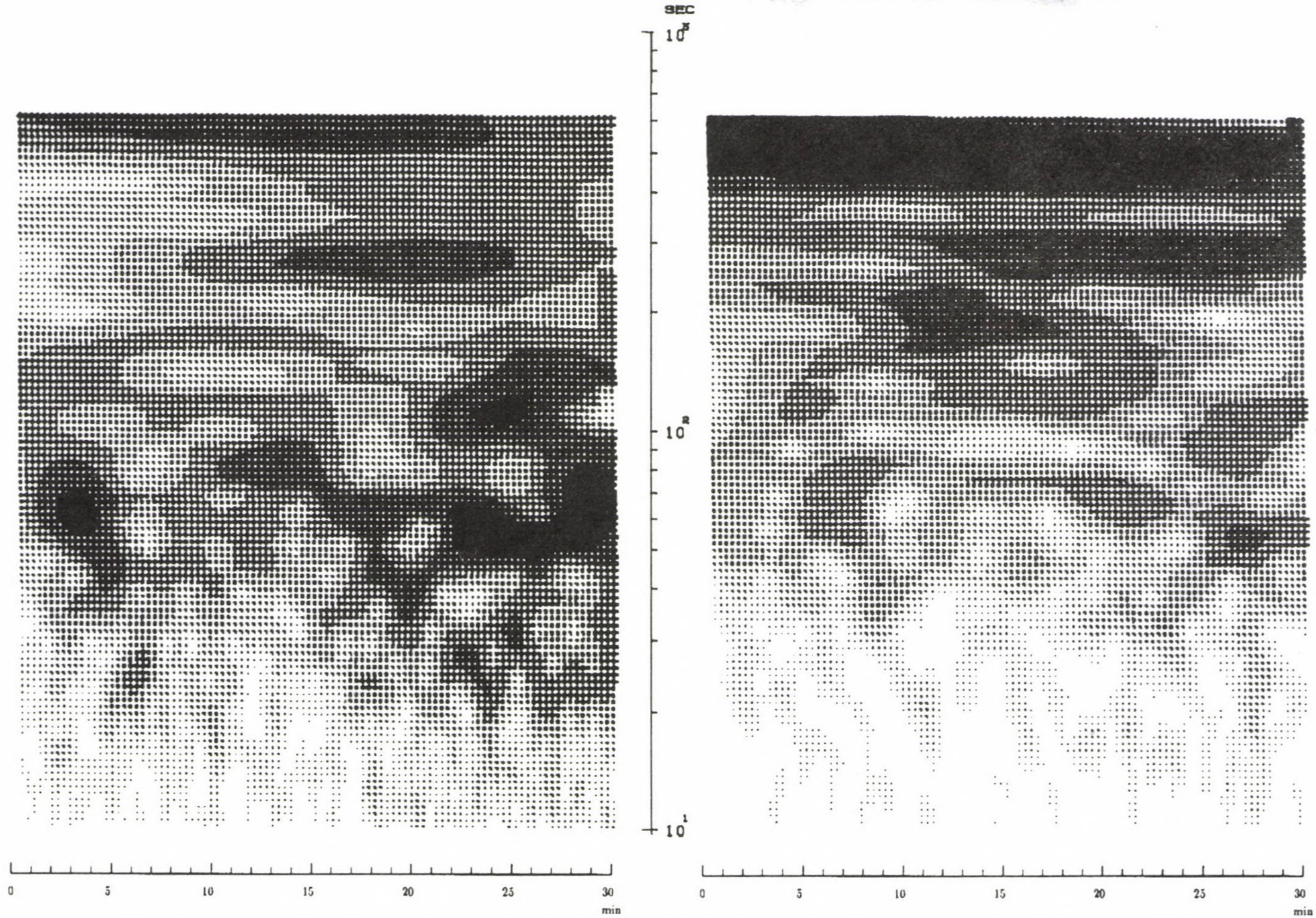


Fig. 1. Dynamic spectrum of the pulsations in the period range 10 s–600 s for Part 2 (August 4, 1990, 0422–0452 UT). Station NCK (left) and BAC (right). Shadings change at powers of 2 (amplitudes)

whereas at BAC, nothing can be seen below 40 s. A strong impulsive NCK event centered around 30 s at 0713 UT is similarly strong at NGK, whereas at BAC, nothing can be seen.

A summary of the complete event can be given as follows: below say 35 to 40 s, the equatorial records have practically no activity, no similarity with low latitude records; the significant difference in longitude, about 6 h should also be hereby considered. The similarity with the station NGK at $L \sim 2.6$ is slightly better than with NCK at $L \sim 2.0$. At longer periods, all three records are more similar, even if certain characteristic changes sometimes are lacking at one of the stations.

The same conclusions, as presented here, were drawn from a comparison of records of a station near Bombay (Dhargapur) in India (Verő et al., 1991) with NCK and NGK. The main facts were: next to no Pc 3 activity at the equatiorial station, some similarity with NGK and only very slight similarity with NCK (in rapid changes of the activity; at DHA, a 24 h record was available, and some slow changes of the activity were similar at all the three stations). It should be added that Dhargapur lies outside of the equatorial electrojet, Baclieu inside of it, nevertheless, the results are very similar.

The interpretation for the present case is correspondingly similar to that in Verő et al., (1991): upstream waves from the pre-shock region cannot penetrate into equatorial regions. Shell (field line) resonances, however, have some similarity with mid-latitude stations at equatorial ones, and in that case, higher latitude stations are more closely correlated with the equatorial ones (records from $L \sim 2$ are less similar to those at $L \sim 1$, than to those at $L \sim 3$).

Reference

Verő J, Holló L, Singh B P 1991: *Acta Geod. Geoph. Mont. Hung.*, 26, 253–262.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 177–183 (1992)

SOME PROBLEMS OF THE 20-YEAR SECULAR VARIATION OF THE GEOMAGNETIC FIELD

LE NGOC THANH[1]

[Manuscript received August 12, 1991]

Based on the conclusion to the common nature and the common source of the 60-year and the 20-year variations of the geomagnetic field, as well as the correlation between the 20-year secular variation of the magnetic field and other natural phenomena such as solar activity, speed of the Earth's rotation and seismic activity, the author reaches the conclusion that all these results furnish a further evidence giving support to Barta's assumption that the geomagnetic short-period secular variation is connected with large-scale movements of the Earth's core. Furthermore, in the paper it is assumed that the source of the 20-year secular variation of the geomagnetic field of global character, like that of the 60-year secular variation, is the circular current determined by Barta.

Keywords: circular current; dispersion; Earth's core; geomagnetic secular; hydromagnetic wave; seismic activity

Introduction

The study of the internal structure of the Earth is one of the fundamental problems of today's geophysics. Since one cannot investigate by direct way the internal structure of the Earth, it is important to study other natural phenomena which are related to the problem to be investigated. The slow change of the main magnetic field of the Earth, geomagnetic secular variation is divided according to Braginsky's classification (1970) into three parts: secular variation with periods of about 8×10^3 years, of 10^3 years, and of 10^2 years and less called geomagnetic short-period secular variation. Based on the study of the geomagnetic secular variation, it is possible to test various hypotheses on the mechanism of generation of the magnetic field of the Earth. In particular, when investigating the geomagnetic short-period secular variation, the internal structure, and moreover the processes taking place in the Earth's interior are to be considered.

Aim of the present paper is to elucidate the results of the 20-year secular variation of the geomagnetic field and other natural phenomena. From these, some conclusions have been drawn related to Barta's hypothesis on the occurrence of the geomagnetic secular variation and the characteristics of the Earth's core.

[1]Department of Geophysics, University Eötvös Loránd, H-1083 Budapest, Ludovika tér 2, Hungary

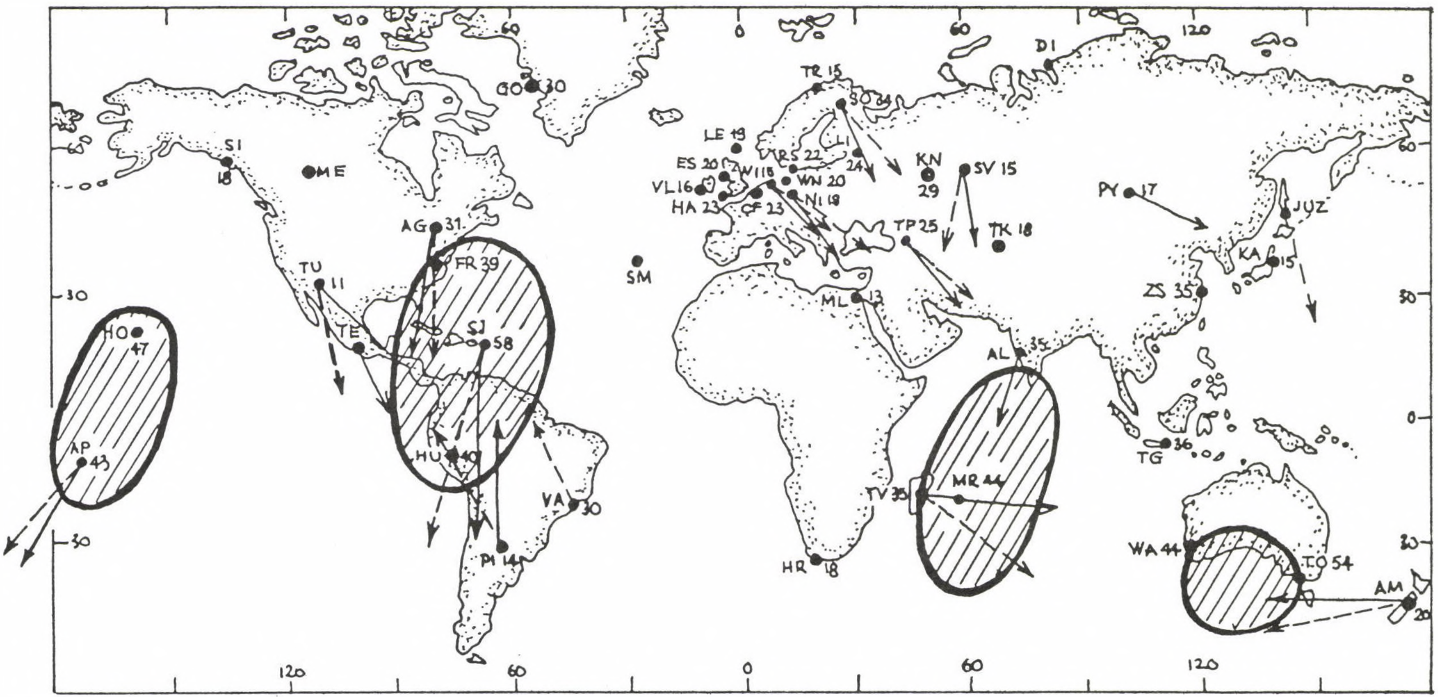


Fig. 1. Spatial distribution of the amplitude of the vertical component of the 20-year secular variation. Numbers – amplitude of the variation in nT; arrows – vectors of polarization of the 20-year (solid lines) and 60-year secular variations (dashed lines) (after Philippov and Rotanova 1988)

The 60-year and the 20-year secular variations of the geomagnetic field

On the basis of the results obtained from various authors, Philippov (1986) has arranged the geomagnetic short-period secular variations into four classes with the characteristic times: 9–12 years, the so-called 11-year secular variation; 18–25 years — the 20-year secular variation; 30–40 years — the 30-year secular variation, and 50–80 years — the 60-year secular variation. Such a dispersion of the periods can be caused by

- — the nature itself of the geomagnetic field and the existence of a temperature gradient in the Earth's liquid core (Le Ngoc Thanh 1991),
- — the various time intervals of magnetic data and the difference in realizing the methods of analysis of the geomagnetic short-period secular variations.

The above facts enable us to assume that the secular variation of the geomagnetic field with a certain period belonging to one of the four classes is the corresponding secular variation.

More recently, geomagnetic secular variation data from 54 observatories allowed Philippov and Rotanova (1988) to make a comparison between the spatial distribution of the 60-year and the 20-year secular variations. From this, the authors concluded that the 60-year and the 20-year secular variations of the geomagnetic field are of the same nature, and consequently they assumed that these components have the same source. An example of the spatial distribution of the amplitude of the vertical component of the 20-year secular variation and the major axes of the ellipses of polarization, in the horizontal plane, of the 20-year and the 60-year secular variations at some observatories are given in Fig. 1.

The period of about 50 years of the geomagnetic field has been since long detected by Barta (1956, 1963, 1981, 1985) and it has particular significance in this theory. Following Philippov's classification, the geomagnetic secular variation with this period of about 50 years may be considered as the 60-year secular variation. One of the important results in Barta's theory is that there exists at the mantle-core boundary a circular current, the centre of which is below Pakistan ($\varphi = 28°02'$ N, $\lambda = 62°17'$ E). This circular current is the source of the global secular variation of the geomagnetic field with a period of about 50 years observed all over the Earth. Another result is that the geomagnetic short-period secular variation is connected with large-scale movements of the Earth's core due to a tide-like effect. It is necessary to emphasize that the latter conclusion is drawn from the correlation of the period of about 50 years, on the basis of experimental data obtained from 1900 to 1950, between geomagnetic short-period secular variations and variations in the movement of the Earth, namely, variations in the speed of the Earth's rotation, changes in amplitude of the oscillation of polar altitude and sea level oscillation (Fig. 2).

The 20-year secular variation of the geomagnetic field, as a matter of fact, has been long studied. The existence of its source within the Earth is now generally accepted (e.g., Alldredge 1977, Jin 1977, Courtillot and Le Mouël 1988). The investigation of the spatial distribution indicates that the 20-year secular variation

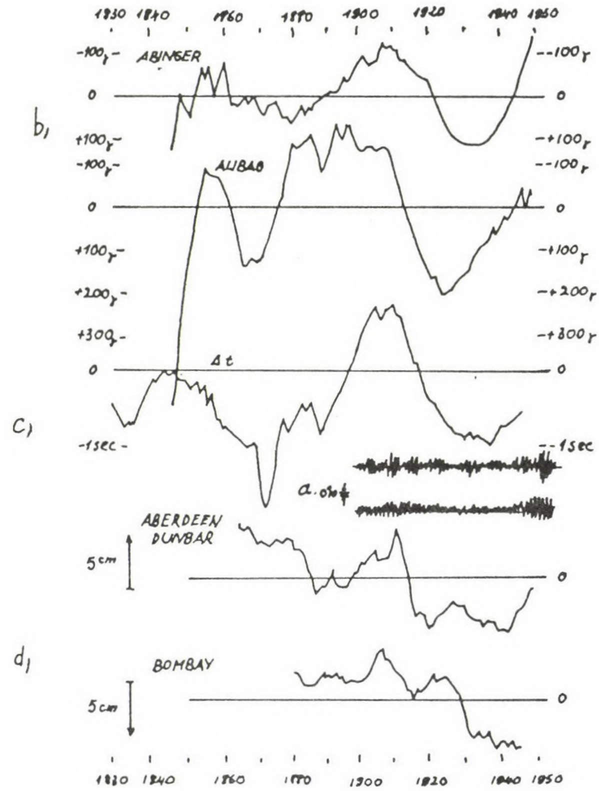


Fig. 2. a) Variation of the amplitude of the oscillation of polar altitude; b) Wave of the geomagnetic field; c) Variation of the Earth's rotation; d) Variation of sea level (after Barta 1963)

of the geomagnetic field consists of two parts: one of them is related to a global source, while the other part is related to local sources, the positions of which are determined by the foci. These foci of the 20-year secular variation are mainly located at low latitudes (Fig. 1). In fact, more than 30 years ago, Barta detected at equatorial magnetic observatories the secular variation of the geomagnetic field with a period of about 25 years (Barta 1956). On the basis of the magnetohydrodynamics of the Earth's liquid core Le Minh Triet (1972, 1983) showed that the 20-year secular variation of the geomagnetic field occurs in consequence of the dispersion of hydromagnetic waves propagating along the magnetic toroidal field in the mantle-core boundary layer. This dispersion of the hydromagnetic waves is caused by the Earth's rotation. The conclusive advantage of this interpretation is the predominant existence of the 20-year secular variation at low latitudes and in the zones near the equator as said above.

Besides the 20-year secular variation of the geomagnetic field, there are other natural phenomena manifesting also the variations with periods of about 25 years. the correlation between their variations and the 20-year secular variation of the geo-

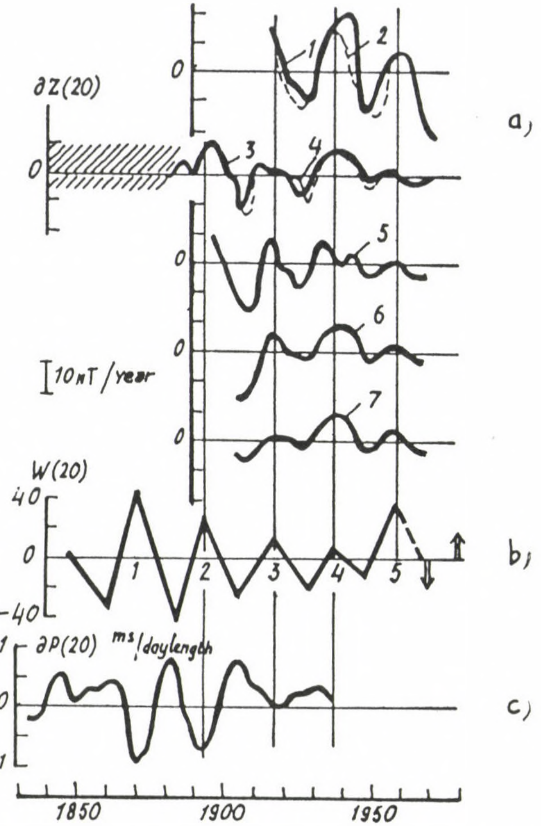


Fig. 3. a) ∂Z: the variation in the vertical components of the geomagnetic field at various observatories. 1. San-Juan, 2. Apia, 3. Hartland, 4. Eskdalemuir, 5. Chambon-la-Forêt, 6. Rude Skov, 7. Niemegk. b) W: sunspot number. c) ∂P: irregular velocity variations of the Earth's rotation (after Rivin 1985b)

magnetic field has become the subject studied in detail. In particular, Rivin (1985a, b) has showed the correlation between the 20-year secular variation of the sunspot activity, the length of day and the intensity of the geomagnetic field (Fig. 3). He interpreted this correlation on the assumption that the Earth's core would execute forced oscillations under the Sun's action.

Considering the correlation between the 24-year variation of the geomagnetic field, detected on the basis of the magnetic data from 1905 to 1974, and the 25-year variation in worldwide seismic energy release, established by experimental data from 1800 to 1974, Kalinin and Rozanova (1984a, b) assume that the variations of these geophysical phenomena are results of a common cause-mechanical oscillation of the Earth's inner core. The 24-year libration of the Earth's poles was attributed by Busse to the mechanical oscillation of the inner core as well (Kalinin and Rozanova 1984a).

Remark and conclusion

From the results analysed above, namely, from the common nature and the common source of the 60-year and the 20-year secular variation of the geomagnetic field, from Barta's conclusion concerning the geomagnetic short-period secular variation, as well as from the correlation between the geomagnetic field and other natural phenomena detected by Rivin, Kalinin and Rozanova, conclusions can be drawn for the 20-year secular variation of the geomagnetic field. All these facts confirm — in connection with Barta's hypothesis — that

1. the geomagnetic short-period variation, including the 60-year and 20-year secular variation is connected with large-scale movements of the Earth's core. In other words, the correlation between the 20-year secular variation of the geomagnetic field and other geophysical phenomena, namely speed of the Earth's rotation and planetary seismic activity, furnishes a further evidence supporting Barta's assumption on the connection between geomagnetic secular short-period variations and the large-scale movement of the Earth's core.

2. it is probable that, a part of the gobal 20-year secular variation of the geomagnetic field is produced by the circular current, as a source, at the mantle-core boundary surface, the centre of which is located under Pakistan.

It is not to be precluded that the appearance of the 20-year secular variation of the geomagnetic field at low latitudes and the zones near the equator is due to the dispersion of the hydromagnetic waves propagating along the magnetic toroidal field in the core-mantle boundary layer under the action of the Earth's rotation.

In addition, these results also allow us to assume that the large-scale movement of the Earth's core could cause the 20-year secular variation of geophysical phenomena. The change in the large-scale movement of the Earth's core varies the speed of the Earth's rotation and stimulates in the upper layer of the Earth's crust seismic activity with a period of about 25 years.

Acknowledgements

I wish to express my special thanks to Professor Gy Barta of Institute of Geophysics, University Eötvös Loránd, for his kind support and valuable comments in improving the manuscript.

The author is deeply obliged to the Hungarian Academy of Sciences for a grant to study at University Eötvös Loránd.

References

Alldredge L R 1977: *J. Geomag. Geoelectr.*, 29, 123–135.
Barta Gy 1956: *Acta Geol. Hung.*, 4, 15–42.
Barta Gy 1963: *Annales Univ. Sci. Budapesti, de Roland Eötvös Nom., Sec. Geol.*, 7, 71–81.
Barta Gy 1981: *Annales Univ. Sci. Budapesti, de Roland Eötvös Nom., Sec. Geol.*, 13, 3–24.
Barta Gy 1985: *Acta Geod. Geoph. Mont. Hung.*, 18, 29–59.
Braginsky S I 1970: *Geomagn. Aeron.*, 10, 221–233.
Courtillot V, Le Mouël J L 1988: *Ann. Rev. Earth Planet. Sci.*, 16, 389–476.

Jin R S 1977: *J. Geoph. Res.*, 82, 828–834.
Kalinin Yu D, Rozanova T S 1984a: *Geomagn. Aeron.*, 24, 113–116.
Kalinin Yu D, Rozanova T S 1984b: *Geomagn. Aeron.*, 24, 266–268.
Le Minh Triet 1972: *Acta Geod. Geoph. Mont. Hung.*, 7, 147–154.
Le Minh Triet 1983: *Acta Geod. Geoph. Mont. Hung.*, 18, 97–106.
Le Ngoc Thanh 1991: *Annales Univ. Sci. Budapesti, de Roland Eötvös Nom.* (in press)
Philippov S V 1986: On problems of investigations of the secular short-period variations. Problems of generation of the magnetic field, Izmiran, USSR Academy of Sci., 63–72.
Philippov S V, Rotanova N M 1988: *Geomagn. Aeron.*, 28, 286–293.
Rivin Yu R 1985a: *Geomagn. Aeron.*, 25, 833–836.
Rivin Yu R 1985b: *Geomagn. Aeron.*, 25, 990–995.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 185-194 (1992)

MFV-FILTERING TO SUPPRESS ERRORS — A COMPARISON WITH MEDIAN FILTERS

B HAJAGOS[1] and F STEINER[1]

[Manuscript received August 17, 1991]

The concept of the filtering based on the most frequent value, the so-called MFV-filtering is defined in this paper, similarly to the median filtering. In contrast to normal linear filtering, both filterings are resistant (i.e. the appearance of outliers does not lead to distortions in the filtered series which may distort the interpretation neither in the case of median-, nor of MFV-filtering), but in the case of the more frequently occurring distribution types of the errors the latters can be more effectively eliminated, as with median filtering.

Keywords: error distribution; error suppression; median filtering; most frequent value; robust methods

1. Different kinds of filtering of time series

Moving averages have been used from time immemorial for smoothing the random variations in the "time series" x_i: the time series according to

$$x_i(\text{mean}) = \frac{1}{2n+1}\sum_{k=-n}^{+n} x_{i+k} \tag{1}$$

represents clearly (or at least more clearly) the essential features of the original time series x_i (in the most simple case e.g. its trend). The length $N = 2n + 1$ is called "size of the window". Similarly to this, we consider as "time series" not only series changing in time, but also in function of any other physical quantity if they are sampled according to this function, thus e.g. well logging data sampled in equidistant depth intervals are also called "time series".

In a next step, simple averages are substituted by weighted averages:

$$x_i(\text{w-mean}) = \sum_{k=-n}^{+n} c_k x_{i+k} \tag{2}$$

("w-mean" is the short form for weighted mean). If $\sum_{k=-n}^{n} c_k = 1$ is fulfilled, one speaks also of moving or running averages similarly to the method described by Eq. (1).

[1]Geophysical Institute, University Miskolc, H-3515 Miskolc, Egyetemváros, Hungary

Akadémiai Kiadó, Budapest

The weights c_i can be selected according to different points of view, and the data series x_i(w-mean) computed by Eq. (2) is — even in the strict mathematical sense of the word — a filtered form of the time series x_i, the operation given by Eq. (2) is a filtering, and the vector c_k is a filter (if the weights c_k decrease linearly in both directions from the central value c_o, the filter is a triangular one). Equation (1) defines obviously a filter, too, only the weights are here identical: $c_k = 1/(2n+1)$.

This simplest case shows most clearly one of the dangers of such a kind of smoothing: if the time series x_i contains outliers, too, then this fact influences in the filtered data series a $(2n+1)$ data long part even if the outlier is not an extremal one. (It is a paradox, but latter case may be less dangerous, as in the case of a value differing by orders of magnitude from the others a change will be present in the filtered series which can be excluded on the basis of practical considerations, while a certain amplitude due to a not extremely differring x_i from the rest of the series may lead to erroneous consequences and deductions from the filtered data series, if the statistics is used without precaution.)

Thus it is not surprising that uptodate statistical works do not propose moving averages first in smoothing data series. Tukey (1977) proposes first medians from three elements each; in a general case of the window length this means the determination of median

$$x_i(\text{med}) = \text{med}\,\{x_{i-n}, x_{i-n+1}, \ldots, X_{i-1}, x_i, x_{i+1}, \ldots, x_{i+n}\} \tag{3}$$

for each i being the search for the medium value in the ordered sample of the length $(2n+1)$.

This procedure is evidently insensitive against outliers moreover it filters the random elements from the time series similarly (or even better) than the strictly mathematical filtering according to Eq. (2). From the point of view of the statistical practice it is correct to call the series x_i(med) a filtered series, even if the filtering after Eq. (3) must not be considered filtering in the strict mathematical sense of the word.

This fact explains that the expression "median filtering" has been introduced relatively late for the operation x_i(med). Tukey did not use this expression in 1977, but Gallagher and Wise already used it in 1981, when they discussed the peculiarities and theory of the median filtering. It has not more sense to discuss the correctness of this expression, similarly e.g. to the meaning of the word "norm" which is also used with an extended meaning, as it was done by Wolf (1983) when he used it for the characterization of deviation systems; such a general concept of the norm helped Steiner (1991) to present basic statistical connections in a large-size table. The introduction of the expression "median filtering" makes it necessary to distinguish traditional filtering by some method; thus the traditional filtering according to Eq. (2) will be called "linear filtering" (in spite of the fact that in the strict mathematical sense this expresion is a pleonasm).

An increasing number of publications deals with the advantages of median filtering, what is hardly surprising for those who are acquainted not only with the resistance of statistical algorithms based on the minimization of the L_1-norm of the deviations (i.e. its low sensitivity against outliers), but also with its robustness (i.e.

the fact that the efficiency of the method depends only slightly on the type of the error distribution; algorithms based on the traditional L_2-norm are, however, far from being robust).

Robust methods do significantly differ in their robustness, too, as shown e.g. by Fig. 58 in Steiner (1988a). We quote here the last sentence from Bednar's Summary (1983) what refers to that paper, but outlines future development, too: *"It represents a preliminary step toward refinement of existing processing techniques in that the further expansion of the use of robust methods in velocity analysis, stacking and other situations in which averages can be replaced by more robust estimators may offer tremenous advantages in extracting more information from the given data."* The advantages to be expected are the greater the higher the robustness is; Steiner (1988a) has shown that the "most frequent value" (the minimum of the so-called P-norm of the deviations, see e.g. a great table in Steiner (1991)), is a very robust method. It is therefore evident to put the question: would not it be even more advantageous in comparison to the median filtering, to compute the most frequent value within the window of the length $(2n+1)$. Dresen et al. (1989) called this method "MFV (**M**ost **F**requent **V**alue) algorithm", thus the filtered values by this method are denoted by $x_i(\mathrm{MFV})$:

$$x_i(\mathrm{MFV}) = \mathrm{MFV}\{x_{i-n}, x_{i-n+1}, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{i+n}\}, \tag{4}$$

the filtering according to Eq. (4) is called "MFV-filtering".

The right hand side of Eq. (4) means the computation method of the MFV-values by a black box similarly to the right hand side of Eq. (3). While in the latter case a few words are enough to define the algorithm, the algorithm described as MFV { } is too complicated to be presented here: this would make the understanding of the present paper more difficult for those who are already acquainted with the MFV-method and use it. Therefore the algorithm to compute the MFV-filtered values, i.e. the values $x_i(\mathrm{MVF})$ are given in the Appendix.

2. Comparison of MFV-filtering and median filtering for some types of error distributions

The efficiency of the methods to produce the data series x_i(filtered), i.e. if $x_i(\mathrm{MFV}), x_i(\mathrm{med})$ or just $x_i(\mathrm{mean})$ would be the most efficient in a certain case in eliminating noise from the given series, depends clearly on the type of the error distribution superimposed on the original series. In order to see most evidently the difference in the noise elimination for different types of filtering, the possible simplest case is to be studied: In that case errors are independent and the errorfree values are constant (the latter condition is fulfilled e.g. in the logging practice by a homogeneous layer).

The dependence on the type should not be decided too quickly: it is theoretically insufficient to consider asymptotic characteristics only. Nevertheless, Table I contains data for large windows N, i.e. for the asymptotic case (for the simplest case defined previously), namely for four types of the distribution, the error ratios for the MFV- and median-filtered series. "Error ratios" are e.g. the ratios of the probable errors (semi-interquartile ranges) q, in the asymptotic case all error definitions yield the same ratios. $k = 2$ denotes the standard variant of the MFV-filtering, $k = 1$ that variant of the MFV-filtering when it is justified to use the Cauchy-model, be it due to the heavy tails of the basic distribution (an example was presented by Landy et al. (1982)), be it due to the non negligible occurrence of outliers (as discussed by Tarantola (1987)). It is comforting to see in that Table — at least for high values of N — that these error ratios are greater than 1 except for Laplace-distribution, i.e. the primary errors are more effectively reduced by MFV-filtering than by the median filtering (namely no author has found the Laplace-distribution to occur dominantly, or even frequently).

Table I. Error ratios of MFV-filtered and median-filtered data for high values of N, for four types of distribution of the errors in the original data series (The efficiency of the MFV-filtering referred to the median-filtering is given in brackets)

	$k = 2$		$k = 1$	
geostatistical	1.1267	(127 %)	1.0728	(115 %)
Gaussian	1.1975	(143 %)	1.0762	(116 %)
Cauchy	1.0472	(110 %)	1.1107	(123 %)
Laplace	0.8904	(79 %)	0.9283	(86 %)

The results of Table I coincide with the values (Steiner 1991) of the statistical efficiencies known for the asymptotic case the geostatistical distribution will be later defined by its density function, Eq. (5). The samples, however, contain in each case a finite number of elements N and the limit where the asymptotical laws are fulfilled within the practical accuracy limits, do change from case to case. Figures 32 and 33 in Steiner (1988a) are examples which show what a difference in the properties of the estimations may occur if the number of elements is low: in these examples the average shows only indications of the asymptotic behaviour at a number of elements of $N = 25$.

The properties of samples with a low number of elements have been studied by Monte-Carlo methods by Andrews et al. (1972). The choice of these eminent authors to study samples with a low number of elements shows that this question is worth investigating from the point of view of the algorithms used, moreover, it also shows that in lack of corresponding theoretical results — with a few exceptions — they can be studied only by Monte-Carlo methods. The investigations in connection with the different kinds of filtering have been also carried out by Monte-Carlo methods. (Cases with $N \to \infty$ can be easily treated analytically using the notion of the asymptotic scatter. The data in Table I were computed by this method as ratios of asymptotic scatters; exact numerical values and definitions are found by Steiner 1990.)

Before a summary of the results, the incorrectness of an opinion often mentioned in the statistical practice is to be emphasized. According to this opinion, it is impossible to conclude from a sample consisting of only a few elements to the type of distribution of the random variable from which it is taken. As a false consequence, it is considered as of no importance for which type of distribution the chosen statistical algorithm is an optimum. The incorrectness of the deduction is evident from the point of view of mathematical statistics (and the following results can be considered as concrete convincing refutals), nevertheless, it must not be resented if those, who apply statistical methods in the practice, and start from the supposition $N \to \infty$ (as anybody else), get irresolute or even think that statistical methods are unapplicable if the sample is small, or even more, very small. Not only the theory of mathematical statistics, but also completely heuristic considerations — such considerations which led e.g. to the definition of the "most frequent value" (Steiner 1990) suppose implicitely or explicitely samples with a high number of elements, even the attribute "most frequent" was born from this concept. The applicability of the algorithm on this basis, as given in the Appendix is, however, not bound to high numbers of elements: computations can be made with the MFV-algorithm already with $N = 3$ (according to Eq. (3), with $n = 1$), too. It may be necessary to use such a small size window: the window size is determined mostly by the variability of the quantity to be detected in the case of both the MFV-filtering and median- and mean-filterings, too.

The minimum window size $n = 1$ ($N = 3$) has been used in the Monte-Carlo investigations together with $n = 2$ ($N = 5$), $n = 4$ ($N = 9$), as well as $n = 9$ ($N = 19$). As the aim of the investigations was to compare the efficiencies of different filters for different types of errors, a set of $(1000 + 2n)$ values of x_i were generated (in the most simple case as described earlier), and as characteristic value the ratio of the semi-interquartile range of the two kinds of filtered series was accepted each consisting of 1000–1000 values, in other words the ratio of the probable errors in the two series was accepted. It is to be remarked that the ratios of other error parameters, e.g. the semi-intersextile range and the mean square error were also observed, but these parameters did not led to conclusions differing from those got by the semi-interquartile range.

The investigations of the efficiency of the filters were started with the geostatistical distribution defined by the density function

$$f_{st}(x) = \frac{3}{4} \cdot \frac{1}{(1+x^2) \cdot \sqrt{1+x^2}}. \tag{5}$$

The result of this detailed Monte-Carlo investigation can be summarized as follows: the advantages of the MFV-filtering are obvious against both the median- and mean-filtering, independently of N. This statement has a special significance as Steiner (1990) has given results concerning the occurrence of different types of distributions (referring to Dutter 1986/87) and found that the most frequently occurring type of the error distribution in geostatistics is that defined by f_{st}.

At first some partial results will be presented. Figure 1 is the histogram of the ratios of the q-values (probable errors) determined in the previously described method: they are computed from two filtered series of 1000 values both (x_i(med) and x_i(MFV)) for $N = 5$. (The abscissa is indicated in a short form, $q_{\text{med}}/q_{\text{MFV}}$).) Only 7 of the 80 values were less than 1, the smallest of them was 0.94, the greatest 1.23. Two thirds of the ratios were between 1.02 and 1.11, with a median at 1.06.

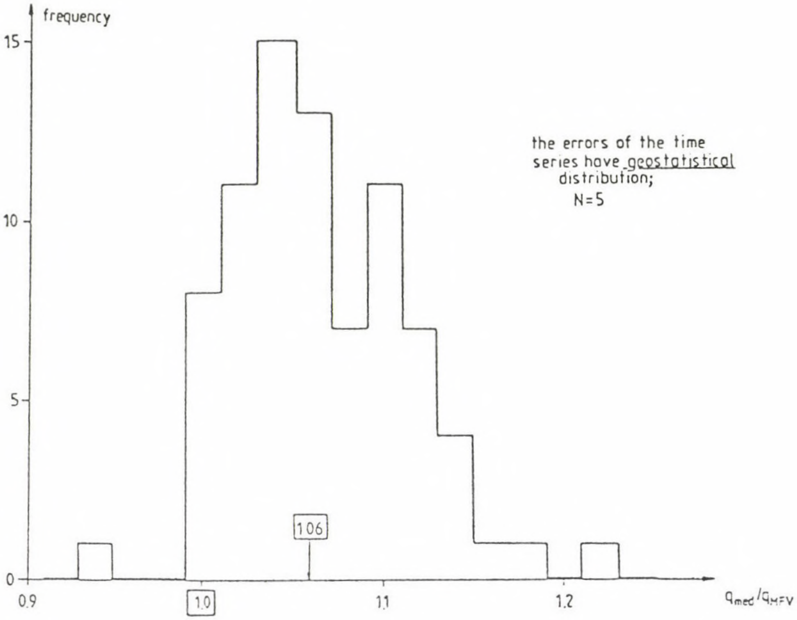


Fig. 1.

This median is smaller than that given in Table I, nevertheless, together with the value for $N = 3$ (1.02) it shows that the empirical values of $q_{\rm med}/q_{\rm MFV}$ tend even at low values of N toward the asymptotic value.

These are the results for the geostatistical distribution — let us see now those for the classical Gaussian distribution. The histogram represented in Fig. 2 for $N = 3$, as well as the even greater value of the median of the q-values (1.16) shows even more significant advantages of the MFV-filtering with respect of the median filtering. By comparing this value of 1.16 with the value in Table I (1.20) the rapid convergence of the ratios to the asymptotic value is evident even at low values of N. (This latter statement is even more valid for the Laplace-distribution, as the Monte Carlo-method yielded a value of 0.90 for the median of the ratios $q_{\rm med}/q_{\rm MFV}$, while the value in Table I is 0.89.)

We conclude the results with those for the Cauchy-distribution. It is interesting to note that the convergence to the asymptotic value of $q_{\rm med}/q_{\rm MFV}$ is here slower than at the geostatistical distribution: values around 1 were obtained even at $N = 5$; at $N = 19$ the empirical value, 1.19, was nearly identical with the asymptotic value, 1.20.

Thus in conclusion, Table I gives correct information on the efficiency of MFV-filtering with respect to median filtering, even in the case of small filter lengths, too. This fact refers not only to the result that MFV-filtering decreases the errors in the overwhelming majority of cases more efficiently than median filtering, but

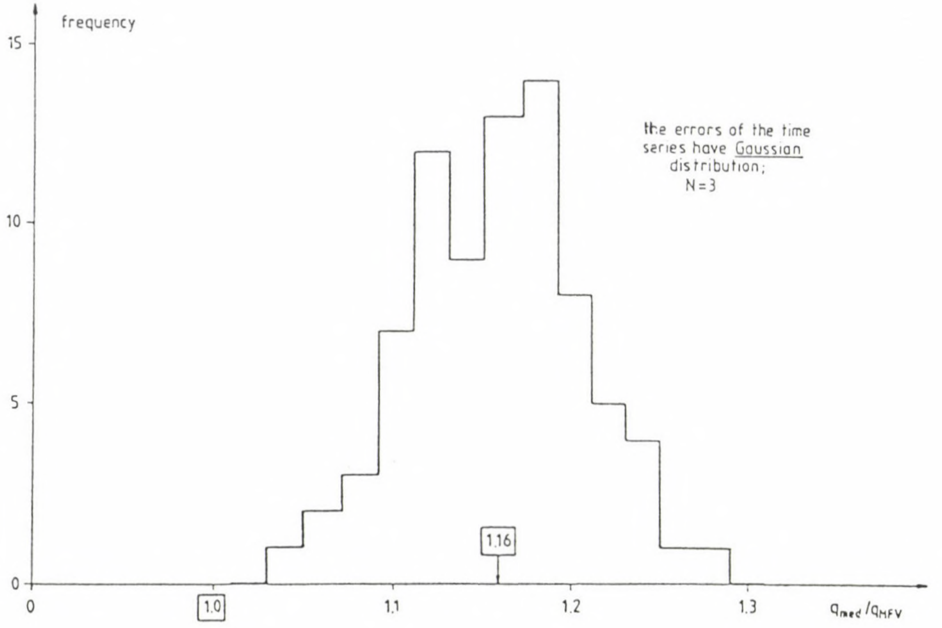


Fig. 2.

also to the numerical values which are more or less acceptable approximations even at $N = 3$, i.e. at the possible minimum size of the filter.

3. Remarks

Equation (2) describes a weighted alternative of Eq. (1) with the weights c_k; the same weights c_k can be used for a weighted MFV, too (Steiner 1982). The time series constructed so is denoted by x_i(w-MFV). The Appendix contains the algorithm for the determination of the filtered data series x_i(w-MFV), too. There is no problem in the determination of the weighted median: after ordering, an element is looked for, for which the sums of the weights before and after that element are less than 0.5, respectively, the average of two neighborous elements is taken, for which the sum of the weights is 0.5 both before the latter and after the former element. An investigation was made for triangular weights (the weights decrease linearly from the central value in both directions) for different window sizes. Since the general advantages of the MFV-filtering appeared in these weighted filterings as well as in the unweighted ones, a detailed account on these results seems to be superfluous.

Similarly to the linear- and median-filterings, MFV-filtering has also a variant with several variables. Steiner and Zilahi-Sebess (1988) propose (p. 41) a two-variable form of the MFV-filtering with bell-shaped weights for the smoothing of gravity maps. As the manuscript of the book has been closed in 1981, the term

"MFV-filter" did not appear in it, while the remark about this type of filtering, namely that it is "a very time-consuming process" lost its validity: the program based on the algorithm of the MFV-filtering, as presented in the Appendix needs more computer time only in relation to the median filtering (and to the linear filtering). The absolute computer time is even for PC-s acceptable, especially if the advantages of the MFV-filtering are taken into account, as they are presented in this paper.

Appendix

How to carry out MFV-filtering?

For the sake of simplicity let us use new indices for those x_i-s which are involved in the computation of a single filtered MFV-value; the new indices are $j = 1$ to N. The most frequent value, M, of these x_j-s is the filtered value x_i (MFV).

A double iteration is needed to compute M, and these iterations yields in addition to M the dihesion ε of the x_j-s (N items). This dihesion is the semi-interquartile range of a Cauchy-distribution (i.e. its probable error) which is most similar to that defined by the data series (Csernyák 1973).

The most simple method for this calculation is the so-called ping-pong iteration, during which on one of the branches M, on the other ε is more and more approximated (in both cases making a single step only) according to the formulas

$$M = \frac{\sum_{j=1}^{N} \frac{x_j}{(k\varepsilon)^2+(x_j-M)^2}}{\sum_{j=1}^{n} \frac{1}{(k\varepsilon)^2+(x_j-M)^2}} \tag{6}$$

$$\varepsilon^2 = 3\frac{\sum_{j=1}^{N} \frac{(x_j-M)^2}{\left[\varepsilon^2+(x_j-M)^2\right]^2}}{\sum_{j=1}^{N} \frac{1}{\left[\varepsilon^2+(x_j-M)^2\right]^2}}. \tag{7}$$

M and ε values at the right side are from the previous pair of iteration steps, those at the left side are new values, the value of k is in the standard case 2. The iteration should be stopped if Eq. (6) yields no more significant change in M (the limit of a significant change is determined for the actual problem to be solved). For the sake of simplicity the approximate values were also denoted by M: the most frequent value M is naturally that value which precisely fulfils Eq. (6) with the dihesion ε fulfilling Eq. (7). It is best to start the ping-pong iteration by using for ε^2 0.75 times the difference of the maximum and minimum values of x_j squared, the initial value of M should be the average or median of the x_j values. The ε-branch should be iterated first. Concerning the number of the iteration steps, it is usual to prescribe the maximum allowed number of the steps in which case the program is stopped by this number (let us say, after 5, 10 or 20 steps) and not by the limit prescribed for M. (The histograms in Figs 1 and 2 were e.g. computed with 5 steps; the initial value of M was the average of the data.) It depends on the actual problem how far the number of steps can be reduced, or even if the number of steps can be anyway limited. Due to the small number of data in the MFV-filtering, a limit (not necessarily a narrow one) should be given even if the computer time for MFV filtering is negligible in an absolute sense, too, or in relation to the computer time of the other tasks of the problem.

If a small part of the N data is very concentrated, much more than characteristic for the complete data set, the iteration according to Eq. (7) may lead to a dihesion characteristic for this small part of the set (the Cauchy distribution is fitted to this small set, see again Csernyák 1973), and in consequence, M will not be characteristic for all the N data. This situation does not occur frequently, nevertheless, it is advisable to organize the program so (especially for high values of N) that such cases should be excluded. The easiest way to do this is to compute after each iteration the effective number of cases being

$$n_{\text{eff}} = \sum_{j=1}^{N} \frac{\varepsilon^2}{\varepsilon^2 + (x_j - M)^2} \tag{8}$$

and to see if it is not less than $N/2$; if this would occur, then ε from the last but one iteration is to be used in the iteration after Eq. (6) to find M.

If weighted MFV-filtering is to be carried out with the weights c_j, then Eqs (6) and (7) are substituted by

$$M = \frac{\sum_{j=1}^{N} \frac{c_j \cdot x_j}{(k\varepsilon)^2 + (x_j - M)^2}}{\sum_{j=1}^{N} \frac{c_j}{(k\varepsilon)^2 + (x_j - M)^2}} \tag{9}$$

$$\varepsilon^2 = 3\frac{\sum_{j=1}^{N} \frac{c_j \cdot (x_j - M)^2}{\left[\varepsilon^2 + (x_j - M)^2\right]^2}}{\sum_{j=1}^{N} \frac{1}{\left[\varepsilon^2 + (x_j - M)^2\right]^2}}. \tag{10}$$

The result is then naturally one element of the weighted MFV-filtering (x_i(w-MFV)).

This method of the ping-pong iteration fulfils the practical demand for the velocity of the convergence (both for x_i(MFV) and x_i(w-MFV)). If the convergence must be notwithstanding improved, it can be done from some pairs of the iterations using *e.g.* Aitken and Steffensen's (1972) method (Steiner (1988b) gave deductions for this method concerning the computation of the most frequent value). In this case both on the M- and ε-branches of the iteration not one, but 3 steps are carried out each one after the other using Eqs (6) and (7) (or in a weighted case, Eqs (9) and (10)). It the results are denoted by M_1, M_2 and M_3, as well as $\varepsilon_1, \varepsilon_2$ and ε_3, then the values

$$M = \frac{M_1 \cdot M_3 - M_2^2}{M_1 + M_3 - 2M_2} \tag{11}$$

$$\varepsilon = \frac{\varepsilon_1 \cdot \varepsilon_3 - \varepsilon_2^2}{\varepsilon_1 + \varepsilon_3 - 2\varepsilon_2} \tag{12}$$

will converge rapidly. In this case special protections are to be included into the program to ensure that the differences should not be less than prescribed limits. The consideration of the number of the effective values (Eq. (8)) is to be made in this case so that if at any value of $\varepsilon_1, \varepsilon_2, \varepsilon_3$ the value n_{eff} gets less than $N/2$, then the iteration is to be continued with this value, as a constant on the M-branch. If $n_{\text{eff}} < N/2$ gets valid with ε from Eq. (12) for the first time, then either ε_3 is to be considered as a constant value, or in the case of more exact calculations, $\varepsilon_4, \varepsilon_5 \dots$ are to be computed with ε_3 by Eq. (7) or

Eq. (10), as long as the relation $n_{\text{eff}} < N/2$ is fulfilled. The last but one value will be the future constant ε.

As a last remark, we propose to use in Eqs (6) and (9), respectively, instead of the standard $k = 2, k = 1$ in the case of heavy tails. (See the data for the Cauchy distribution in Table I.) Moreover, it is advisable to use the median as starting M-value instead of the average which may be far from the most frequent value due to outliers or/and due to the tails of the primary distribution, as in that case too many steps of the iteration would be necessary. (The number of the iteration steps on the ε-branch may be anyway increased significantly in such a case by the fact that the initial value of ε may be by orders of magnitude greater than the correct one.) In the case of primary errors of Cauchy distribution, the test computations were started from the median (with $k = 1$) in the present study, too.

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.

Bednar J B 1983: *Geophysics*, 48, 1598–1610.

Csernyák L 1973: *Acta Geod. Geoph. Month. Hung.*, 8, 317–401.

Dresen L, Csókás J, Dobróka M, Gyulai Á, Ormos T 1989: In: Proceedings of the 34th International Geophysical Symposium, MGE, Budapest

Dutter R 1986/87: Mathematische Methoden in der Montangeologie. Vorlesungsnotizen, Leoben

Gallagher N C, Wise G L 1981: In: IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-29, No. 6, December

Landy I, Lantos M 1982: Publ. of the Univ. of Miskolc, Series A, Mining, 37, 29–32.

Steiner F 1982: Publ. of the Univ. of Miskolc, Series A, Mining, 37, 17–28.

Steiner F 1988a: *Geophysical Transactions*, 34, 139–260.

Steiner F 1988b: Publ. of the Univ. of Miskolc, Series A, Mining, 45, 71–90.

Steiner F 1990: Principles of geostatistics (in Hungarian). Tankönyvkiadó, Budapest

Steiner F (Ed.) 1991: The Most Frequent Value. Akadémiai Kiadó, Budapest

Steiner F, Zilahi-Sebess L 1988: Interpretation of Filtered Gravity Maps. Akadémiai Kiadó, Budapest

Tarantola A 1987: Inverse Problem Theory, Elsevier

Tukey J K 1977: Exploratory Data Analysis. Addison-Wesley, Reading, Massachusetts

Wolf H 1983: *Vermessung, Photogrammetrie, Kulturtechnik*, 7, 233–240.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 195–202 (1992)

A NORM CONNECTED WITH ROBUST ESTIMATION

L CSERNYÁK[1]

[Manuscript received October 25, 1991]

A norm is defined being in connection with a robust estimation method used in geostatistics and gives theses for that norm.

Keywords: most frequent value; norm; robust estimation

If the density function of the random variable ξ is g and its distribution function G, then its second moment is:

$$m_2 = \int_{-\infty}^{\infty} t^2 g(t)\, dt \tag{1}$$

which gets the form

$$m_2 = \int_0^1 \left(G^{-1}(x)\right)^2 dx = \|G^{-1}\|^2_{L^2}$$

by substituting $x = G(t)$. The connection between the second moment and L^2 is well known, and it can be expressed otherwise, too: be the function f interpreted in the interval [0;1] quadratically integrable (in Lebesgue's sense) and be the distribution function of its values G. Then:

$$\|f\|^2_{L^2} = \int_0^1 f^2(x)\, dx = \int_{-\infty}^{\infty} t^2 G(dt) = \int_{-\infty}^{\infty} t^2 g(t)\, dt. \tag{1a}$$

(For the last equality the existence of the density function should be supposed, too.)

These connections can be generalized by substituting t^2 in Eq. (1) by any continuous function φ of t if the corresponding integrals exist:

$$\int_{-\infty}^{\infty} \varphi(t) g(t)\, dt = \int_0^1 \varphi\left(G^{-1}(x)\right) dx.$$

[1]University of Miskolc, H-3515 Miskolc, Egyetemváros, Hungary

Accordingly Eq. (1a) gets the form:

$$\int_0^1 \varphi(f(x))\,dx = \int_{-\infty}^{\infty} \varphi(t)G(dt) = \int_{-\infty}^{\infty} \varphi(t)g(t)\,dt.$$

Steiner's team introduced a robust estimation method (Steiner 1991). Here a norm will be defined on the set of functions integrable in the interval [0;1] which leads to the norm P_k yielding the parameters of this estimation (see *e.g.* the Table at the end of Steiner's (1991) book).

Be f a function quadratically integrable on the interval [0;1] (in Lebesgue's sense) ($f \in L^2[0;1]$) and

$$F(\varepsilon) = \int_0^1 \frac{\varepsilon^{\frac{3}{2}}}{\varepsilon^2 + f^2(x)}\,dx; \qquad \varepsilon > 0. \tag{2}$$

If function F has any maximum place in the interval $[0,\infty]$ and the number of this places is finite, then be $\varepsilon(f)$ the greatest of them. If F has no maximum place in the interval of interpretation (it can be shown that in that case it is monotonously non-increasing), then be $\varepsilon(f) = 0$. (The notation $\varepsilon(f)$ will be often substituted by ε if no misunderstanding is possible.) As it can be easily shown in such a case F is a derivable function of ε, i.e. it may have a maximum only there where

$$F'(\varepsilon) = \frac{\sqrt{3}}{2}\int_0^1 \frac{3f^2(x) - \varepsilon^2}{\left[\varepsilon^2 + f^2(x)\right]^2}\,dx = 0. \tag{3}$$

Thesis 1. If function f is quadratically integrable on the interval [0,1], then

$$\varepsilon^2(f) \le 3\int_0^1 f^2(x)\,dx = 3\|f\|^2_{L^2}.$$

Proof. Equation (3) yields for $\varepsilon > 0$ with some rearrangement:

$$\varepsilon^2 = 3\frac{\displaystyle\int_0^1 \frac{f^2(x)}{\left[\varepsilon^2 + f^2(x)\right]^2}\,dx}{\displaystyle\int_0^1 \frac{1}{\left[\varepsilon^2 + f^2(x)\right]^2}\,dx}.$$

It is sufficient to prove that

$$\frac{\int\limits_0^1 \frac{f^2(x)}{\left[\varepsilon^2 + f^2(x)\right]^2}\, dx}{\int\limits_0^1 \frac{1}{\left[\varepsilon^2 + f^2(x)\right]^2}\, dx} \leqq \frac{\int\limits_0^1 f^2(t)\, dt}{\int\limits_0^1 1 \cdot dt}.$$

By multiplying both sides of the equation with the denominators:

$$\int\limits_0^1\int\limits_0^1 \frac{f^2(x)}{\left[\varepsilon^2 + f^2(x)\right]^2}\, dx\, dt \leqq \int\limits_0^1\int\limits_0^1 \frac{f^2(t)}{\left[\varepsilon^2 + f^2(x)\right]^2}\, dx\, dt.$$

Thus it should be shown that

$$\int\limits_0^1\int\limits_0^1 \frac{f^2(x) - f^2(t)}{\left[\varepsilon^2 + f^2(x)\right]^2}\, dx\, dt \leqq 0. \tag{4}$$

Be given the unit square $T = \{(x,t) | 0 \leq x \leqq 1;\ 0 \leq t \leqq 1\}$. In the point $(a,b) \in T$ and in the point $(b,a) \in T$ the integrands have opposite signs or it is in both points zero. That means that by mirroring with respect to the diagonal of the unit square, the expression after the integral changes its sign in every point (on the diagonal the values are zero). Be partial sets of the unit square:

$$T_1 = \{(x,t)\,|\,(x,t) \in T,\ f^2(x) > f^2(t)\}$$

$$T_2 = \{(x,t)\,|\,(x,t) \in T,\ f^2(x) < f^2(t)\}$$

(The integrand is positive on T_1, negative on T_2.) According to previous, by mirroring T_1 with respect to the diagonal, the points of T_2 are obtained. To prove the inequality (4) it is sufficient to show that

$$\iint\limits_{T_1} \frac{f^2(x) - f^2(t)}{\left[\varepsilon^2 + f^2(x)\right]^2}\, dx\, dt \leqq \iint\limits_{T_2} \frac{f^2(t) - f^2(x)}{\left[\varepsilon^2 + f^2(x)\right]^2}\, dt\, dx.$$

Due to the definition of T_1 one has:

$$\iint\limits_{T_1} \frac{f^2(x) - f^2(t)}{\left[\varepsilon^2 + f^2(x)\right]^2}\, dx\, dt < \iint\limits_{T_1} \frac{f^2(x) - f^2(t)}{\left[\varepsilon^2 + f^2(t)\right]^2}\, dx\, dt =$$

$$= \iint\limits_{T_2} \frac{f^2(t) - f^2(x)}{\left[\varepsilon^2 + f^2(x)\right]^2}\, dt\, dx.$$

The thesis has been proved.

The statistical counterpart of this thesis is Thesis 3 in Csernyák and Steiner (1991).

Thesis 2. If $f \in L^2[0,1]$ and c is an arbitrary real number, then:

$$\varepsilon(cf) = |c|\varepsilon(f).$$

Proof: If $c \neq 0$, on the basis of Eq. (2), then the greatest maximum of the function

$$F(\varepsilon(cf)) = \int_0^1 \frac{\varepsilon^{\frac{3}{2}}(cf)}{[\varepsilon^2(cf) + c^2 f^2(x)]^2}\,dx = \frac{1}{|c|^{\frac{5}{2}}} \int_0^1 \frac{\left(\frac{\varepsilon(cf)}{|c|}\right)^{\frac{3}{2}}}{\left[\frac{\varepsilon^2(cf)}{c^2} + f^2(x)\right]^2}\,dx$$

occurs at the place $\frac{\varepsilon(cf)}{|c|} = \varepsilon(f)$, and if $\varepsilon(f) = 0$, then $\frac{\varepsilon(cf)}{|c|}$ is also equal zero. If $c = 0$, the statement remains correct, as according to Eq. (2), $\varepsilon(0) = 0$, as in such a case, $F(\varepsilon) = \varepsilon$ is a monotonously decreasing function.

Example 1:

Let us see now an example. Be

$$f(x) = \begin{cases} c & \text{if } \; 0 \leqq x < x_0 < 1 \\ 0 & \text{if } \; x_0 \leqq x \leqq 1 \end{cases}$$

Equation (2) gets in this case the form:

$$\frac{1}{2}\sqrt{\varepsilon}\left[x_0 \frac{3c^2 - \varepsilon^2}{(\varepsilon^2 + c^2)^2} - (1 - x_0)\frac{1}{\varepsilon^2}\right] =$$
$$= \frac{-\varepsilon^4 + \varepsilon^2(5x_0 c - 2c^2) - (1 - x_0)c^4}{2\varepsilon^{\frac{3}{2}}(\varepsilon^2 + c^2)^2} = 0\,.$$

The nominator is zero, if

$$\varepsilon^2 = \frac{c^2(5x_0 - 2) \pm \sqrt{c^4(5x_0 - 2)^2 - 4(1 - x_0)c^4}}{2}\,. \tag{5}$$

The condition for the existence of the solution is that the determinant is non-negative:

$$(5x_0 - 2)^2 - 4(1 - x_0) = 25x_0^2 - 16x_0 \geqq 0.$$

It follows that F has a maximum place on the semi-axis $\varepsilon > 0$ on if

$$x_0 > \frac{16}{25},$$

thus function f may be zero not more than on 9/25 of the interval. It can be easily shown that in such a case F does have a maximum place at the right hand side root of Eq. (5) (being a positive one).

If $x_0 = 1$ (i.e. $f(x) = c$ for all $x \in [0,1]$), then Eq. (5) yields $\varepsilon^2 = 3c^2$. This result shows that the estimation for ε from Thesis 1 cannot be improved.

In this example, if f is equal zero on an interval longer than 9/25, then F is monotonously decreasing, thus $\varepsilon(f) = 0$. This result can be generalized.

Thesis 3. If the function $f \subset L^2[0,1]$ is equal zero on a set $H \subset [0,1]$ having a measure greater than 9/25, then $\varepsilon(f) = 0$.

Proof: According to the suppositions, be f on the set H equal zero and on the set $[0,1] \setminus H$ differing from zero. Without loosing generality it can be supposed that H is an interval and $H = [0,a]$. Then Eq. (4) has the form.

$$\begin{aligned} F'(\varepsilon) &= \int_0^a \frac{3 \cdot 0^2 - \varepsilon^2}{[\varepsilon^2 + 0^2]^2}\, dx + \int_a^1 \frac{3f^2(x) - \varepsilon^2}{[\varepsilon^2 + f^2(x)]^2}\, dx = \\ &= -\frac{a}{\varepsilon^2} + \int_a^1 \frac{3f^2(x) - \varepsilon^2}{[\varepsilon^2 + f^2(x)]^2} = \\ &= \int_a^1 \left\{ \frac{3f^2(x) - \varepsilon^2}{[\varepsilon^2 + f^2(x)]^2} - \frac{a}{(1-a)\varepsilon^2} \right\} dx = \\ &= \int_a^1 \frac{-\varepsilon^4 + \varepsilon^2 f^2(x)(3-5a) - a f^4(x)}{(1-a)\varepsilon^2[\varepsilon^2 + f^2(x)]^2}\, dx = 0. \end{aligned}$$

The denominator of the integrand is positive, the nominator is a second order polynomial of ε^2 (of $f^2(x)$, too); plotted this function, a downward open parabola is obtained. The integral can be zero only if the nominator has positive values, too. Therefore the discriminant, i.e. the expression

$$f^4(x)[(3-5a)^2 - 4a] = f^4(x)25a - 34a + 9 \tag{6}$$

has to be positive. Taking into account that $0 \leqq a < 1$, the expression in Eq. (6) is positive only in the interval

$$\left[0, \frac{9}{25}\right].$$

Thus, if the value a is greatern than 9/25, then $F'(\varepsilon) < 0$ for all $\varepsilon > 0$, i.e. $\varepsilon(f) = 0$. (Csernyák and Steiner (1991) published an analogous result for data systems.)

Let us now define the norm indicated in the title.

DEFINITION: Be $k > 0$ an arbitrary real number. The norm P_k of the function $f \in L^2[0,1]$ is the number

$$\|f\|_{P_k} = \begin{cases} \dfrac{k}{\sqrt{3k^2+1}}\varepsilon(f) \exp\left\{\frac{1}{2}\int\limits_0^1 \ln\left(1 + \frac{f^2(x)}{k^2\varepsilon^2(f)}\right) dx\right\} & \text{if} \quad \varepsilon(f) > 0 \\ \lim\limits_{\varepsilon \to 0} \dfrac{k}{\sqrt{3k^2+1}}\varepsilon(f) \exp\left\{\frac{1}{2}\int\limits_0^1 \ln\left(1 + \frac{f^2(x)}{k^2\varepsilon^2(f)}\right) dx\right\} & \text{if} \quad \varepsilon(f) = 0 \end{cases}$$

Example 2:

Let us compute the norm P_k of the function

$$f(x) = c \qquad x \in [0,1]$$

a) Let us consider first the case, when $c \neq 0$. According the Example 1 $\varepsilon(c) = \sqrt{3|c|}$, thus

$$\begin{aligned}||c||_{P_k} &= \frac{k}{\sqrt{3k^2+1}}\sqrt{3|c|}\exp\left\{\frac{1}{2}\int_0^1 \ln\left(1+\frac{c^2}{k^2 3c^2}\right)dx\right\} = \\ &= \frac{k}{\sqrt{3k^2+1}}\sqrt{3|c|}\cdot\sqrt{1+\frac{1}{3k^2}} = |c|.\end{aligned}$$

b) If $c = 0$, then $\varepsilon(c) = 0$, thus

$$||0||_{P_k} = \lim_{\varepsilon\to 0}\frac{k}{\sqrt{3k^2+1}}\varepsilon\exp\left\{\frac{1}{2}\ln\left(1+\frac{0}{k^2\varepsilon^2}\right)dx\right\} = 0.$$

Thesis 4. If the function $f \in L^2[0,1]$ is zero on a set of positive measure, and $\varepsilon(f) = 0$, then $||f||_{P_k} = 0$.

Proof: Let us split the interval [0,1] into two parts: be

$$H_1 = \{x | f(x) = 0\} \quad \text{and} \quad H_2 = \{x | f(x) \neq 0\}.$$

It the measure of H_2 is $m(H_2) = 0$, then following the line of thoughts in part b) of Example 2 we get the thesis. If the measure of H_2 differs from zero, then

$$\begin{aligned}||f||_{P_k} &= \lim_{\varepsilon\to 0}\frac{k}{\sqrt{3k^2+1}}\varepsilon\exp\left\{\frac{1}{2}\int_{H_1}\ln\left(1+\frac{0}{k^2\varepsilon^2}\right)k + \int_{H_2}\ln\left(1+\frac{f^2(x)}{k^2\varepsilon^2}\right)dx\right\} = \\ &= \frac{k}{\sqrt{3k^2+1}}\exp\left\{\lim_{\varepsilon\to 0}\frac{1}{2}\int_{H_2}\left[\frac{2}{m(H_2)}\ln\varepsilon + \ln\left(1+\frac{f^2(x)}{k^2\varepsilon^2}\right)\right]dx\right\} = \\ &= \frac{k}{\sqrt{3k^2+1}}\exp\lim_{\varepsilon\to 0}\frac{1}{2}\int_{H_2}\left[\ln\varepsilon^{\frac{2}{m(H_2)}}\cdot\left(1+\frac{f^2(x)}{k^2\varepsilon^2}\right)\right]dx.\end{aligned}$$

As $\dfrac{2}{m(H_2)} > 2$

$$\lim_{\varepsilon\to 0}\varepsilon^{\frac{2}{m(H_2)}}\left(1+\frac{f^2(x)}{k^2\varepsilon^2}\right) = 0 \qquad (x \in H_2)$$

(it converges to zero in a monotonously decreasing way). It follows immediately that

$$||f||_{P_k} = 0.$$

This result means that P_k may be zero not only if the function is zero. Thus one of the traditional demands against norms is not fulfilled.

The so-called triangular inequality does not hold, either. By namely

$$f_1(x) = \begin{cases} 1 & \text{if} \quad 0 \leqq x < \frac{1}{2} \\ 0 & \text{if} \quad \frac{1}{2} \leqq x \leqq 1 \end{cases}$$

$$f_2(x) = \begin{cases} 0 & \text{if} \quad 0 \leqq x < \frac{1}{2} \\ 1 & \text{if} \quad \frac{1}{2} \leqq x \leqq 1 \end{cases}$$

On the basis of Theses 3 and 4:

$$||f_1||_{P_k} = ||f_2||_{P_k} = 0.$$

On the basis of Examples 1 and 2, however,

$$||f_1 + f_2||_{P_k} = 1$$

holds.

Thesis 5. If $||f||_{P_k}$ does exist, then $||cf||_{P_k}$ does exist, too, and

$$||cf||_{P_k} = |c| \cdot ||f||_{P_k}$$

Proof: Thesis 2 has shown that $\varepsilon(cf) = |c|\varepsilon(f)$, therefore

$$\begin{aligned} ||cf||_{P_k} &= \frac{k}{\sqrt{3k^2+1}} |c| \varepsilon \exp\left\{ \frac{1}{2} \int_0^1 \ln\left(1 + \frac{c^2 f^2(x)}{k^2 c^2 \varepsilon^2(f)}\right) \right\} = \\ &= |c| \frac{k}{\sqrt{3k^2+1}} \varepsilon(f) \exp\left\{ \frac{1}{2} \int_0^1 \ln\left(1 + \frac{f^2(x)}{k^2 \varepsilon^2(f)}\right) \right\} = \\ &= |c| \cdot ||f||_{P_k}. \end{aligned}$$

Finally let us consider (as a matter of special interest) the next example.

Example 3:

Let us consider the functions

$$f(x) = x^\alpha \qquad \alpha \geq 2 \qquad x \in [0,1]$$

Using the transformation $x = \varepsilon^{1/\alpha} \cdot t$, function $F(\varepsilon)$ in Eq. (2) can be given in the following form:

$$F(\varepsilon) = \varepsilon^{\frac{1}{\alpha} - \frac{1}{2}} \cdot \int_0^{\varepsilon^{-1/\alpha}} \frac{1}{1+t^{2\alpha}} \, dt.$$

By deriving it after ε:

$$F'(\varepsilon) = \varepsilon^{-\frac{3}{2}} \cdot \left[\varepsilon^{\frac{1}{\alpha}} \left(\frac{1}{\alpha} - \frac{1}{2} \right) \int_0^{\varepsilon^{-1/\alpha}} \frac{1}{1+t^{2\alpha}} \, dt - \frac{1}{\alpha} \frac{\varepsilon^2}{1+\varepsilon^2} \right] < 0\,;$$

as F is monotonously decreasing, thus $\varepsilon = 0$. The norm P_k is to be computed as a limit value:

$$\begin{aligned}
\|x^\alpha\|_{P_k} &= \lim_{\varepsilon\to 0} \frac{k\varepsilon}{\sqrt{3k^2+1}} \exp\left\{\frac{1}{2}\int_0^1 \ln\left(1+\frac{x^{2\alpha}}{k^2\varepsilon^2}\right) dx\right\} = \\
&= \lim_{\varepsilon\to 0} \frac{k}{\sqrt{3k^2+1}} \exp\left\{\frac{1}{2}\int_0^1 \ln\left(\varepsilon^2+\frac{x^{2\alpha}}{k^2}\right) dx\right\} = \\
&= \frac{k}{\sqrt{3k^2+1}} \exp\left\{\frac{1}{2}\int_0^1 \ln\left(\frac{x^{2\alpha}}{k^2}\right) dx\right\}.
\end{aligned}$$

Some easy manipulations yield:

$$\|x^\alpha\|_{P_k} = \frac{1}{\sqrt{3k^2+1}} \cdot \frac{1}{e^\alpha} > 0\,.$$

That means that if $\varepsilon(f) = 0$, $\|f\|_{P_k}$ is not necessarily 0. It is to be remarked that

$$\|x^\alpha\|_{L^2} = \frac{1}{\sqrt{2\alpha+1}}$$

therefore Hajagos's conclusion (1992) in item b), namely that $\|f\|_{P_k} \le \|f\|_{L^2}$, can be checked easily, even in form of an inequality, moreover it is valid in the form

$$\|x^\alpha\|_{P_k} < \frac{\sqrt{5}}{e^2} \cdot \|x^\alpha\|_{L^2} = 0.3 \cdot \|x^\alpha\|_{L^2}$$

too.

References

Csernyák L 1991: In: The most frequent value. F Steiner ed., Akadémiai Kiadó, Budapest

Csernyák L, Steiner F 1991: In: The most frequent value. F Steiner ed. Akadémiai Kiadó, Budapest

Hajagos B 1992: *Acta Geod. Geoph. Month. Hung.* (present issue)

Steiner F ed. 1991: The most frequent value. Akadémiai Kiadó, Budapest

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 203–206 (1992)

CONNECTION BETWEEN THE NORMS P_k AND L_2 USED IN STATISTICS

B HAJAGOS[1]

[Manuscript received October 25, 1991]

It is shown that the norm P_k of the deviations is less than their L_2-norm. An important consequence is that convergence after L_2 is also the convergence after P_k.

Keywords: convergence; L_2 statistical norm; P_k statistical norm; robust estimation

By introducing the robust P norm of the deviations (see e.g. the last Table at Steiner 1991) a problem arises about the connection of this norm with the usual ℓ_2 or L_2 norms in the space in which they exist.

The present paper proves the following thesis on this connection.

Be $\epsilon > 0$, $x_i (i = 1, 2, \ldots, n)$ a statistical sample, and $f(x)$ a density function and

$$p_k = \epsilon \exp \left[\frac{1}{2n} \sum_{i=1}^{n} \ln \left[1 + \frac{1}{k^2 \epsilon^2} (x_i - a)^2 \right] \right] \tag{1a}$$

$$P_k = \epsilon \exp \left[\frac{1}{2} \int_{-\infty}^{\infty} \ln \left[1 + \frac{1}{k^2 \epsilon^2} (x - a)^2 \right] f(x)\, dx \right] \tag{1b}$$

respectively, are the P-norms in cases of discrete and continuous distributions and the norms

$$\ell_2 = \left[\frac{1}{n} \sum_{i=1}^{n} (x_i - a)^2 \right]^{1/2} \tag{2a}$$

$$L_2 = \left[\int_{-\infty}^{\infty} (x - a)^2 f(x)\, dx \right]^{1/2} \tag{2b}$$

should exist, then it is true that

$$p_k^2 \le \epsilon^2 + \frac{1}{k^2} \ell_2^2 < \left(\epsilon + \frac{1}{k} \ell_2 \right)^2 \tag{3a}$$

$$P_k^2 \le \epsilon^2 + \frac{1}{k^2} L_2^2 < \left(\epsilon + \frac{1}{k} L_2 \right)^2 \tag{3b}$$

[1]Geophysical Institute, University Miskolc, H-3515 Miskolc, Egyetemváros, Hungary

a) Equation (3a) for the discrete case can be proven elementarily:
Lemma 1. If $a \neq b$ are positive real numbers and n is a natural number, then

$$\frac{n \log a + \log b}{n+1} < \log \frac{na+b}{n+1} \tag{4}$$

Proof: According to Bernoulli's inequality:

$$\left(\frac{na+b}{a(n+1)}\right)^{n+1} = \left(1 + \frac{b-a}{a(n+1)}\right)^{n+1} > 1 + (n+1)\frac{b-a}{a(n+1)} = \frac{b}{a}$$

therefore

$$\frac{na+b}{n+1} > a\left(\frac{b}{a}\right)^{\frac{1}{n+1}} = a^{\frac{n}{n+1}} b^{\frac{1}{n+1}}$$

and taking the logarithms of both sides, Eq. (4) is obtained.
Lemma 2. If $a_i > 0$ are real numbers, $i = 1, 2, \ldots, n$, then

$$\frac{1}{n}\sum_{i=1}^{n} \log a_i \leq \log \frac{1}{n}\sum_{i=1}^{n} a_1 \tag{5}$$

The proof can be carried out by induction after n.
It follows from Lemma 1 that for $n = 2$:

$$\frac{\log a_1 + \log a_2}{2} \leq \log \frac{a_1 + a_2}{2}$$

(the equality is true only if $a_1 = a_2$).
Let us suppose that Eq. (5) is true for $n = m$, then for $n = m+1$ one gets:

$$\frac{m}{m+1}\sum_{i=1}^{m+1} \log a_i = \frac{m}{m+1}\left(\frac{1}{m}\sum_{i=1}^{m} \log a_i\right) + \frac{1}{m+1}\log a_{m+1}$$

using at first the supposition, then Lemma 1, this expression is less than, or equal to,

$$\frac{m}{m+1}\left(\log \frac{1}{m}\sum_{i=1}^{m} a_i\right) + \frac{1}{m+1}\log a_{m+1} \leq \log \frac{1}{m+1}\sum_{i=1}^{m+1} a_i \,.$$

Equation (5) holds therefore for all n-s.
(This thesis follows also from Jensen's inequality, see Jensen (1906)).
Be $a_i = 1 + \left(\frac{x_i - a}{k\epsilon}\right)^2$, then from Lemma 2:

$$\frac{1}{n}\sum_{i=1}^{n} \ln\left[1 + \left(\frac{x_i - a}{k\epsilon}\right)^2\right] \leq \ln \frac{1}{n}\sum_{i=1}^{n}\left[1 + \left(\frac{x_i - a}{k\epsilon}\right)^2\right]$$

and the latter is equal to $\ln\left[1 + \frac{1}{k^2\epsilon^2}\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2\right]$, by raising the basis e to this power and by multiplying it with ϵ^2 the form of Eq. (3a) of the thesis is obtained.

b) Proof of Eq. (3b):

It is advantageous to substritute $t = F(x)$ in Eqs (1b) and (2b), where $F(x)$ is the distribution function. By introducing $\varphi = F^{-1}$, more simple and more generally interpretable formulas are obtained:

$$P_k = \epsilon \exp \left[\frac{1}{2} \int_0^1 \ln \left[1 + \frac{1}{k^2 \epsilon^2} \varphi^2(t) \right] dt \right]$$

$$L_2 = ||\varphi||_{L_2} = \left[\int_0^1 \varphi^2(t)\, dt \right]^{1/2}$$

By comparing them with the norm P_k defined by Csernyák (1992), one gets:

$$||\varphi||_{P_k} = \frac{k}{\sqrt{3k^2+1}} P_k, \tag{6}$$

if ϵ is the dihesion in the defining formulas.

Let us start from Jensen's inequality (Edwards et al. 1965), as it is called in the functional analysis,

$$H \left[\int_0^1 g(t) dt \right] \leq \int_0^1 (g(t))\, dt$$

where g is a positive measurable function on the interval [0,1] and H is (from bottom) a convex non-negative and monotonously increasing function on $(0, \infty)$.

If $H(t) = e^t$, then the inequality has the form:

$$e^{\int_0^1 g(t)\, dt} \leq \int_0^1 e^{g(t)}\, dt, \text{ e.g., } \int_0^1 g(t)\, dt \leq \ln \int_0^1 e^{g(t)}\, dt\,.$$

Be $g(t) = \ln \left[1 + \frac{1}{k^2 \epsilon^2} \varphi^2(t) \right]$, then

$$\int_0^1 \ln \left[1 + \frac{1}{k^2 \epsilon^2} \varphi^2(t) \right] dt \leq \ln \int_0^1 \left[1 + \frac{1}{k^2 \epsilon^2} \varphi^2(t) \right] dt$$

thus:

$$\exp \int_0^1 \ln \left[1 + \frac{1}{k^2 \epsilon^2} \varphi^2(t) \right] dt \leq \int_0^1 \left[1 + \frac{1}{k^2 \epsilon^2} \varphi^2(t) \right] dt\,.$$

By multiplying both sides with ϵ^2, Eq. (3b) is obtained. If both sides are multiplied by $\frac{k^2 \epsilon^2}{3k^2+1}$, and by taking Eq. (6) into account the result is obtained in the form

$$||\varphi||^2_{P_k} \leq \frac{k^2}{3k^2+1} \left[\epsilon^2 + \frac{1}{k^2} ||\varphi||^2_{L_2} \right]. \tag{7}$$

This is more general than Eq. (3b), as it is also valid if $\epsilon = 0$.

Consequences and remarks:

a) The sharpness of the inequality $P_k^2 \leq \epsilon^2 + \frac{1}{k^2}L_2^2$ is proved by the relations $P_k \geq \epsilon$ and $P_\infty = \epsilon$ where the latter two can be easily proven.

b) According to Thesis 1 in Csernyák (1992), $\epsilon^2 \leq 3||\varphi||_{L_2}^2$ and thus Eq. (7) yields:

$$||\varphi||_{P_k}^2 \leq \frac{k^2}{3k^2+1}\left(3+\frac{1}{k^2}\right)||\varphi||_{L_2}^2 = ||\varphi||_{L_2}^2$$

therfore the convergence after the norm L_2 means also convergence after the norm P_k.

c) If that P_k, where a is the most frequent value and ϵ the dihesion (Steiner 1991), is considered as uncertainty U_k then the inequalities

$$U_k^2 \leq \left(3+\frac{1}{k^2}\right)L_2^2, \quad \text{i.e.,} \quad U_k \leq \sqrt{3+\frac{1}{k^2}}$$

are obtained. For example $U \equiv U_2 \leq \sqrt{3.25}L_2 = 1.8L_2$.

d) L_2 means in the previous the quadratic deviation with respect to the most frequent value and it is equal with the scatter in the case of symmetrical distribution. In such a case $U_k \leq \sigma\sqrt{3+\frac{1}{k^2}}$. The inequality cannot be made more sharp, as in the case of a distribution defined by two equal Dirac-δ-s, $U_k = \sigma\sqrt{3+\frac{1}{k^2}}$, e.g. $U = 1.8\sigma$.

References

Csernyák L 1992: *Acta Geod. Geoph. Mont. Hung.*, 27, 195–202.
Edwards R E 1965: Functional Analysis. Holt, Rinehart, Winston
Jensen J L W V 1906: *Acta Math.*, 30, 175.
Steiner F (ed.) 1991: The most frequent value. Akadémiai Kiadó, Budapest

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 207–218 (1992)

POLARISATION OF THE SOLAR WIND CONTROLLED PULSATIONS

T A Plyasova-Bakounina[1] and J W Münch[2]

[Manuscript received March 3, 1992]

An analysis of the polarisation of the solar wind controlled Pc_{sw} 2–4 during day time was carried out for the high latitude stations Ny Alesund and Bear Island. A change of the sign of the vector rotation of pulsations near local noon was not observed. It is concluded that Alfvén type waves from the solar wind are responsible for the Pc_{sw} 2–4 activity. A comparison of the waves in the solar wind (HEOS-1) and on the ground (Borok) has revealed no consistent change of the polarisation sign. A tentative model for upstream wave entry into the magnetosphere is discussed.

Keywords: geomagnetic pulsation; magnetosphere; polarisation; solar wind; upstream waves

A class of the solar wind controlled pulsations Pc_{sw} (the subscript sw indicates this influence) exists among geomagnetic pulsations (5–100 sec). The frequency of these oscillations is controlled by the magnitude of the interplanetary magnetic field (IMF) according to the experimentally found formula: $T = 160/B$, where T is the period of Pc2–4 in sec, B is the magnitude of IMF in nT (Troitskaya et al. 1971). Their intensity depends on the orientation of the IMF and a combination of some other parameters of the interplanetary medium (Bolshakova and Troitskaya 1968, Plyasova-Bakounina 1972, Greenstadt 1973, Webb and Orr 1976, Verő and Holló 1978, Plyasova-Bakounina et al. 1978, Golikov et al. 1980, Greenstadt and Olson 1976, 1979, Greenstadt et al. 1979). In contrast to pulsations Pc_{mg}2–4 which are of innermagnetospheric origin they exhibit a high-latitude (near cusp) maximum of the intensity along a meridian (Plyasova-Bakounina et al. 1986, Morris and Cole 1987). Similar results about the high latitude entry of Pc3 energy were obtained by Lanzerotti et al. (1986) and Engebretson et al. (1986).

Up to now the source of these waves is not completely discovered. Mainly two alternative mechanisms for the generation of these waves are proposed:

a) Waves in the solar wind (Gul'elmi 1974, Greenstadt 1973, Kovner et al. 1976, Varga 1980),

b) Waves at the magnetopause (Southwood 1968, Chen and Hasegawa 1974, Golikov et al. 1980, Pu and Kivelson 1983) and

c) a combination of both (Yumoto et al. 1984, 1985).

[1]Institute of Physics of the Earth, B. Gruzinskaya 10, 123810 Moscow, Russia
[2]University of Siegen, Siegen, FRG

One of the most simple opportunities to answer the question — which of the two mechanisms may be responsible for the generation of the solar wind controlled pulsations — would be an analysis of the geomagnetic wave fields recorded simultaneously by three satellites in the three different regions — solar wind, magnetosheath and magnetosphere. Greenstadt et al. (1983) analysed wave field data recorded by two satellites being positioned in the magnetosheath and in the magnetosphere. They have convincingly shown that waves pass from the magnetosheath to the magnetosphere. But this result does not allow to answer the question of the location of the source of these oscillations — at the magnetopause or in the solar wind? Engebretson et al. (1987) analysed the wave field data in the solar wind and in the magnetosphere but did not answer this question as well.

Observations recorded at three satellites within the magnetosphere arranged along a radius of the magnetosphere could also help to solve this problem. An analysis of the type of attenuation of the wave signal with distance would indicate indirectly the type of the source. One source — the Kelvin-Helmholtz instability (KHI) — generates a surface wave, while the other source — upstream waves — are considered as compressional waves reaching to, and passing through, the magnetopause. Both waves differ in type as well as in rate of attenuation with distance from the magnetopause. The surface wave attenuates exponentially with distance, the compressional wave does not attenuate at all if we consider it a plane wave and consider the medium without dissipation. Detailed analyses of this topic were conducted by Wolfe et al. (1989).

Is it possible to solve this problem without observing geomagnetic waves at satellite positions and use only ground-based observations? We believe that a positive answer to this question can be achieved by investigating the polarisation of geomagnetic pulsations.

Chen and Hasegawa (1974) considering effects of the KHI at the magnetopause have shown that the sign of rotation of the polarisation ellipse of the generated waves is different before and after noon. The resonance characteristics of the magnetosphere affect the spatial distribution of the polarisation in resonance regions (Southwood 1968). Therefore we will consider the change of the polarisation of the Pc3–4 pulsations at (high latitude) stations the resonance oscillations of which correspond to an other frequency band of micropulsations, to Pc5. Though the high harmonics except the fundamental one can be generated simultaneously at one and the same field tube, as it was demonstrated from satellite data, nevertheless, the most typical oscillations corresponding to the high latitudes are in the frequency range rather of Pc5 than of Pc4 or Pc3.

By studying the direction of rotation of the polarisation ellipse of the Pc3–4 pulsations at different local times, information can be obtained about the change (or lack of change) of the sign of rotation near local noon. The polarisation reversal (or lack of the reversal) near local noon would be indicative for (or against) a KHI source at the magnetopause. The local Alfvén velocity for upstream waves which propagate antisunward, is 1500 km/sec versus 200 km/sec for the KHI waves. Consequently the waves from upstream sources would have azimuthal wavelengths λ, which are an order of magnitude longer than those due to a KHI. Hence, due to the big λ for upstream waves, actually there will be no polarisation reversal near noon.

A great number of papers are devoted to the study of the polarisation of Pc3–5 pulsations. The most complete review of these papers was given by Ansari and Fraser (1985). Samson et al. (1971), and Olson and Rostoker (1978) showed that Pc4–5 wave vectors at high latitudes (70°) have left direction of rotation before local noon and right after noon. Such a clear situation is not observed for Pc3–4 pulsations at middle latitudes. In contrast to the paper of Lanzerotti et al. (1974) who showed that Pc3 wave vectors change sign of rotation near local noon, Green (1976) and Mier-Jedzzejowich and Southwood (1979, 1981) reported absence of the sign change near local noon. They concluded that the polarisation of Pc3 changes irregularly during daytime for low latitudes as it does for high latitudes (Zelwer and Morrison 1972, Lanzerotti et al. 1981). Additionally Lanzerotti et al. (1981) reported that before noon the polarisation of the Pc3 waves shows left rotation, but afternoon there is no clear regular pattern.

Thus, there is no simple picture of the change of the Pc3–4 polarisation near noon which corresponds to the resonance theory of generation of geomagnetic pulsations. The results of the studies seem to be contradictory.

In the present work we investigate the relationship between the polarisation of Pc3–4 at two high latitude stations and local time. Our work differs from the previous ones as we deliberately restrict the analysis to pulsations controlled by the solar wind $Pc_{sw}3$–4. This restriction to Pc_{sw} in the class of Pc3–4 helps to get rid of an influence which is introduced by pulsations of innermagnetospheric origin, and may lead more directly to the localization of the source of the solar wind controlled pulsations.

In addition we consider how the polarisation of the waves is changed passing from the solar wind region to the ground. Since a similar investigation has not been carried out before, even a rough estimation of this parameter is of scientific interest.

Data and analysis

Geomagnetic pulsation data from the stations Ny Alesund ($\Phi = 74°$, $\Lambda = 127°$) and Bear Island ($\Phi = 71°$, $\Lambda = 119°$) were available during IMS August-September 1978. The magnetometers and digital recording devices were described by Münch et al. (1975). The phase and amplitude response characteristics of the magnetometers are identical for both stations and for the H- and D-components. Chart records of the H- and D-components produced with a speed 15 mm/min were used for visual selection of $Pc_{sw}2$–4 pulsation activity. We have restricted our initial selection to events in which the field components at both stations oscillated in a quasi-regular way.

Furthermore the selection was restricted to events in the time range 0800–1500 UT for which IMF data were available and periods corresponded to the experimental formula $T = 160/B$ within the limit ± 10 %.

Table I lists investigated events of Pc3–4 pulsations before and after noon, their start and end time, averaged period T within a ± 10 % deviation limit, and T_{calc}

— the period calculated according to the quoted formula. The IMPJ data of the IMF-magnitude B (5 sec averaged plots) were available for the calculation of $T_{\text{calc}} = 160/B$.

Table I.

No.	Day	Time Interval Local Time	T	T_{calc}
1.	17.08.78	08:00–08:06	25	22
2.	17.08.78	08:08–08:12	25	22
3.	17.08.78	08:17–08:24	25	22
4.	17.08.78	08:28–08:34	25	22
5.	17.08.78	08:36–08:46	25	22
6.	17.08.78	08:58–09:06	25	22
7.	17.08.78	09:06–09:12	25	22
8.	07.08.78	09:06–09:12	42	41
9.	17.08.78	09:43–09:48	22	22
10.	17.08.78	09:55–10:00	22	22
11.	17.08.78	10:03–10:09	22	22
12.	21.08.78	10:36–10:41	37	41
13.	21.08.78	10:53–11:11	37	41

No.	Day	Time Interval Local Time	T	T_{calc}
1.	07.08.78	12:00–12:12	42	41
2.	07.08.78	12:53–13:00	42	41
3.	18.08.78	13:02–13:22	25	25
4.	27.08.78	13:00–13:08	17	19
5.	27.08.78	13:09–13:26	19	19
6.	12.08.78	14:12–14:18	18	18
7.	12.08.78	14:22–14:30	18	18
8.	18.08.78	14:46–14:56	25	25

Table II. The ratio of the number of the polarisation ellipses of both signs LH/RH

Station	Time	
	Before noon	After noon
NYA	1.1 $\left(\frac{20}{17}\right)$	0.9 $\left(\frac{13}{14}\right)$
BI	1.3 $\left(\frac{17}{13}\right)$	1.6 $\left(\frac{10}{6}\right)$

Figure 1 shows an example of data recorded at Ny Alesund together with the polarisation ellipses for each cycle. These data are typical of those studied during August-September 1978.

Figure 2 gives an overview of the distribution of the polarisation ellipses during

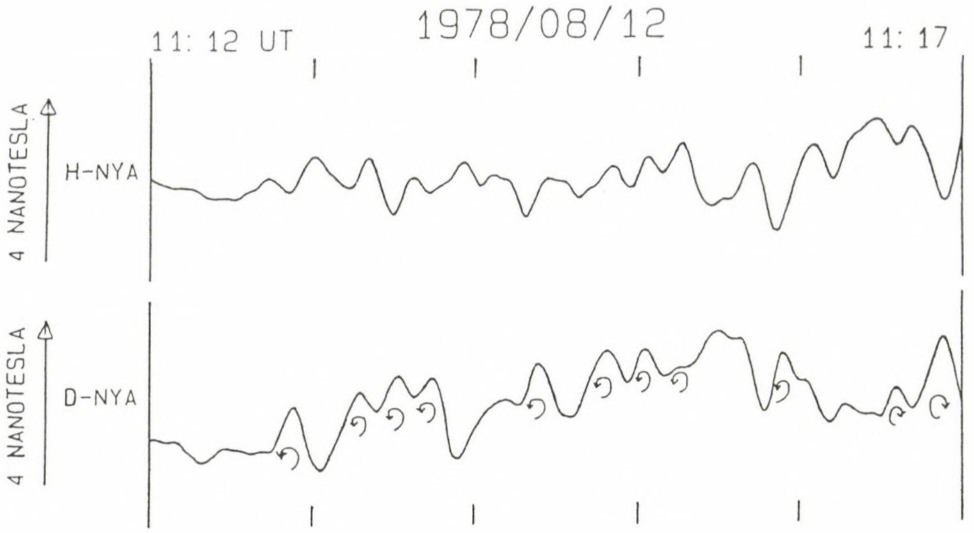


Fig. 1. Pulsations recorded at Ny Alesund

daytime for all events. The direction of the rotation of the ellipses varies irregularly. Table II shows the number of ellipses with left (LH) and right (RH) hand direction of the rotation (in brackets LH/RH) and their ratio. One can see from the table that the ratio of the number of polarisation ellipses of each sign are approximately equal before and after noon.

We have used magnetic wavefield data of HEOS-1 and ground-based pulsation data of the station Borok ($\Phi = 53°$, $\Lambda = 123°$) to investigate the change of the polarisation of the waves.

The IMF measurements were obtained from the memory mode of the HEOS-1 magnetometer which provided intermittent coverage for periods of $\sim$ 18 min duration, at a sampling rate of one vector every 1.5 sec. The frequency response of the instrument was flat at frequencies less than $\sim$ 1 Hz, and the resolution of the field component measurements was $\pm$ 0.25 nT (Hedgecock 1975).

The ground station data from the Borok observatory were available as chart records produced with a scale 30 mm/min from an induction magnetometer. The instrument has a flat frequency response throughout the period range of interest. The amplitudes have been corrected for the instrumental response.

Pulsations with periods in the range 10–100 sec are frequently observed upstream from the bow shock (Fairfield 1969, Russell et al. 1971). During the period December 1968 – April 1969 HEOS-1 was in the sunward hemisphere, upstream from the bow shock, at a magnetospheric latitude in the range $20° - 50°$ for a substantial part of each orbit. A considerable variety of pulsation activity was noticed

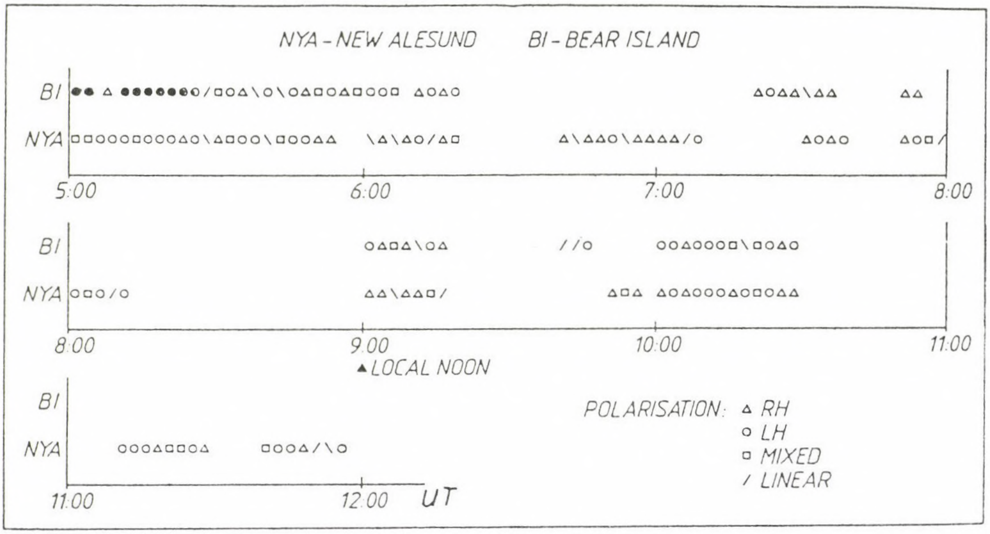


Fig. 2. Distribution of the polarisation ellipses during daytime for Ny Alesund and Bear Island

but we have restricted the initial selection to events in which the field components at the satellite oscillated in a quasi-regular way with an amplitude of several nT. Restricting the analysis furthermore to events which occurred in the time range 0300–1000 UT — most suitable for observing Pc3–4 at Borok —, we obtained 13 events. Simultaneous pulsation activity was absent at Borok during four of these events; among the remaining nine selected for detailed study we restricted the analysis to oscillations which started abruptly at some point in the 18 min record "burst of pulsations". These measurements provided an estimate of the dominant period and an indication of the time delay between the onsets of pulsation activity at the satellite and at Borok. Dynamic spectra covering the period range 6–60 sec were calculated for the remaining events. The individual spectral estimates from this analysis overlapped, with a bandwidth such that the -3 dB (0.707 amplitude) points on the spectral response curve of each estimate coincided with the centre frequencies of the two neighbouring estimates. The generation of spurious responses was avoided by removing linear trends from the initial data sequenced and by using a time window with a cosine profile (e.g. Bendat and Piersol 1971). The frequency and time windows, Δf and Δt resulting from any dynamic spectral analysis are related such that $\Delta f\ \Delta t = 1$. We have compared spectra with frequency resolutions $\Delta f/\Delta t$ of 0.2 and 0.5 corresponding to time resolutions $\Delta t/t$ of 5 and 2, to verify that the observed features at not artefacts of the mathematical procedure. Thus, the calculated spectra represent the real dynamic features of relatively narrow band signals.

A dynamic spectrum for one of the selected events is displayed in Fig. 3. The contours in the dynamic spectra correspond to the amplitudes 1, 0.63, 0.39, while the empty region represents amplitudes < 0.24. The closely similar dominant periods

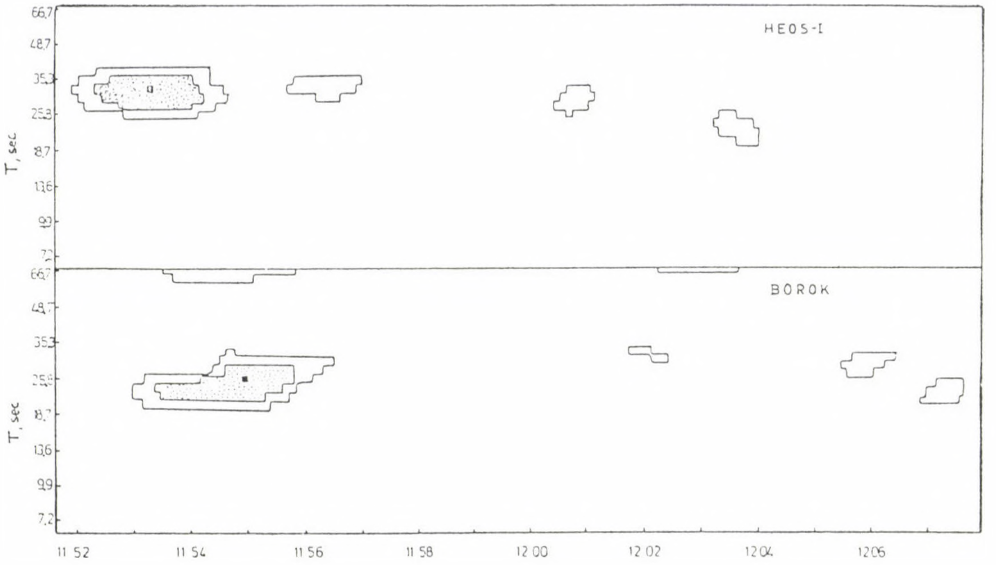


Fig. 3. Dynamic spectra of the pulsation activity at HEOS-1 and at Borok during the event of 23 December 1968. The contours represent normalised amplitudes of 1, 0.63, 0.39

at the satellite and at the ground station, illustrated in these figures, confirm the results of an earlier analysis (Plysova-Bakounina 1972).

The parameters $T_{\rm HEOS}$, $T_{\rm BOR}$, τ are deduced by inspection of these dynamic spectra. They are listed in Table III: T is the period corresponding to the dominant frequency component in the signal; $T_{\rm calc}$ is the period calculated according to the formula $T = 160/B$, where B is the average magnitude of IMF, τ the time delay HEOS-Borok.

Table III.

No.	Date	Time interval	$T_{\rm BOR}$ sec	$T_{\rm HEOS}$ sec	$T_{\rm calc}$ sec	τ min	$K_{\rm p}$
1.	23.12.68	11:51–12:07	26	30	27	2	3–
2.	18.12.68	10:02–10:20	27	30	32	1	1+
3.	02.04.69	06:00–06:15	16	16	17	6	2–

Figure 4 shows ellipses of polarisation for three bursts of the solar wind controlled pulsations at the satellite and on the ground. Polarisation hodograms were plotted, taking into account the time-delay τ of signals. Hodograms were calculated for the first 450 pairs of points, what corresponds approximately to the first 10 cycles of

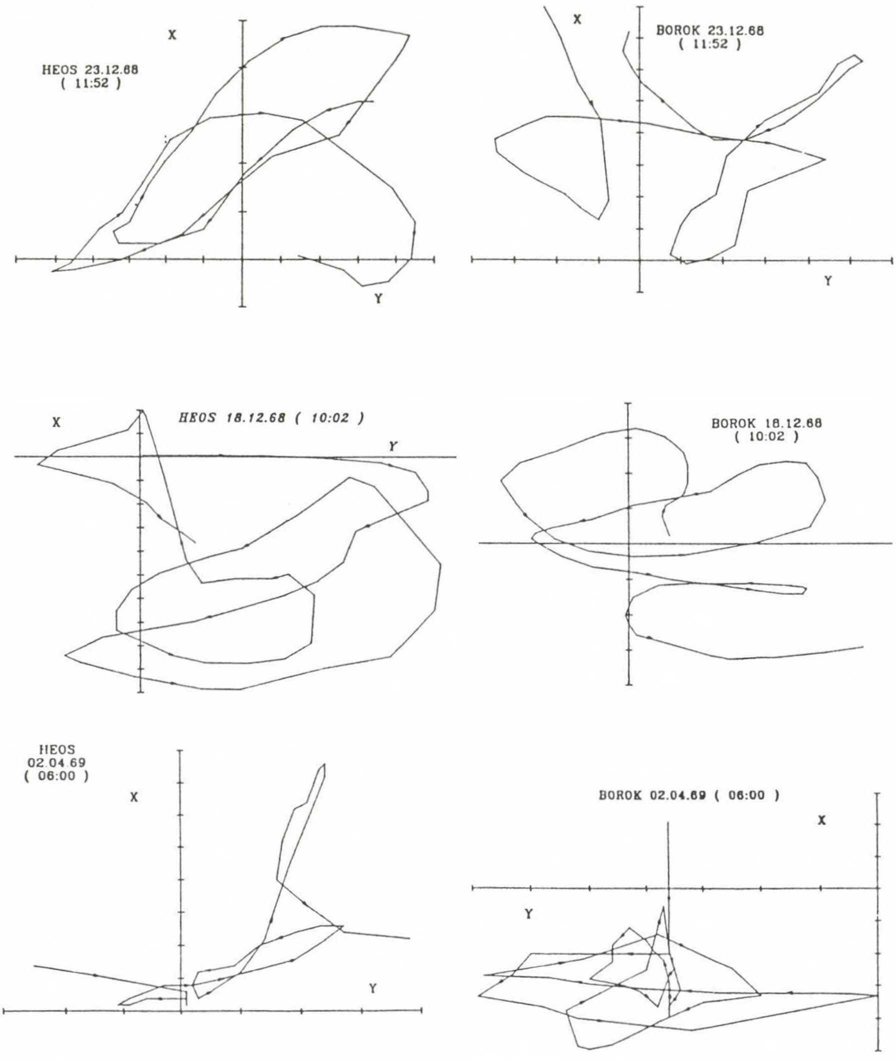


Fig. 4. Ellipses of polarisation for three bursts of pulsations in the solar wind (HEOS-1) and on the ground (Borok)

the oscillation, a few of which are displayed in the figure. The direction of the rotation of the waves changes irregularly both in the solar wind and on the ground. A correlation between the polarisation of the signal at the moment of the waves' onset on the solar wind and the polarisation at the moment of arrival on the ground surface is not observed. The time-delay for events between events in space and on the ground depends on the distance between the satellite position and the bow shock (see Plyasova-Bakounina et al. 1978).

Discussion

The equal ratio of the number of polarisation ellipses with opposite signs of rotation both in the morning and afternoon support the idea of mechanisms of Pc_{sw} generation, which are not related to the Kelvin-Helmholtz instability at the magnetopause. In this case an alternative mechanism is the beam instability of reflected protons in the solar wind.

This inference is not trivial. Indeed, the question of the separation of the waves from two sources — waves in the solar wind and surface waves at the magnetopause (instability of tangential discontinuity) is rather complicated itself. The difficulties of distinguishing pulsations from two sources are the following ones:

First, according to the linear theory, both mechanisms produce spectra of the waves approximately in the same frequency range; second, nonlinear mechanisms which work both before the bow shock and at the magnetopause have rather weak influcence upon the frequency which corresponds to the maximum of the spectrum of the excited wave, the position of which is defined in fact by linear mechanisms; and finally, the orientation of the IMF (decrease of the azimuthal angle) affects approximately similarly the development of both instabilities (kinetic one in the solar wind and hydromagnetic one at the magnetopause; Golikov et al. 1980).

The experimental result that the direction of rotation of the polarisation ellipse of $Pc_{sw}3$ varies randomly with local time agrees with the result, obtained earlier (Green 1976, Mier-Jedrzejowich and Southwood 1979, 1981) for the whole class of Pc3 pulsations at middle latitudes. This result, according to our opinion, is a consequence of the fact that pulsations, which were chosen for the analysis, were accidentally solar wind controlled pulsations. It is not surprising at all, because, as shown by Verő (1986), the majority of pulsation events (80 percent) are solar wind controlled pulsations for middle latitudes.

Let us consider the two results:

1. Pc_{sw} have their maximum intensity at high latitudes as shown by Plyasova-Bakounina et al. (1986),

2. Pc_{sw} are generated by upstream waves (present study).

From both facts we conclude that the upstream waves penetrate into the magnetosphere and are guided by magnetic fields lines, e.g. Alfvén type, as it was originally shown by Fairfield (1969). For the explanation of the unhindered penetration of upstream waves through the magnetopause and subsequent propagation to the low latitudes some authors (Wolfe et al. 1985, Yumoto et al. 1984, 1985) consider the penetrating upstream waves as magnetosonic ones (see e.g. Fig. 8 in Yumoto et al. 1985). Though both types of MHD-waves are observed in the solar wind, generation of magnetosonic ones is more effective than of Alfvén ones

(Gul'elmi 1979, Hoppe and Russell 1983). But, according to this model the maximum intensity would occur at the equator, what is not confirmed experimentally. The absence of the equatorial maximum could be explained (Kwok and Lee 1984) by the mechanism of the transformation of the magnetosonic waves into Alfvén ones, but for its realisation specific conditions of the reconnection are required; they are not always fulfilled. Therefore this question should be investigated separately.

The assumption of Yumoto et al. (1984) that Alfvén mode Pc3 pulsations only exist at high latitudes is contradictory to the findings of Plyasova-Bakounina et al. (1986) who have demonstrated that Pc_{sw} are observed simultaneously from high to middle latitudes. This means transmission of energy of waves across L-shells.

What could the mechanism of transmission be? We speculate that Alfvén waves generated by the upstream source modulating precipitation of electrons is a cause of the modulation of the brightness of the day auroras. The effect of the Pc3 period range pulsations on the day auroral brightness in described by Chernous et al. (1986). It seems that the coincidence of the spatial positions and the maxima of number of occurrence both of day auroral pulsations and $Pc_{sw}2$–4 pulsations is not accidental. The particle precipitation, modulating the conductivity of the ionosphere, provokes the appearance of a current system which is closed through middle to low latitudes. Due to the oscillations of such a current system the middle to low latitude solar wind controlled pulsations are generated. The close idea about the high latitude wave entry is suggested by Engebretson et al. (1987). We think that the Alfvén mode of upstream waves play a more important role than previously supposed.

A comparison of the waves in the solar wind and on the ground at the midlatitude station has shown that pulsations, both in the solar wind and on the ground, do not change sign of the polarisation ellipse in a consistent manner. A relationship between the polarisation of the signal on the ground and the polarisation of the signal in the solar wind just after its generation could not be established. Being aware of the fact that resonance characteristics of the magnetosphere may change the polarisation of the signal during propagation from high to middle latitudes, we believe that the change of polarisation of the propagating signal is the same for each of the selected events, since the position of the footprint of the fieldline of the station Borok relative to plasmapause did not change considerably due to small K_p-values.

Conclusion

Trying to find out the type of source which is responsible for the Pc_{sw} generation, we have indicated that the problem may be solved even without a special array of satellites which are simultaneously measuring the wave fields in three different regions: solar wind, magnetosheath and magnetosphere.

Supposing Pc3–4 polarisation is the principal characteristic of the definition of the mechanism of the pulsation source, we have investigated the change of Pc3–4 polarisation with local time. In order to exclude an influence upon the polarisation by resonance effects of the magnetosphere and to avoid confusion with the pulsations of innermagnetospheric origin, we restricted our investigations of the Pc activity to high latitude stations where we selected only those pulsations the period of which agrees with $T = 160/B$. An analysis of the relationship between the polarisation ellipses of $\mathrm{Pc_{sw}}3$–4 at high latitudes and local time has shown that a change of the rotation sign is not observed near local noon. It obviously indicates that the generation of the solar wind controlled pulsations is not related to the Kelvin-Helmholtz instability at the magnetopause. The source of $\mathrm{Pc_{sw}}$ are upstream waves in the solar wind, generated by two-beam instability of protons, reflected from the bow shock, as was suggested by Gul'elmi (1974). The comparison of the waves recorded simultaneously in the solar wind and on the ground taking into account the time delay due to propagation have not revealed a change of the polarisation sign in a consistent manner.

By stating that the source of solar wind controlled Pc are upstream waves and the position of the high latitude maximum of $\mathrm{Pc_{sw}}$ intensity is near the cusp, we infer that the mode of hydromagnetic waves, transmitting energy from the solar wind into magnetosphere is Alfvén type.

We speculate that the transmission of energy from the high to the middle and low latitudes is realized by the creation of a corresponding current system.

Acknowledgements

We thank Goddard Space Flight Centre for providing the IMPJ-data. We are grateful to Dr P C Hedgecock for providing of HEOS-1 data.

References

Ansari I A, Fraser B J 1985: *Indian J. Phys.* 59B, 151–178.

Bendat J S, Piersol A G 1971: Random Data. Wiley Interscience, New York

Bolshakova O V, Troitskaya V A 1968: *Dokl. AN SSSR*, 180, 343–346.

Chen L, Hasegawa A 1974: *J. Geoph. Res.*, 79, 1024.

Chernous S A, Vorobjow V G, Tagirov V R, Rothwell P, Grivs C, Lanchester B 1986: TV and photometrical observations of the region between the discrete arcs and pulsations in day auroras. Proceedings of International Symposium "Polar Geomagnetic Phenomena", Suzdal, USSR

Engebretson M J, Meng C I, Arnoldy R L, Cahill L J Jr. 1986: *J. Geoph. Res.*, 90, 8909.

Engebretson M J, Zanetti L J, Potemra R A, Baumjohann W, Lühr H, Acuna M H 1987: *J. Geoph. Res.*, 92, A9, 10053–10062.

Fairfield D M 1969: *J. Geoph. Res.*, 74, 3541–3553.

Golikov Yu V, Plyasova-Bakounina T A, Troitskaya V A, Chernikov A A, Pustovalov V V, Hedgecock P C 1980: *Planet. Space Sci.*, 28, 525.

Green C A 1976: *Planet. Space Sci.*, 24, 79.

Greenstadt E W 1973: Field-determined oscillations in the magnetosheath as possible source of medium-period, daytime micropulsations. Proceedings of the Conference on Solar Terrestrial Relations. D Venkatesan ed. University of Calgary, Calgary, Alta, p. 515.

Greenstadt E W, Olson J V 1976: *J. Geoph. Res.*, 81, 5911.

Greenstadt E W, Singer H J, Russell C T, Olson J V 1979: *J. Geoph. Res.*, 84, 527.

Greenstadt E W, Olson J V 1979: *J. Geoph. Res.*, 84, 1493.

Greenstadt E W, Mellot M M, McPherron R L, Russell C T, Singer H J, Knecht D J 1983: *Geophys. Res. Lett.*, 10, 8, 659–662.

Gul'elmi A V 1974: *Space Sci. Rev.*, 16, 331.

Gul'elmi A V 1979: MHD-waves in the environment plasma (in Russian). Nauka

Hedgecock P C 1975: *Space Sci. Instrum.*, 1, 61.
Hoppe M M, Russell C T 1983: *J. Geoph. Res.*, 88, 2021–2028.
Kovner M S, Lebedev V V, Plyasova-Bakounina T A, Troitskaya V A 1976: *Planet. Space Sci.*, 24, 267.
Kwok Y C, Lee L C 1984: *J. Geoph. Res.*, 89, 10697.
Lanzerotti L J, Fukunishi H, Chen L 1974: *J. Geoph. Res.*, 78, 4648.
Lanzerotti L J, Medford L F, MacLennan C G, Hasegawa T 1981: *J. Geoph. Res.*, 86, 5500.
Lanzerotti L J, MacLennan C G, Medford L V, Carpenter D L 1986: *J. Geoph. Res.*, 91, A1, 375.
Mier-Jedrzejowich W A C, Southwood D J 1979: *Planet. Space Sci.*, 27, 617.
Mier-Jedrzejowich W A C, Southwood D J 1981: *J. Atm. Terr. Phys.*, 43, 911.
Morris R J, Cole K D 1987: *Planet. Space Sci.*, 35, 1437.
Münch J W, Wilhelm K, Troitskaya V A, Kasak B N, Baranskiy L N 1975: Observations of geomagnetic micropulsations on a mid-latitude East-West-profile. Kleinheubacher Berichte, Bd. 18, 123.
Olson J V, Rostoker G 1978: *J. Geoph. Res.*, 83, 2841.
Plyasova-Bakounina T A 1972: *Geomagnetism i Aeronomija*, 4, 12.
Plyasova-Bakounina T A, Troitskaya V A, Golikov Y V, Hedgecock P C 1978: *Planet. Space Sci.*, 26, 547–553.
Plyasova-Bakounina T A, Troiskaya V A, Münch J W, Gauler H F 1986: *Acta Geod. Geoph. Month. Hung.*, 21, 143–153.
Pu Z-Y, Kivelson M G 1983: *J. Geoph. Res.*, 88, 841.
Russell C T, Childers D D, Coleman P J Jr. 1971: *J. Geoph. Res.*, 76, 845.
Samson J C, Jacobs J A, Rostoker G 1971: *J. Geoph. Res.*, 6, 3675.
Southwood D 1968: *Planet. Space Sci.*, 16, 587.
Troitskaya V A, Plyasova-Bakounina T A, Gul'elmi A V 1971: Relationship between Pc2–4 pulsations and the interplanetary magnetic field (in Russian). Dokl. AN SSSR, 197, 1312.
Varga M 1980: *J. Atm. Terr. Phys.*, 42, 365–369.
Verő J 1986: *Journal of Geophysics*, 60, 106–119.
Verő J, Holló L 1978: *J. Atm. Terr. Phys.*, 40, 857–865.
Webb D, Orr D 1976: *J. Geoph. Res.*, 81, 5941.
Wolfe A, Meloni A, Lanzerotti L J, MacLennan C G, Bamber J, Venkatesan D 1985: *J. Geoph. Res.*, 90, 5117.
Wolfe A, Uberoi C, Russell C T, Lanzerotti L J, MacLennan C G, Medford L V 1989: Penetration of hydromagnetic energy deep into the magnetosphere. Abstracts of IAGA meeting
Yumoto K, Saito T, Tsurutani B T, Smith E J, Akasofu S-I 1984: *J. Geoph. Res.*, 89, 9731–9740.
Yumoto K, Saito T, Akasofu S-I, Tsurutani B T, Smith E J 1985: *J. Geoph. Res.*, 90, 6439–6450.
Zelwer R, Morrison H F 1972: *J. Geoph. Res.*, 77, 674.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 219–229 (1992)

COMPARISON OF MAGNETOTELLURIC MODELLING RESULTS WITH AN ELLIPTIC CYLINDER MODEL

E LANNE[1]

[Manuscript received April 14, 1992]

In magnetotelluric soundings, superficial resistivity variations may magnify or diminish apparent resistivities even by decades. These shifts are difficult to correct by forward modelling. Berdichevsky and Dmitriev (1976) have presented relationships for the effects of a vertical cylinder which is located in the first layer of a K-type layer structure and which in cross-section is an ellipse. The relationships are based on the so-called long wavelength approximation (LWA). In the interpretation of MT soundings, the relationships were helpful if the real structures correspond to the model adequately. In this work, some published modelling results are compared with the cylinder model. If the LWA condition holds, the equivalent cylinder model can be found for fairly complicated structures. The parametres (*i.e.* resistivity contrasts and geometry) correspond well with the ones of the original model. If the LWA condition fails, the distortions are relevant but are usually significantly weaker.

Keywords: inverse problem; magnetotelluric methods; static shift; three-dimensional models

Introduction

Magnetotelluric sounding curves measured in different directions usually deviate from each other. The reason for this are variations in lateral resistivity. The effects caused by these variations depend on the measurement directions. Furthermore, the superficial bodies shift the resistivity curves, so that the view of large scale structures may become strongly distorted. The 3-D interpretations are crude because even forward modelling methods are still laborious. One way to facilitate the interpretation is to use approximative solutions which may be adjusted by more exact methods.

Berdichevsky and Dmitriev (1976) have examined a model in which a vertical cylinder is situated in the first layer of a K-type layer structure. (A short review on the model is presented in the Appendix. However, it is assumed, that the reader is acquainted with the original paper). They have presented simple formulae for the distortions of the K-type sounding curves. The model is $2\frac{1}{2}$-dimensional (*i.e.* it is limited in the y-direction but it is symmetric with respect to the xz-plane). These formulae offer an opportunity to solve and correct the shifts of sounding curves (Lanne 1987, 1990).

However, the ideal K-type structures are rare in practice. Therefore, it is useful to compare the model with observations taken from different types of known

[1]Geological Survey of Finland, Regional Office, P.O.Box 77, SF-96101 Rovaniemi, Finland

structures. In practice, the only acceptable test materials are numerical modelling results. Park et al. (1983) and Wannamaker et al. (1984) have published 3-D results which resemble K-type structures. The model of Ting and Hohmann (1981) consists only of a conductive prism which lies in the half-space. It represents a case which differs considerably from the K-type structure. Another diverse model is the vertical cylinder which continues downwards and has a square cross-section (Reddy et al. 1977). In the next two examples, the resistivity decreases downwards from the surface. Sternberg et al. (1988) have presented distorted curves of H-type while the Q-type layer models of Zhang (1989) are covered by a thin sheet which include conductive strips.

Method

From the published sounding curves (Park et al. 1983 and Wannamaker et al. 1984), the apparent resistivity values were sampled, and they were scaled to the AMT-domain for the inversion program. Though the models include more than three layers, the equivalent three-layer inversions were easy to determine. The accuracy of calculated apparent resistivities was usually better than the line width in drawings. From the layer models got by inversion, the S- and z-distortions were calculated. In addition, based on distortions, the parametres of the cylinder model were determined. Though only the relative parameters (say, a ratio of axes $r_{a/b} = a/b$) can be determined, absolute parameter values are presented for the sake of clarity. For example, the axis a has been fixed constant while b is based on the formula $b = r^*_{a/b}a$. In the original model, the material parameter is the ratio of S_1-values. Supposing that the cylinder penetrates the first layer, the S_1-ratio simplifies to the conductivity (or resistivity) contrast. In the treated modelling results, the resistivity contrasts were calculated down to the insulating layer.

The papers of Ting and Hohmann (1981) and Reddy et al. (1977) represent apparent resistivities only at constant frequencies. In these cases, the distortions of resistivities were calculated, and then they were interpretated either to be S- or z-distortions. From the curves of Sternberg et al. (1988), only the descending parts were utilized. The distortions were determined from the ratios of apparent resistivities. Zhang (1989) represents the distortions as diagrams in which the magnitudes and principal directions are expressed by vectors. I have interpretated the vectors to L- or T-directions and — distortions respectively.

Equivalent cylinder models were determined by a computer program in which the model parameters can be optimized when the distortion coefficients have been given.

Comparison of models

Model: Park et al. (1983)

This model consists of 6 layers and includes an L-shaped body in the surface (Fig. 1). The resistivity of the insulating intermediate layer is 25 times larger than the surface. Thus the layer structure resembles K-type. The z-distortions presume

S_0/S_i-values of 0.10–0.22 which coincide with a theoretical value of 0.10 (Fig. 2). Though the shape of the body differs significantly from the ellipse, the elongations of ellipses characterize the form of the body.

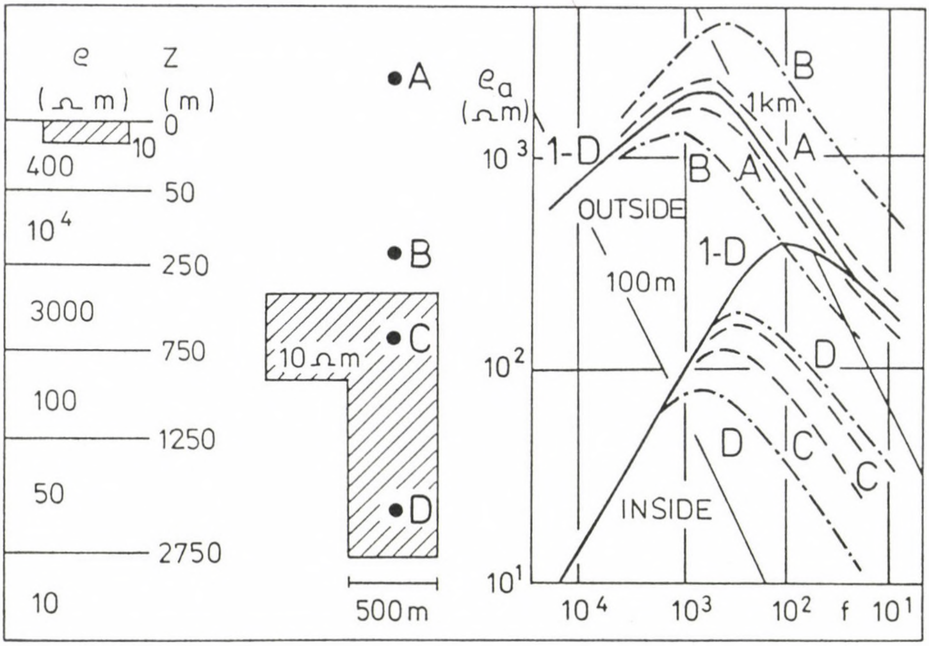


Fig. 1. Model of Park et al. (1983) and modelled sounding curves (scaled to AMT domain)

Model: Wannamaker et al. (1984)

Figure 3 presents the model and the sounding curves at points A–D. The frequency domain of the original curves is 0.001–100 Hz. The intermediate layer is 10 times more resistive than the surface. The conductive body is in the first layer, but it is not just in the surface. The conductivity constrast (or merely the S-contrast) needed for the cylinder model is about one hundred. The LWA does not hold at high frequencies. For example, a wave length corresponding to the thickness of the body takes place at 5 Hz and the width of the body occurs at 0.05 Hz. S-distortions particularly occur at frequencies higher than 0.1 Hz. Because points A–C are situated on the axes of symmetry, the correspondence to the cylinder model is satisfactory. However at D, near to the corner of the sheet, the correspondence is more ambiguous.

Table I presents S_1- and z_2-values gotten from layer model inversions and corresponding distortions. Further, it contains parameters of such elliptic cylinders which cause equivalent distortions to z_2. The resistivity contrast, the ratio of axes, and the distance have been adjusted as near as possible to the quantities of the original body.

The S-distortions of T-soundings were comparable to the effects of a quite similar cylinder model, but the distortions in L-soundings are altogether too weak when compared with the effects of a comparable cylinder model. The cylinder models which explain S-distortions in both directions simultaneously differ so much from the original model that it is meaningless to present them.

Table I. Layer model inversions and distortions of model parameters. Parameters of an elliptic cylinder corresponding to z-distortions. ρ_0 is set to 400 Ωm. The values in brackets are corresponding parameters of the sheet model

Station	A		B		C		D	
Model S_1, z_2								
$S_{1,1-D}$	463		4.81		4.81		4.81	
S_{xy}, S_{yx}	–	–	4.86	4.13	5.24	4.18	4.92	3.95
$K_{S,L}$, $K_{S,T}$	–	–	1.01	0.859	1.09	0.869	1.02	0.821
$Z_{2,1-D}$	30.7		33.4		33.4		33.4	
Z_{xy}, Z_{yx}	3.8	2.2	8.3	45.3	12.0	70.9	9.0	97.6
$K_{z,L}$, $K_{z,T}$	0.123	0.072	0.248	1.356	0.359	2.123	0.269	2.922
Cylinder								
ρ_i	19.0	(4)	15.3	(4)	5.4	(4)	4.0	(4)
a	18.0	(18)	18.0	(18)	10.7	(7)	7.7	(-)
b	10.0	(7)	8.0	(7)	19.0	(18)	21.0	(-)
x, y	0	(0)	10.0	(10)	21.0	(21)	21.8	(-)

Model: Ting and Hohmann (1981)

The model consists only of a conductive square prism which is situated in the half-space. Its dimensions are 1000 m × 2000 m × 2000 m. The upper side is 250 m from the surface. The resistivities of the body and half-space are 5 Ωm and 100 Ωm respectively. The apparent resistivities have been calculated for frequencies of 0.1 Hz (Fig. 4) and 10 Hz. On the prism, a reference resistivity of 7.5 Ωm is used. This corresponds to the 1-D layer model with the resistivity at 0.1 Hz. It cannot be concluded directly from the model whether the distortion is either of the S- of z-type.

On the higher frequency, the LWA doesn't hold because the wave length in the body is about 1.1 km. Outside the body, the distortions are weaker than the LWA implies. On the body, both apparent resistivity values are too large if they are compared with the value of the 1-D model. Therefore, the results associated with the frequency of 10 Hz will not be discussed further.

Table II presents the calculated apparent resistivity values ρ_L and ρ_T with their distortions $r_L = (\rho_L/\rho)^{1/2}$ and $r_T = (\rho_T/\rho)^{1/2}$. By using the square roots, the distortions should correspond either to the S- or z-distortions. Table II presents

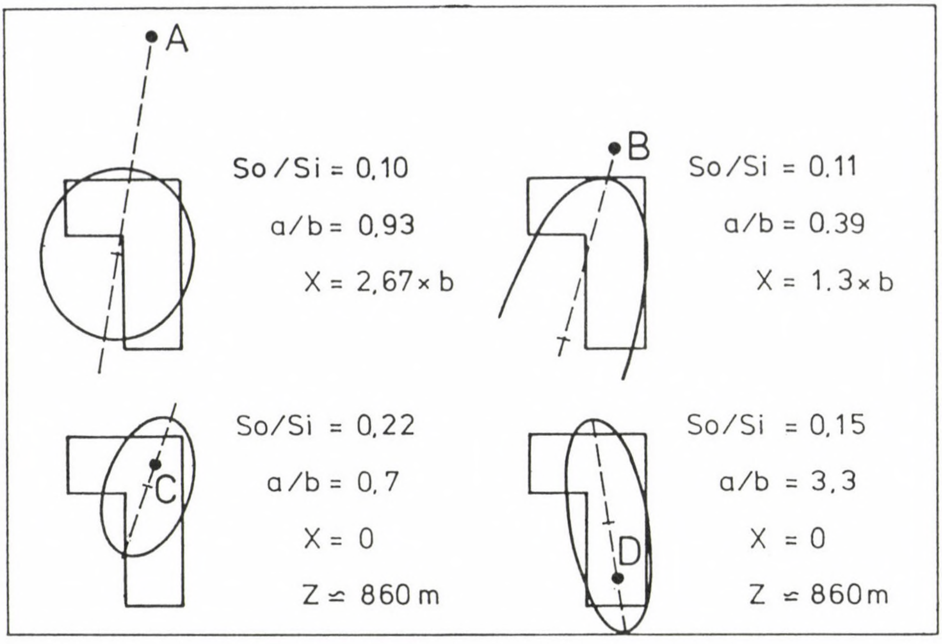


Fig. 2. Elliptic cylinders which cause equivalent effects as the L-shaped body (Lanne 1987)

examples of such cylinder parameters which explain the distortion to z-type. As in the previous model, the resistivity contrasts, the ratios of axes, and the relative distances have been adjusted to match the parameters of the prism. On the body, the resistivity contrast and ratio of axes are unique. Outside the body, the distance is an extra parameter which allows several solutions. If the distortion is presumed to be of the S-type, the cylinder deviated more from the prism model (*e.g.* at $x = 1000$, $y = 0$ ρ_i should be 2.9 Ωm [correct value 7.5 Ωm] and $a = 1100$ m and $b = 684$ m [1000 m and 500 m]).

As an example of more constrained modelling, corrected resistivities ρ_{corr} have been presented in Table II. The ratio of axes and the area of the ellipse were set equivalent with the prism. The only determined parameter was the S_0/S_i-contrast. The geometric averages of apparent resistivities were corrected by appropriate coefficients.

Model: Reddy et al. (1977)

The model consists of a vertical prism, which has a square cross-section and which is unlimited downwards. The resistivity contrast is 10, the frequency 0.1 Hz, and the side length is 16 km. Thus the LWA is not fulfilled. Surely the tendency of these distortions is reasonable outside the body, but they are altogether so small that they could not be interpreted by realistic cylinder models. Inside on the body, the apparent resistivities vary around the resistivity of the body, though according

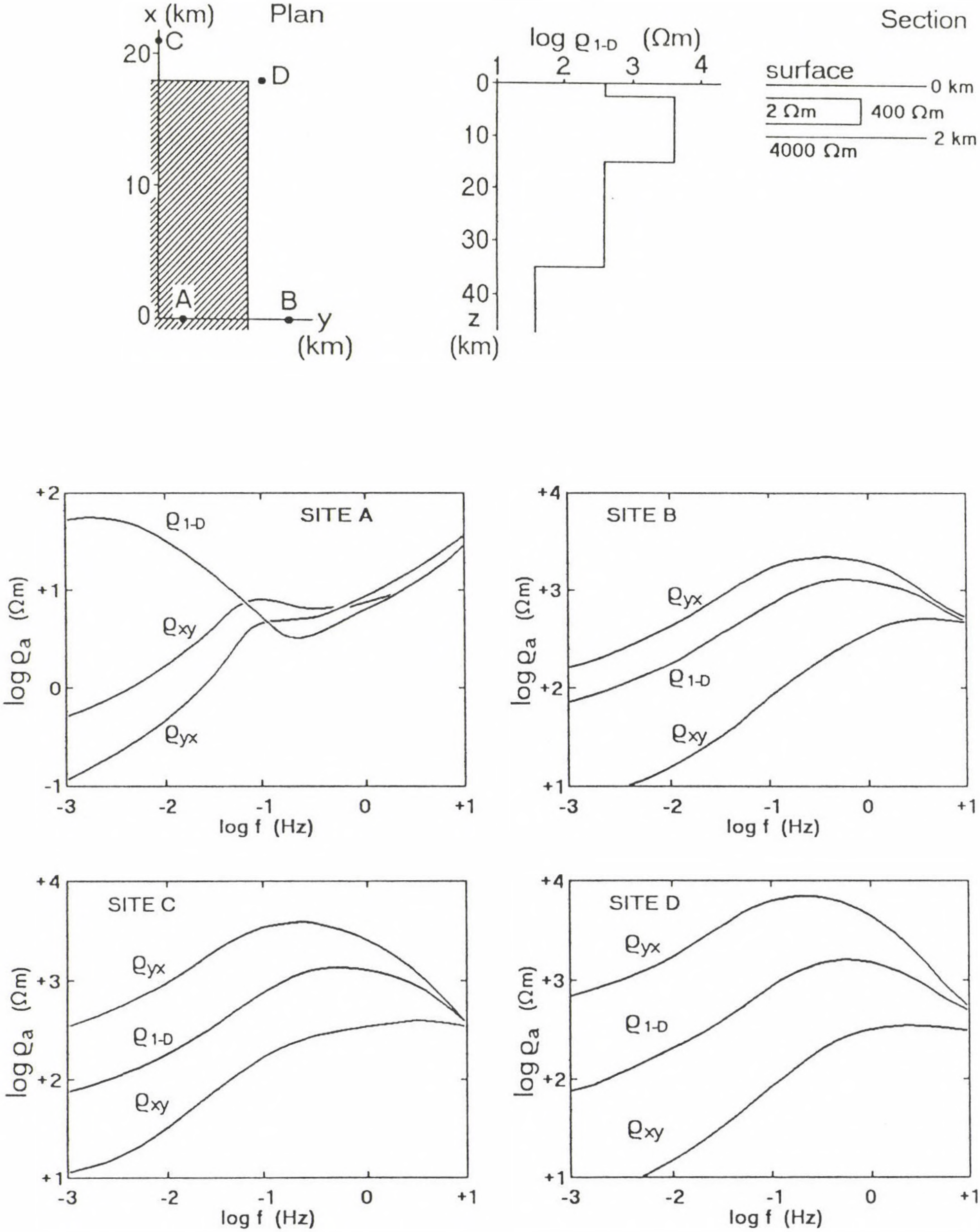


Fig. 3. Model of Wannamaker et al. (1984) and modelled sounding curves. Note: the curves are at directions of the tipper strike

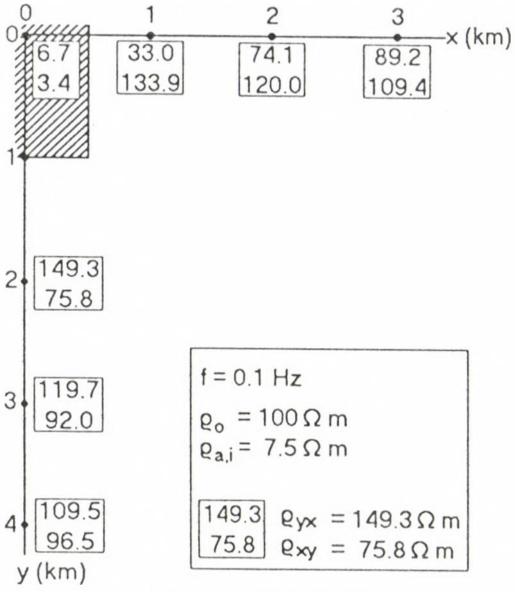


Fig. 4. Model of Ting and Hohmann (1981) and apparent resistivities used

Table II. Apparent resistivities and their interpretation to z-distortions. The parameters of elliptic cylinders and corrected apparent resistivities (see text). The corresponding parameters of the sheet model: at points 1–4, a = 1000 m, b = 500 m, at points 5–7 a = 500 m, b = 1000 m. s = distance (m)

Point	1	2	3	4	5	6	7
Model							
x	0	1000	2000	3000	0	0	0
y	0	0	0	0	2000	3000	4000
ρ_L	6.7	33.0	74.1	89.2	75.8	92.0	96.5
ρ_T	3.4	133.9	120.0	109.4	149.3	119.7	109.5
r_L	0.945	0.574	0.861	0.944	0.871	0.959	0.982
r_Z	0.673	1.157	1.095	1.046	1.221	1.094	1.046
Cylinder							
ρ_i	65	8.6	29	43	7.5	7.5	7.5
a	4200	1500	1500	1500	581	448	386
b	500	437	717	851	1091	1197	1170
s	0	1000	2000	3000	2000	3000	4000
ρ_{corr}	18.9	88.4	101.9	101.3	99.5	101.5	101.2

to the elliptic cylinder model, both apparent resistivities should be "too small".

Model: Sternberg et al. (1988)

The 1-D model consists of a conductive intermediate layer (Fig. 5). The distortive body is a conductive slab (40 m × 40 m × 4 m) in the surface. In the sense of the 1-D model, it has no practical influence on the presented frequency domain (0.001–100 Hz). The modelled sounding curves have shifted up and down depending on the directions and calculation sites. Certainly the shifts have proper trends, but the determined parameter values of the cylinder model deviate quite remarkably from those of the slab model (Table III). The distortions proved merely to be of the z-type rather than the S-type.

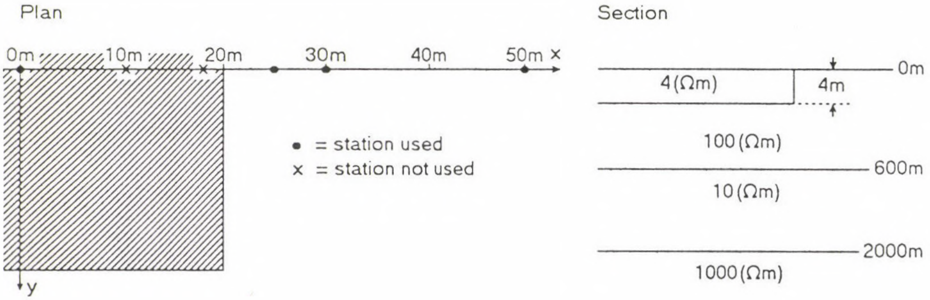


Fig. 5. Model of Sternberg et al. (1988) and sites used

Table III. Distortions of apparent resistivity curves. Parameters of an elliptic cylinder corresponding to z-distortions. ρ_0 is set to 100 Ωm. The values in brackets are corresponding parameters of the slab model

Station	0		25		30		50	
$K_{z,L}$, $K_{z,T}$	0.354	0.354	0.58	1.79	0.72	1.41	0.91	1.07
Cylinder								
ρ_i	21.5	(5)	4.0	(5)	8.0	(5)	5.0	(5)
a	20.0	(20)	10.1	(20)	12.8	(20)	26.0	(20)
b	20.0	(20)	20.0	(20)	20.1	(20)	20.0	(20)
x	0	(0)	24.0	(25)	30.0	(30)	76.0	(50)

Model: Zhang (1989)

Zhang's thin sheet models include one or several bodies in the surface. The thin sheet model is somewhat ambiguous in the sense of curve types because one can buy or sell the resistivity of the sheet by the cost of layer thickness. The discussion on the LWA condition is senseless as well because the resistivity is zero in principle. Figure 6 represents one of his models. On the figures, the parameters of equivalent cylinder models have been added. The distortions are of the proper kind, but the corresponding cylinder models differ remarkably from the strip model. In the majority of other cases, the directions and distortions of impedance tensors could also be explained qualitatively by single elliptic cylinders.

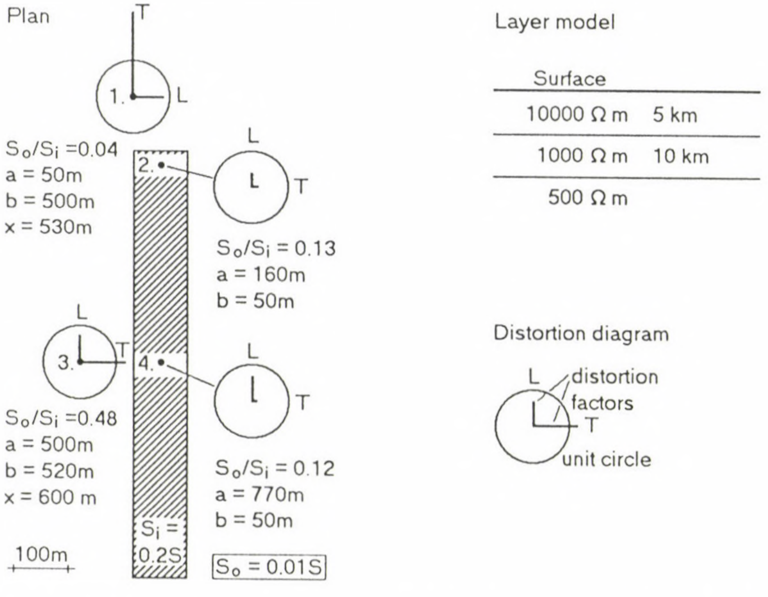


Fig. 6. Model of Zhang (1989) and cylinder model parametres interpretated

Conclusions

The cylinder model seems useful because equivalent solutions were found in all those cases which simplify to K-type. Usually, the parameters of the cylinders were comparable with those of the original model. If the shape of the body differs from the ellipse, it is valuable to try ellipses which are visible in the same sector as the original body.

Because some of distortions are very sensitive for model parameters, they should be determined on the basis of sounding observations (*i.e.* the forward modelling of poorly known bodies may lead to conclusions which are as faulty as the ones based on the original data). If one wants to apply forward modelling, it is best to use to T-soundings for which it succeeded adequately. L-soundings seem more involved because, for example, the resistivity contrasts in the models proved to be too large. However the L-soundings alternate less than T-soundings from one point to another.

Though the model of Ting and Hohmann differs from the principal K-type, the cylinder model also functioned well. The distortions were merely of the z-type rather than the S-type.

The LWA condition seems to be a more strict demand for the elliptic cylinder model. Especially in wide-band soundings, the LWA usually fails. The LWA is the demand which the model of Reddy et al. didn't fulfill. Though some solutions have a correct tendency outside the body, the model is merely an example of a case to which the cylinder model should not be applied.

If the resistivity decreases downwards (type Q and the beginning of type H) the cylinder model gives only indicative results. The models of Sternberg et al. and Zhang show that z-distortions seem preferable to S-distortions. The contribution of vertical currents should be tested in these cases.

Only a few 3-D results have been published. I did not find results which represent the effects of insulating bodies. Further, there were no examples of A-type layer structures. Because the A-type curve conforms with the beginning of the K-type curve, the distortions should be of the S-type. The LWA should be tested by models which match to the elliptic cylinder model as accurately as possible.

Appendix

Elliptic cylinder model

The K-type apparent resistivity curve consists of an ascending S-part and a descending z-part. The concepts S and z come from the layer parameters $S_1 = z_1/\rho_1$ and z_2, which mainly control the parts of the sounding curve. The model of Berdichevsky and Dmitriev (1976) consists of a vertical, elliptic cylinder which is situated in the first layer of a K-type layer structure (Fig. 7). The model considers the galvanic S-effect, current channeling, and induction. Berdichevsky and Dmitriev have presented the formulae for the distortion coefficients $K_{S,L}$, $K_{S,T}$, $K_{z,L}$ and $K_{z,T}$. The indexes L (longitudinal) and T (transversal) refer to the directions in which the cylinder is located. In the 2-D case, they coincide with E- and H-polarizations respectively. The observation point may lie inside or outside of the body. The ellipse may be either elongated ($a > b$) or flattened ($a < b$). Further, it may be conductive ($\rho_i < \rho_0$) or resistive ($\rho_i > \rho_0$). The formulae are so-called long wave length approximations (LWA).

The analysis of formulae shows that they are functions of the following dimensionless quantities: ratio of axes (a/b), resistivity contrast (ρ_i/ρ_0), and relative distance (x/b). Hence, the absolute size of the body or the magnitude of the resistivity have no influence on the distortions if the LWA holds. The relative parameters advance the use of nomograms in the study of distortions. One can compute auxiliary distortion coefficients from the original ones. The most useful ones are the ratios of distortions $R_S = K_{S,L}/K_{S,T}$ and $R_z = K_{z,L}/K_{z,T}$. Because the ratios are independent of the absolute values of S_1 or z_2, the corresponding ratios can be determined from the inversion results of L- and T-soundings or directly from the appropriate parts of sounding curves. The simultaneous interpretation of intersecting soundings is based just on these ratios (Lanne 1990).

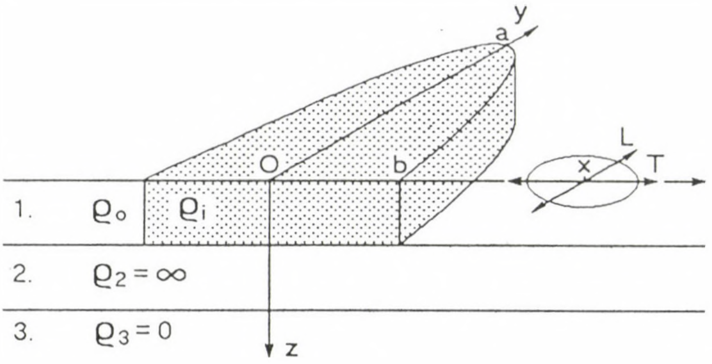


Fig. 7. Vertical elliptic cylinder in K-type layer structure (Berdichevsky and Dmitriev 1976)

References

Berdichevsky M N, Dmitriev V I 1976: In: Ádám A ed.: Geoelectric and Geothermal Studies. Budapest, Akadémiai Kiadó, 165–221.

Lanne E 1987: Lateraalisen pintarakenteen huomioonotto AMT-luotausten tulkinnassa. XIII Geofysiikan päivät Oulussa. Geofysiikan seura, Oulu, 33–38.

Lanne E 1990: Ristikkäisten AMT-mittausten tulkinta. Geologian tutkimuskeskus, Geofysiikan osasto, Open file report Q 16.2/24.33/90/1, p. 13.

Park S K, Orange A S, Madden T R 1983: *Geophysics*, 48, 1402–1405.

Reddy I K, Rankin D, Phillips R J 1977: *Geophys. J. R. astr. Soc.*, 51, 313–325.

Sternberg B K, Washburne J C, Pellerin L 1988: *Geophysics*, 53, 1459–1468.

Ting S C, Hohmann G W 1981: *Geophysics*, 46, 182–197.

Wannamaker P E, Hohmann G W, Ward S H 1984: *Geophysics*, 49, 1517–1533.

Zhang P 1989: In: Magnetotelluric study of the Siljan impact region: techniques and results. Ph.D. Thesis, Acta Univ. Upps., p. 21.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 231–239 (1992)

WATERTABLE VARIATIONS CONNECTED TO THE EARTHQUAKE IN BÉKÉS

T VARGA[1] and P VARGA[2]

[Manuscript received May 4, 1992]

Eleven watertable monitoring wells (WMW) situated on the epicentral area of the Békés earthquake (22.06.1978, $I_o = 6$) show level variations which can be attributed to seismic stressfield variations.

Keywords: Békés earthquake; seismic stress; watertable change

An extended scientific literature deals with hydrological problems related to the earthquake activity (Chi-Yu King 1985, Roeloffs 1988). There is a tendency to use the watertable variations in seismically active areas for the prediction of large earthquakes (*e.g.* Asada 1982, Mogi 1985). A lot of attempts are known to model the tectonic processes prior the quakes and to describe the local mechanisms using hydrogeological data (Rudnicki and Tze-Chi Msu 1988). Monitoring systems based on watertable observation are in operation, too (Wang 1985, Asteriadis and Contadakis 1992). The interpretation of the watertable records is a difficult problem due to the complexity of the stress accumulated in the epicentral area, and its effect in a complex hydrogeological medium (Roeloffs 1988).

So far only the description of watertable variations connected to strong earthquakes from active seismic areas are known. Since the establishing of the Hungarian network of watertable monitoring wells (WMW) started in 1936, the following earthquakes with epicentral intensity $I_0 > 6$ were used in our study:

In the present paper the behaviour of the watertable due to medium and small seismological events in Hungary is described by means of the use of observed data.

Time	Place	I_0	Uncertainty in determination of the epicentre (in km)
10.06.1937	Tarcal	6.0	+/ − 20
30.09.1942	Tápiósüly	6.0	+/ − 20
12.01.1956	Dunaharaszti	8.0	+/ − 10
31.03.1956	Pákod	6.0	+/ − 5
22.06.1978	Békés	6.0	+/ − 10
15.08.1985	Berhida	6.5	+/ − 5
27.10.1989	Bérbaltavár	6.0	+/ − 5

(Zsíros et al. 1988)

Geographical distribution of the above events is shown in Fig. 1, where circles around the epicentres — the epicentral areas — are proportional to the uncertainties in the determination of the epicentres.

[1]Eötvös Loránd Geophysical Institute of Hungary, H-1145 Budapest, Columbus u. 17–23, Hungary

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

Akadémiai Kiadó, Budapest

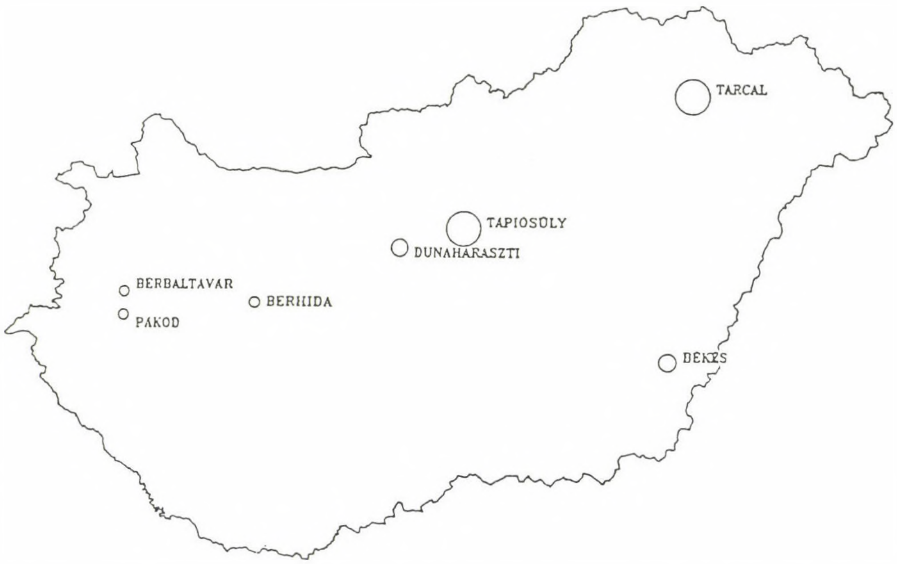


Fig. 1. Epicentral areas of Hungarian earthquakes since 1937 with epicentral intensities equal to or bigger than 6

The number of WMW is about 3500 in Hungary at present, but only twenty of them are on epicentral areas and "pass through an event" what means in our case that a respective WMW has a two years long record as well before as after the earthquake in question (Table I). In this respect the most interesting is the Békés earthquake, since in its epicentral area eleven WMWs can be found having the mentioned observation period of four years.

The existing watertable records were investigated from different points of view:

1. at first the records obtained during the days just before and after the Békés event

2. the correlation of the temporal variation observed in WMWs and the precipitation on the basis of five year long records

3. the anomalies of five year long records in the time interval of the earthquake itself.

1. Watertable data measured one week before and one week after the quake are listed in Table II (observations were carried out every three days). It can be seen that 1–5 cm level variations within this time interval are usual, but one day after the earthquake in the WMWs Békés-377, Békéscsaba-417, Békéscsaba-434, Békéscsaba-1567, Doboz-442, Gerla-439, level variations of 8–39 cm were observed exactly in the same moment and direction (decrease). The reason of the absence of such anomalies in the remaining wells can be explained — beside hydrological reasons — with the observation technics of the water level in the WMWs. On the

Table I. Watertable monitoring wells (WMW) on the epicentral areas of Hungarian earthquakes in the period 1937–1989

Earthquake		Watertable monitoring well	
Time	Place	Number	Name
10.06.1937	Tarcal	1705	Tokaj
		1506	Tiszalök
30.09.1942	Tápósüly	628	Jászberény
12.01.1956	Dunaharaszti	2081	Alsónémedi
		672	Gyál
31.03.1956	Pákod	1212	Zalaszentgrót
22.06.1978	Békés	374	Békés
		377	–"–
		381	–"–
		417	Békéscsaba
		428	–"–
		429	–"–
		431	–"–
		434	–"–
		1567	–"–
		442	Doboz
		439	Gerla
15.08.1985	Berhida	1155	Papkeszi
		1176	Ösi
27.01.1989	Bérbaltavár	1160	Nagytokaj

Table II. Technical data of the watertable monitoring wells (WMW) on the epicentral area of the Békés earthquake

Watertable monitoring well (WMW)				
Well Number	Name	Elevation (m)	Elevation of the top of the well's tube (m)	Length of the tube in the well (cm)
374	Békés	85.81	86.01	700
377	Békés	85.39	85.65	750
381	Békés	85.20	85.59	600
417	Békéscsaba	86.58	86.99	668
428	Békéscsaba	50.12	90.26	790
429	Békéscsaba	92.39	92.74	654
431	Békéscsaba	90.82	91.05	839
434	Békéscsaba	86.55	86.85	605
1567	Békéscsaba	86.15	86.84	722
442	Doboz	90.04	90.32	850
439	Gerla	85.29	85.74	800

basis of these data and those of the literature these changes cannot be explained by another (*e.g.* meteorological) event.

2. In order to provide the seismo-tectonic origin of the unusually big level variations linear correlation analysis has been carried out for the time interval 01.01.1976.–31.12.1980 between the time series of water levels (monthly mean value) and the monthly rainfall quantities most influencing the watertable. The possi-

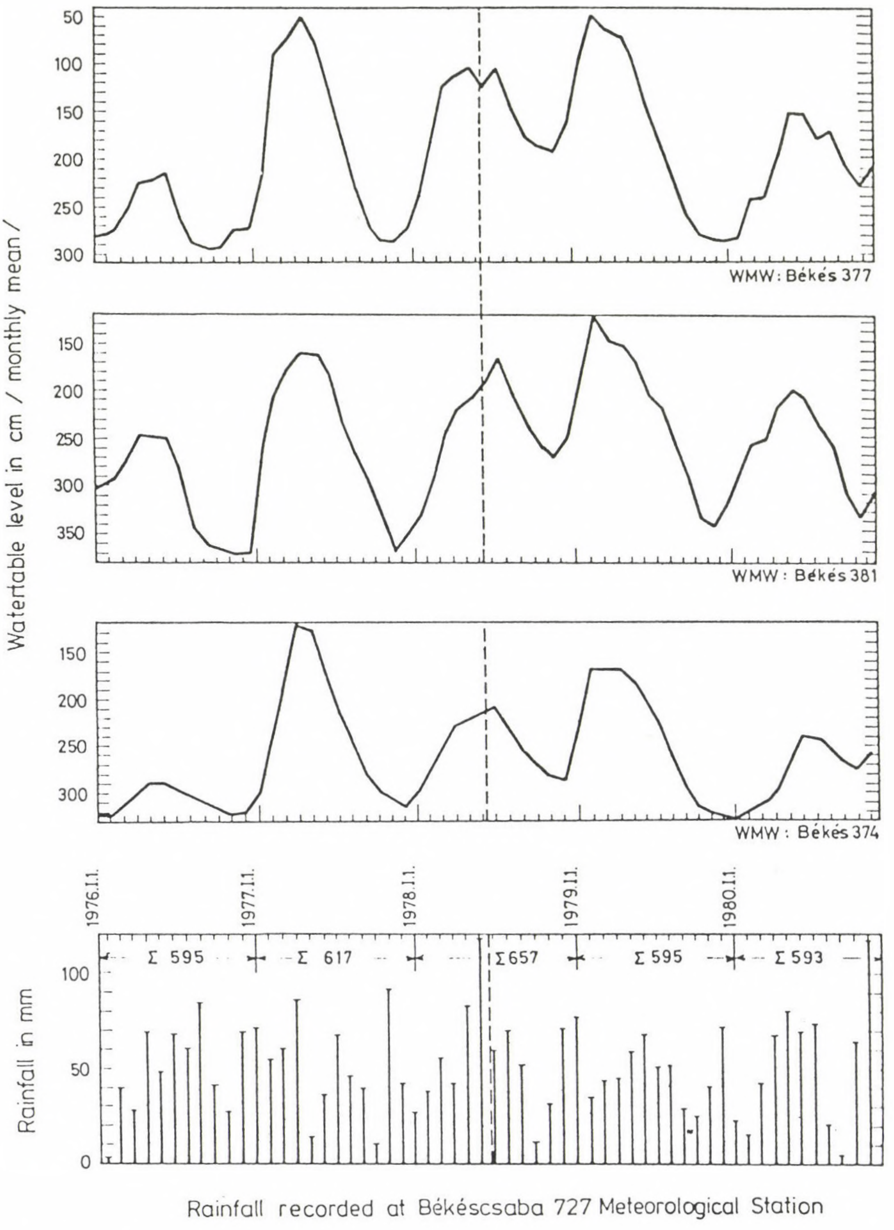


Fig. 2. a. Temporal variation of the watertable and the precipitation quantity on the epicentral area of the Békés earthquake. Σ 600 — annual quantity of precipitation; ¦ — time of the Békés earthquake; WMW — watertable monitoring well

ble influence of temperature and airpressure variations and the effect of the water evaporation have been neglected. The statistical analysis was made in the following

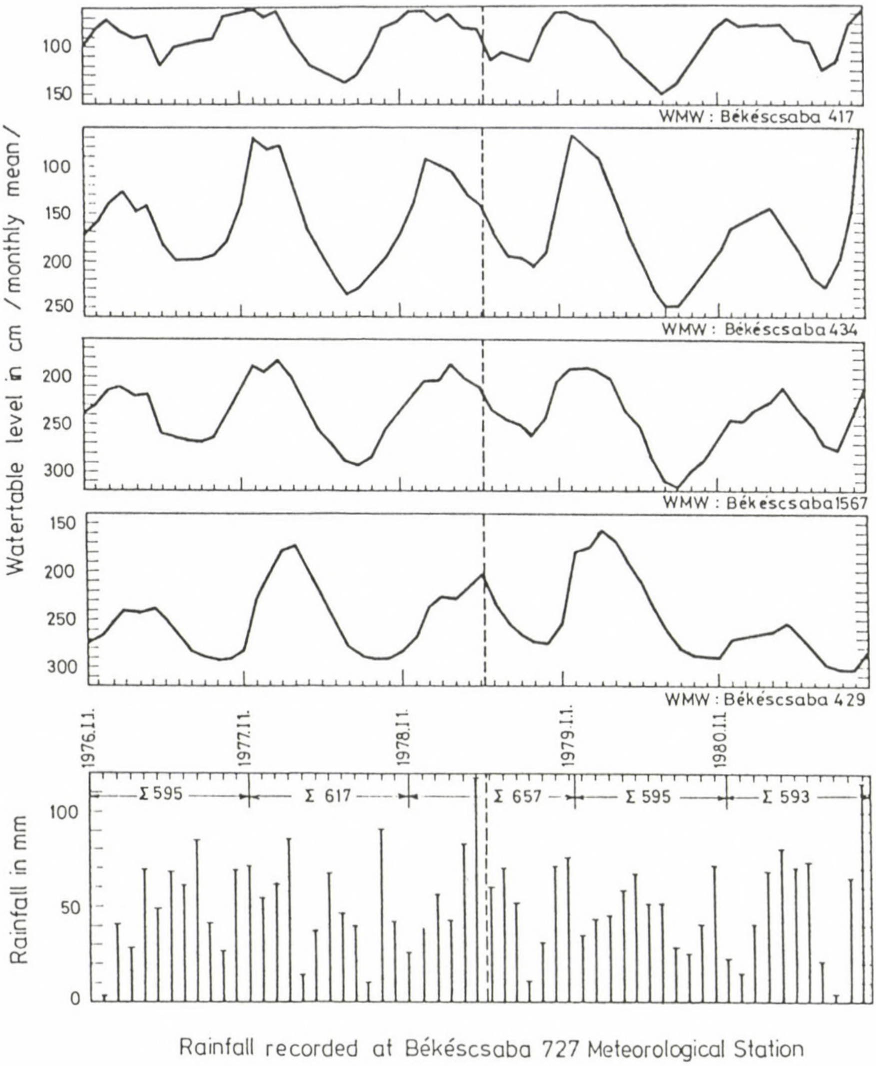


Fig. 2. b

way:

- The temporal behaviour of the watertable in the studied area shows a seasonal variation with minimum in November-December and maximum in April-May. For the time intervals between these extrema (4–11 values each), the correlation was calculated separately;

- the correlation for the total period in case of the WMWs listed above lies between 0.36 and 0.69.

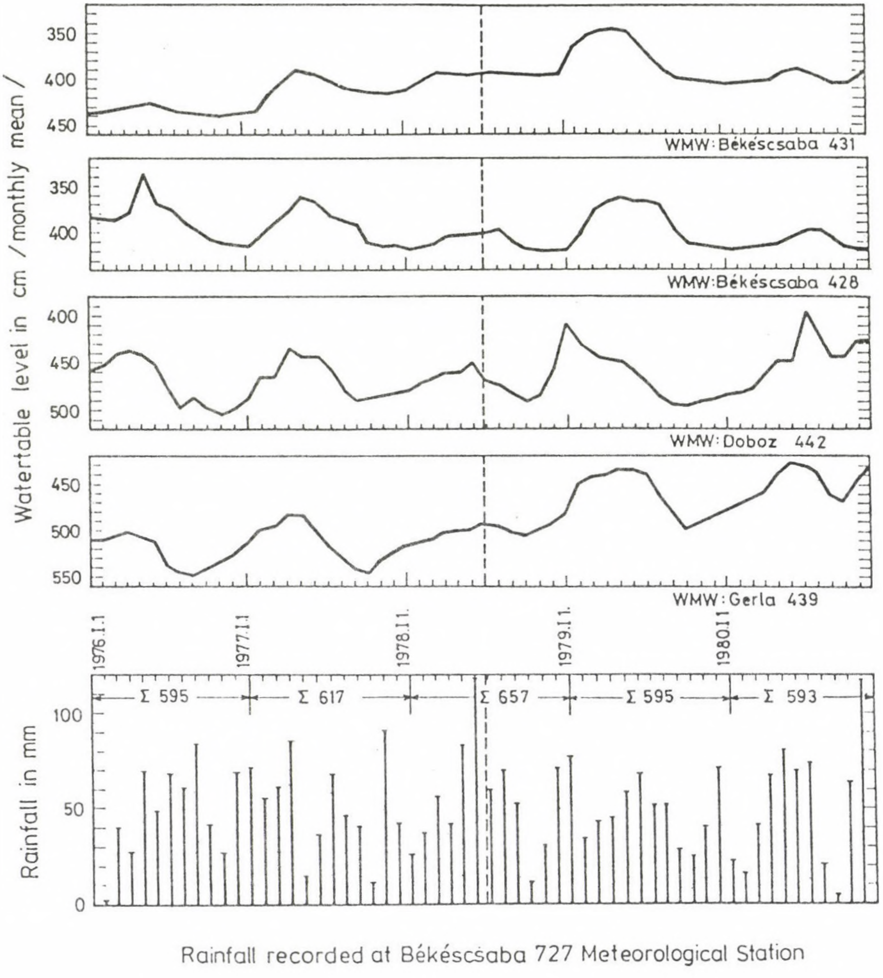


Fig. 2. c

After the exclusion of the time intervals including the time of the Békés earthquake and one period just before the event (both periods are characterised with correlation factors less than 0.5) the correlation between the watertable level and the precipitation quantity increased to 0.80–0.92, for all the other intervals of the studied four years.

The statistical investigation shows that the level of the watertable in the time of the quake or even before became independent from the time distribution of precipitation which usually determines the temporal changes of the underground waterlevel. In this time interval the watertable is a function of slow stress accumulation.

3. The time variations of the watertable and the precipitation is demonstrated in Fig. 2 for the case of all 11 investigated wells. One can see on the first look that

Table III. Level of the water in the watertable monitoring wells (WMW) during the weeks before and after Békés earthquake

Watertable monitoring well (WM)				
Time of the Békés earthquake	WMW Number	WMW Name	Day of observation	Level of the water in the well (cm)
22.06.1978	374	Békés	14.06.1978	219
			17.06.1978	217
			20.06.1978	215
			23.06.1978	211
			26.06.1978	209
			29.06.1978	208
	377	Békés	14.06.1978	123
			17.06.1978	126
			20.06.1978	132
			23.06.1978	137
			26.06.1978	140
			29.06.1978	102
	381	Békés	14.06.1978	195
			17.06.1978	192
			20.06.1978	189
			23.06.1978	185
			26.06.1978	180
			29.06.1978	177
	417	Békéscsaba	14.06.1978	72
			17.06.1978	79
			20.06.1978	87
			23.06.1978	92
			26.06.1978	97
			29.06.1978	61
	428	Békéscsaba	14.06.1978	403
			17.06.1978	403
			20.06.1978	403
			23.06.1978	403
			26.06.1978	403
			29.06.1978	403
	429	Békéscsaba	14.06.1978	211
			17.06.1978	209
			20.06.1978	209
			23.06.1978	209
			26.06.1978	210
			29.06.1978	205

the annual trends-which in case of seismically not disturbed years simply follow the changes in precipitation quantity — show in case of seismic activity an anomaly which increases with the depth of water level, but is independent from precipitation. The decrease of the underground waterlevel experienced in connection with the Békés earthquake refers to an increase of the effective porosity in the deeper layers of the studied area, i.e. to the dilatational extension. This conclusion must be taken into consideration when the tectonical processes causing the Békés earthquake, are studied.

A further investigation of the phenomenon described above needs an extension

Table III. (contd.)

Watertable monitoring well (WM)				
Time of the Békés earthquake	WMW Number	WMW Name	Day of observation	Level of the water in the well (cm)
22.06.1978	431	Békéscsaba	14.06.1978	395
			17.06.1978	395
			20.06.1978	395
			23.06.1978	395
			26.06.1978	395
			29.06.1978	395
	434	Békéscsaba	14.06.1978	128
			17.06.1978	128
			20.06.1978	133
			23.06.1978	138
			26.06.1978	143
			29.06.1978	133
	1567	Békéscsaba	14.06.1978	191
			17.06.1978	202
			20.06.1978	207
			23.06.1978	212
			26.06.1978	217
			29.06.1978	178
	442	Doboz	14.06.1978	460
			17.06.1978	459
			20.06.1978	457
			23.06.1978	462
			26.06.1978	466
			29.06.1978	445
	439	Gerla	14.06.1978	494
			17.06.1978	497
			20.06.1978	500
			23.06.1978	502
			26.06.1978	502
			29.06.1978	494

of the area monitored by WMWs. This is first of all valid for the epicentral areas of earthquakes mentioned in the beginning of this paper where only one WMW was run before and after the seismological events.

In the scientific literature (e.g. review paper of Roeloffs 1988) the relation of the earthquakes and watertable variations are only discussed on a static basis. In future investigations — based on the knowledge got in Hungary — the model used in this paper will be improved by the introduction of hydrodinamical parameters.

References

Asada T ed. 1982: Earthquake prediction techniques. Their application in Japan. University of Tokyo Press

Asteriadis G, Contadakis M-E 1992: The validity of a network for the monitoring of the level and temperature of the shallow underground water in earthquake prediction. Paper presented at the EGS Assembly, Edinbourgh

Chi-Yu King ed. 1985: Earthquake hydrologie and chemistry. Birkhäuser

Mogi 1985: Earthquake prediction. Academic Press
Roeloffs E A 1988: *PAGEOPH*, 126, 177–209.
Rudnicki J W, Tze-Chi Hsu 1988: *J. Geoph. Res.*, 93, 3275–3285.
Hydrographic yearbook (in Hungarian) 1936–1988: VITUKI
Wang C 1985: *PAGEOPH*, 122, 215–217.
Zsíros T, Mónus P, Tóth L 1988: Hungarian earthquake catalog (456–1986). Budapest

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 241-249 (1992)

COSMICAL INFLUENCES ON THE GEOLOGICAL TIME

H J TREDER[1] and W SCHRÖDER[2]

[Manuscript received May 15, 1992]

The paleogeophysics of the earth's surface is not at last determined by cosmic influences, in particular by solar radiation. There is a possibility of secular variation of the solar constants in connection with such a one of earth's orbit elements.

Furthermore, long-period solar constant variations (with periods covering 300 million years) may be considered. No doubt that other solar activities (corpuscular radiation, *etc.*) are highly variable quantitites.

Keywords: change of the gravity constant; galactic dynamics; length of day; length of year; Maunder minimum; solar activity; solar nuclear fusion

I

Serge v Bubnoff referred in his fundamental book, "Basic problems of geology" (1959) to the great geologic cycles which imply according to his opinion specific problems for the concept of time in geology. Among others Bubnoff emphasized the great geotectonic and paleobiological cycles with lengths of 200 to 300 million years and in this connection he mentioned the possibility that geological processes get accelerated by a secular shortening of the astronomical year. For him this question was connected with the main problem, with actualism in geology: he basically defended the actualist point of view, but he remarked very correctly that the universal validity of the physical principles did not exclude that these principles acted on the earth in different conditions in different times.

This question can be discussed further in the framework of cosmical physics. Physics starts naturally from the validity of the physical laws everywhere and at every moment in the universe. Theoretical physics formulates this insight mathematically by Einstein's general principle of relativity which states that nature's laws are in a general form independent of conditions of their action. These conditions change locally the processes of development determined by the laws of nature. They change universally and globally due to evolution of the universe reflected by the notion "cosmological age of the universe".

The age of the earth, *i.e.* 4 to 5 billion years is comparable with this age of the universe amounting to 15 to 20 billion years. The earth, and especially the physics of the earth's surface (and correspondingly the climates of geological epochs) is influenced by solar irradiation. The solar irradiation is on the one hand depending

[1]D-1590 Potsdam, Rosa-Luxemburg-Strasse 17a, Germany
[2]D-2820 Bremen-Roennebeck, Hechelstrasse 8, Germany

on the actual stage of development of the sun, on the other hand on the distance between earth and sun. The terrestrial climate is thus significantly a function of the length of the day, that is why both movements of the earth, its path around the sun and its rotation around the axis represent important data for geology and paleogeophysics by their constant or variable character.

The most important effect of the sun is at the earth's surface the electromagnetic radiation in visible (and in neighborous ultraviolet and infrared) range. It is, however, often emphasized that corpuscular radiation of the sun and solar wind might possess and effect *e.g.* on the development of life on the earth, too.

The universal validity of the physical laws which determine both the development of the sun, the motion of the earth as well as physical processes on the earth, is by no means only a theoretical insight of physics, but also an empirically well proven fact. The empirical problem is mostly summarized as the problem of temporal constancy of certain fundamental physical constants playing eminent roles in physics and chemistry, such as the electric elementary charge e, Plank's action constant h, Sommerfeld's fine structure constant $e^2(hc)^{-1}$, masses of stable elementary particles *etc.* In accordance with theory, accurate measurements of experimental physics prove with incredibly high precision the constant value of the quantities G in the sense that the temporal variability is much less than a relative value

$$\dot{G}/G < 10^{-11} \text{ per year.}$$

As great secular changes consist of small momentary changes, this relation means at the same time the long time constancy of these quantities.

This temporal constancy is proven further by astrophysics of extragalactical objects. Due to the limited value of the speed of light c astrophysicists investigate simultaneously spectra of objects from galaxies which are in distances of several billion light years, therefore the physical state of matter there corresponds to conditions several billion years backwards. With the high accuracy of just the spectroscopy it can be shown that laws of quantum mechanics and atomary constants which determine spectra have remained the same in time.

II

The constant of nature which has basic importance in connection with v. Bubnoff's problem of the variability of the lengths of day and year in geological times is just that which may still contain both theoretically and practically, too, a slight variability *i.e.* Newton's gravity constant f. The important role of the gravity constant depends on the one hand on the extreme weakness of gravitational attraction which have practically no influence on the structure of atoms and which can be therefore hardly tested by methods of atomic physics, on the other hand the universal character of gravity leading to equal effect on all masses and energies and therefore the relative importance of the processes is hardly influenced by it.

Present precision gravimetry and the high precision determination of the distances of natural and artificial celestial bodies in the solar system using the methods of radar astronomy and extraterrestrial space research show that it is possible that the gravity constant may change due to "age of universe", with the estimation

$$|\dot{f}/f| = 10^{-10}\,\text{per year}.$$

Such a change of the gravity constant is not necessarily existing. It is, however, according to our present knowledge possible, but the accuracy of the measurements is low enough to allow both an increase and a decrease of the gravity in cosmical scales of time (it is not to be expected in reasonable time that significantly higher precision will be achieved on the constancy or variability of f. It is namely not the inner accuracy of the measurement methods which determines the accuracy of the measurement here but a series of disturbing effects which result from a complex of effects to be considered. Nevertheless, an experimental decision about the proportionality of f with the age of the universe is possible, at least with an accuracy of an order of magnitude).

Dirac's suggestion some 50 years ago led to a speculation about field theory. He supposed a secular decrease of f according to which (using an age of the universe of 15×19^9 years) one has:

$$\dot{f}/f \approx -2/3 \cdot 10^{10}\,\text{per year}.$$

Newton's gravity constant used earlier to be according to Dirac greater than today and it was at the beginning of the existence of the earth (5 billion years before present) by about one third more than today. Celestial mechanics tells us that with decreasing **f** the distance **r** between sun and earth *i.e.* also the rotation period of the earth around the sun increase and correspondingly the length of the year, T too:

$$-\dot{r}/r = \dot{f}/f; \quad \dot{T}/T = -2\dot{f}/f = 4/3 \cdot 10^{-10}\,\text{per year}.$$

Simultaneously the solar constant S does change with the distance according to Lambert's law, too:

$$\dot{S}(r)/S = -2\dot{r}/r = 2\dot{f}/f.$$

All these are purely celestial mechanics. But the physical theories on the construction and development of the sun, on the production of the solar heat by nuclear fusion in the sun's interior tell us that the radiation temperature of the sun, and consequently the solar constant depend even more on **f**. This dependence of the solar constant is:

$$S(f) \sim f^7 \qquad S(r,f) \sim r^{-2} f^7$$

according to which the solar constant would be after Dirac's hypothesis

$$\dot{S}/S = 9\dot{f}/f = -6 \cdot 10^{-10}\,\text{per year}.$$

That means that a secular decrease of the gravity constant would mean that the length of the year had been earlier shorter than today and the solar constant S had been in previous geological times much greater than today, too. It would be twice the present value in Archean what is geologically excluded. Further a prolongation of the year, $T > 0$ in the earth's history contradicts geological facts discussed by v. Bubnoff. According to them the length of the year, expressed in days has diminished:

$$\dot{T}/T \approx -10^{-10} \text{ per year.}$$

(A mere secular change of the gravity constant cannot evidently explain geological great cycles which have to be connected with long period effects.)

It should be remarked that a change of the length of the year as related to the length of the day may also mean that the length of the day **d**, as measured in absolute physical time scale has continuously increased. It may, however, also mean that the length of the year has actually decreased as meaured again in absolute physical time scale. There are physical hypotheses for both possibilities which would explain them. An increase of the length of the day in geological times during the earth's history according to the empirical formula

$$\dot{d}/d \approx 10^{-10} \text{ per year}$$

could result from tidal friction which causes a transfer of the rotation momentum of the earth on the momentum of the movement of the moon on its path. This mechanism has been studied especially by H G Darwin; it is even today much debated and extremely difficult to extrapolate it to long time intervals, as its effect depends strongly on the actual structure of the shelf seas.

A hypothesis which would cause a realistic decrease of the length of the year and with it, a secular decrease of the distance earth-sun, $\dot{r} < 0$, (and also moon-earth) is the Mach-Einstein doctrine on the relativity of the inertia. This principle by Mach leads if formulated in the language of the analytical mechanics and applied to the actual expanding universe to the following estimates of the changes of **r** and **t** (referred to the present):

$$5 \cdot 10^{-11} \leq \dot{r}/r < 10^{-10} \text{ per year}$$

$$10 \cdot 10^{-10} \leq \dot{T}/T < 2 \cdot 10^{-10} \text{ per year}$$

which implies an increase of the solar constant as

$$10^{-10} \leq \dot{S}/S < 2 \cdot 10^{-10} \text{ per year.}$$

It is similarly possible that both effects, an increase of the length of the day, $\dot{d} > 0$, and a decrease of the length of the year, $\dot{T} < 0$, occur simultaneously and sum up.

It is interesting to note that the principle of the relativity of inertia leads — if applied to the actual cosmical neighbourhood of the earth — to a long period influence of the effectivity of the terrestrial gravity on the physical and chemical processes on the earth.

The solar system rotates in a time of 250 million years around the centre of the Milky Way system in an elliptical parth and it is affected during periastron by a somewhat stronger gravity potential than during apastron. The inertia principle tells us, however, that during periastron terrestrial gravity was somewhat less than during apastron. The approximative values of the parameters of the galactic dynamics yield values of the periodic change of the effective terrestrial gravity g within 250 million years with a relative amplitude of about $\Delta g/g$ of -10^{-7}.

This is a very small effect but in some theories of the earthquake occurrence a (hypothetic) effect being by one order of magnitude less in the terrestrial gravity with a yearly amplitude of $\Delta g/g \approx \pm 10^{-8}$ is included as provoking factor. It is interesting to note that this supposition would physically interpret the often emphasized approximative equality of the great cycles studied by v. Bubnoff and by others (Schindewolf, Umbgrove, Schwinner) and of the rotation period of the galaxy.

III

This time interval of parameter propter 300 million years may influence the earth from the universe by other ways, mainly by a paleoclimatological way. Till quite recently astrophysical models of stars told us that (supposing a temporally constant value of the gravity constant **f**) the energy production on the sun and with it, the temperature of the sun and consequently — according to the radiation laws by Planck, Wien und Stefan-Boltzmann — the quality and quantity of the solar electromagnetic radiation and the solar constant **S** did not change during the existence of the earth. As the expected lifetime of the sun is greater than 10 billion years, the sun has had during the lifetime of the earth, *i.e.* during the last 4 billion years always the same radiation (based on these astronomically well founded models).

The source of the radiation energy is nuclear fusion — hydrogen into helium — in the solar interior at a central temperature of about 15 million °C. The radiation energy produced in the sun's interior is many times absorbed and re-emitted during the path to the sun's surface and it is finally transformed to a radiation temperature of 6000°C. This process of the transfer of radiation from the sun's interior to its surface has a duration which is governed by a fundamental thermodynamical quantity of solar physics, by the Helmholtz-Kelvin period of about 20 million years. This period is a purely thermodynamical quantity and is absolutely independent of the way of energy production and heat transfer. That is why it could be deduced some 100 years ago by Helmholtz and Kelvin (who did not know anything about nuclear physics).

In addition to the electromagnetic radiation, neutrinos are also emitted from the solar interior according to the laws of nuclear fusion. This process carries 2 percent of the radiation energy. Neutrinos are particles with many similar properties to the photons of light, first of all they have the speed of light. In contrary, however, they have very weak interaction with matter and are therefore not absorbed in the solar interior. Thus they arrive in two seconds from the solar centre to the sun's surface. Would it be possible to record the solar neutrino current at the earth's surface, one could look into the central part of the sun where nuclear fusion takes place.

Without going into complicated details of nuclear physics, it should only be remarked that up to now solar neutrinos could not be detected on the earth's surface — in spite of ensured theoretical expectations — though such experiments have been carried out in the last 10 years with instruments being able to record neutrinos produced by terrestrial nuclear reactions.

A possible explanation of this disturbing paradox is that the sun does not operate at present as a nuclear fusion reactor. Then two possible sources of the solar heat radiation remain:

1. an existing heat supply which is going to be exhausted, or

2. the contraction of the solar interior due to the exhaustion of the heat which transforms potential energy into heat by the method described by Helmholtz and Kelvin.

The picture resulting from these conditions for the sun is that the sun is no "continuous burner", but it works like a diesel-motor. During time intervals estimated theoretically to about 250 million years the sun operates as a nuclear fusion reactor and ensures so the equilibrium. This fusion-reactor has a burning temperature of about 15 million°C. The sun makes during this time a little too much and this leads to an expansion of the solar interior, *i.e.* it carries out labour and therefore it gets cooler. The central temperature in the sun decreases, the strongly temperature-dependent fusion-reaction stops quickly and cannot cover the energy balance. There is, however, a period of a few times 10 million years when the sun produces radiation — according to the Helmholtz-Kelvin-period — with the original solar constant due to its energy reserve, but it does not produce any observable neutrinos. Then follows a period (of similarly about 10 million years) when the solar radiation decreases, too. During these periods the sun's interior contracts what results in heating up the sun (according to the old thermodynamical theory by Helmholtz and Kelvin) in about 20 million years and correspondingly it starts again to work as a fusion reactor. Now it lasts again about 10 million years till the sun reaches its original radiation intensity. During long periods of undisturbed nuclear fusion (some hundred million years) the solar constant is higher than during short intervals of switched off energy production: during a Helmholtz-Kelvin period of about 20 million years the solar constant is less than during normal periods longer by one order of magnitude.

This (qualitatively too) completely hypothetical model of the sun as a diesel-motor (based solely on the lack or too low quantity of the observed solar neutrinos) would seemingly explain why paleoclimatology found that the average temperature of the earth had been significantly higher in long geological intervals of about 300 million years than today, that even poles had been free of ice caps, and these long warm intervals had been interrupted by shorter intervals of lower temperature. During these periods poles were covered by ice and ice-ages might occur.

We would be so just in an age when the nuclear fusion is interrupted for such a long time that the sun significantly cooled down and presently a contraction in a Helmholtz-Kelvin period warms it up. Even if all this is astronomically very speculative, it is the more interesting for paleogeophysics as it could yield some proof for a verification of this new solar model. The previous apodictical statement about the invariability of the solar constant during the history of the earth cannot be sustained without further proof.

IV

The solar-terrestrial physics s.s, *i.e.* the effect of the solar activity on geomagnetism and on the terrestrial atmosphere (together with the indirect effect on the biosphere *etc.*) are much less stationary than the solar constant. The data of the present solar activity, its 2×11 years cycle (Bjerknes discovered 1926 the pole change) and their intensity expressed by the relative sunspot numbers cannot be extrapolated in geological timescales, even if this would be suggested by some monographs on solar physics and by some theories on the magnetohydrodynamic causes of solar magnetism.

Magnetohydrodynamics yields a realistic estimate of the periodicity of solar magnetism and of the solar activity controlled by it in function of the construction and rotation of the sun. It cannot, however, give a realistic estimate of the absolute value of these effects. (This is a consequence of the linearization of the hydromagnetic equations.) Solar magnetism is a very peculiar phenomenon of the solar atmosphere whose excitation seems to depend on many random factors.

Sunspots were discovered by Johann and David Fabricius, G Galilei, J Scheiner and Th Harriot. The sunspot numbers have been systematically recorded since 1749, and the 11-years cycle was discovered in the previous century by H Schwabe. Rudolf Wolf succeeded to continue the series of sunspot numbers backwards till into the 17th century using single observations. Ancient sources and medieval chronics — from the Far-East, too — contain observations of big sunspots which could be seen by naked eye. This was mentioned among others by A von Humboldt in Volume 3 of his Kosmos.

It has been much discussed in the last years that much less sunspots have been observed in the 17th century than suggested. This had been noted by expert observers such as J Havelius, J Flamsteed, D Cassini and E Halley. They remarked that in the second half of the 17th century and in the first decade of the 18th century — *i.e.* during five or six 11-years periods — practically no sunspots were observed. Thus J Havelius (1611–1687) observed the last sunspot of his life before 1650. This fact has been mentioned in the first German "Physical Vocabulary" by J S T Gellert in 1791; Humboldt's Kosmos contains further data.

These facts led J A Eddy to suppose a general long-time solar activity minimum, the so-called "Maunder-minimum" (1645–1715). Initially Eddy's supposition seemed to be confirmed as there were scarcely sunspot data in the sources he surveyed. Since then, however, detailed studies by Gleisberg, Landsberg, Legrand and Schröder have shown that his conclusion was too hasty. It is evidently true that less sunspots have been observed then, but this could be due to special circumstances. The Maunder-minimum followed the thirty years war (1618–1648) when observations were hardly possible or got destroyed. That means that an anyway scarce quantity of sources had been further reduced. It is to be added that the observations of that time were complemented by single observations from the series by Lalande, the family Kirch in Berlin, Liebknecht *etc.* The reconstruction gives now a picture differring from that suggested by Eddy.

Auroras are caused by particles emitted by the active sun. At the beginning of the modern times many descriptions are found on auroras in leaflets (Schröder 1984) and in "Neue Zeitungen" (new journals). These were collected *e.g.* by Fritz (1873) in a catalogue. It is a significant fact which is also found in Gehler's vocabulary under the item Northern Light that after 1600, *i.e.* in the second half of the 17th and at the beginning of the 18th century practically no auroras were observed not only in Scandinavia, but also in Greenland. These facts are based on such eminent observers, as *e.g.* Celsius and Halley. The latter saw the first aurora of his life as a sexagenarian in 1716. Anyway, the causal connection auroras — solar activity had not been known. The two phenomena were observed independently and the corresponding data series are therefore independent, too.

Eddy considered the frequency of occurrence of auroras as an indicator for the very low solar activity during the Maunder-minimum. This situation has, however, very significantly changed in consequence of studies by Landsberg, Legrand and Schröder: during the Maunder-minimum we find practically the same periodicities as earlier. It is, however, not to be forgotten that in this interval (1645–1715) there are several years when no data are available on sunspots or auroras. It is to be decided if this lack is due to lack of observations, due to the loss of original records or due to irregularities of the sun.

One can start from the fact that in the near past there were time intervals, too when the solar activity was very low and when this activity had short-time variations. It is therefore wrong to presuppose that solar activity has been always the same in the geological past, namely equal to that today. In this sense solar activity must not be taken as a fixed astrophysical quantity for the paleogeophysics.

References

Bahcall J W 1990: *Sci. Am.*, 262, Wr 5, 26–33.

Bjerknes V 1926: *Astrophys. J.*, 64, 93.

v. Bubnoff S 1949: Einführung in die Erdgeschichte. Mitteldeutscher Verlag, Halle

v. Bubnoff S 1959: Grundprobleme der Geologie. 3. Aufl., Akademie-Verlag, Berlin

Darwin H G 1911: Ebbe und Flut. 2. Aufl., Teubner, Leipzig

Eddy J A 1976: *Science*, 192, 1189–1202.

Eddy J A 1983: *Solar Phys.*, 89, 195–202.

Ertel H 1935: Einsteins kosmologische Konstante und der Zusammenhang von Atom- und kosmischen Konstanten im Evolutionskosmos. Berliner Ber. Nr. I

Gutenberg B ed. 1951: Internal Constitution of the Earth. Dover, New York

v. Helmholtz H 1984: Vorträge und Reden. Vol. I, II. Vieweg, Braunschweig

Lord Kelvin 1887: The Sun's Heat. Roy. Inst. Dis., London

Kraus E 1971: Die Entwicklungsgeschichte der Kontinente und Ozeane. 2. Aufl., Akademie-Verland, Berlin

Kuchowicz B 1975: The Solar Neutrino Discrepance. 14. Int. Cosmic Ray Conf., München, 6, 2129–2132.

Landsberg P E 1980: *Arch Met. Geophys. Bioklim.*, B 28, 181–191.

Maunder E W 1922: *British A.A.J.*, 32, 140–145.

Munk W H, Macdonald G J F 1960: The Rotation of the Earth. University Press, Cambridge

Schindewolf O H 1950: Der Zeitfaktor in der Geologie und Palaeontologie. Hirzel, Stuttgart

Schröder W 1984: Das Phänomen des Polarlichtes. Wiss. Buchgesellschaft, Darmstadt
Schröder W 1988: *Meteorol. Atmos. Phys.*, 38, 245–251.
Steenbeck M, Treder H J 1984: Möglichkeiten der experimentellen Schwerkraftforschung. Akademie-Verlag, Berlin
Treder H J 1969: *Gerlands Beiträge*, 78, 1–26.
Treder H J 1972: Die Relativität der Trägheit. Akademie-Verlag, Berlin; Atomizdat, Moskau 1975
Treder H J 1990: *Berliner Berichte*, 9 N
Treder H J, v. Borzeszkowski H H, v.d. Merwe A, Yourgrau W 1980: Fundamental Principles of General Relativity. Plenum, New York
Umbgrove J H F 1947: The Pulse of the Earth. 2. Aufl., Haag

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 251–282 (1992)

SOLAR AND AURORAL ACTIVITIES DURING THE SEVENTEENTH CENTURY

J P LEGRAND[1], M LE GOFF[2], C MAZAUDIER[3], W SCHRÖDER[4]

[Manuscript received July 4, 1992]

The dark spots on the solar disk visible by eye and the aurorae, including some of them stretching down towards tropics, are the most visible manifestations of the solar activity. Since more than one century, we know that this activity follows a rather regular periodic cycle of eleven years, but with an intensity, measurable in spot number, which is sufficiently variable to speak of "small" or "large" solar cycles.

Actually we observe a series of large cycles, on the contrary at the end of the 17th century this activity was incomparably weak during several decades, at the epoch of the rise of the instrumental astronomy in Europe.

This period of weak intensity which has attracted the attention of the German astronomer Spoerer as early as 1890, is known by the name of "Maunder Minimum".

What are we knowing exactly about this epoch?

Were the tools used accurate enough?

Did the climate conditions, known to be cloudy, get the better of the astronomer's assiduity?

What can we bring from the analysis of more ancient data, dating from the beginning of the 17th century at the time of astronomical telescope invention?

This paper makes the state on all these historical questions and highlights the original documents at the light of more recent knowledge on the running of our star.

Keywords: auroral activity; history of geophysics; Maunder minimum; solar activity

The Spoerer minimum

"It is just one-third of a century since Prof. F W G Spoerer, the veteran observer of sunspots, published two important papers with regard to the "sunspot cyle", in which he drew attention to evidence pointing to a long continued "damping-down" of the solar activity which began in the middle of the seventeenth century. It fell to my duty to supply a short note on these two papers to the Council of the Royal Astronomical Society for their Annual Report for the year 1890, and four years later, I gave a fuller account of the first paper, in knowledge for August 1, 1894 (p. 173).

It appears to me that this discovery of Spoerer has not received as much attention as it deserves, and I would ask permission from the Association to summarise the main facts."

[1]CNRS-INSU, 4 avenue de Neptune, F-94107 Saint-Maur Cedex, France

[2]CNRS-Labor. de Geomagnetisme, 4 avenue de Neptune, F-94107 Saint-Maur Cedex, France

[3]CRPE, 4 avenue de Neptune, F-94107 Saint-Maur Cedex, France

[4]Geophysikalische Station, Hechelstrasse 8, D-2820 Bremen-Roennebeck, Germany

It is in the previous terms that the astronomer E W Maunder began his paper "The prolonged minimum 1645–1715" published in 1922 by the "British Astronomical Association Journal" in which he gave a description of the sunspots and aurorae observed at this epoch, showing the reality of this weak solar activity period.

Then in 1976, Eddy published a new paper on this subject confirming the thesis of the prolonged minimum. Curiously he named his paper "The Maunder minimum", even this discovery belongs to Spoerer, according to the English astronomer.

After Eddy's paper, several textual criticisms were published especially by Gleissberg and Damboldt (1979), Kopeczky and Kuklin (1987), Landsberg (1980), Link (1978), Schröder (1979, 1988), Zhen-Tao (1982), some of them gave a clear indication that the solar activity, even weak during this period, has already followed the well known solar cycle variation. And, as a consequence, there was not a long minimum of sunspots. In particular, the frequency of aurora appearance during the 17th century was used for these studies.

But during these periods of stammering instrumental astronomy, the sunspot and aurora data were certainly not homogenous because there was no regular auroral watch. This fact does not allow their interpretation in terms of activity level.

In this study, all the observations available have been collected and discussed in order to determine their reliability and especially the exhaustivity. Then a detailed synthesis, made at the light of the recent progress made in the knowledge of the solar source of the geomagnetic activity (Legrand and Simon 1989, Simon and Legrand 1989), allows us to show the reality of the desorders which occurred in the "Solar machine running working", precisely during Louis XIV's reign ("Sun King").

The sunspot activity during the seventeenth century

The solar data available for this century of evolution of the astronomy of precision can be classified following these periods:

- the period of discovery of sunspot activity with the help of astronomical telescope (1610–1645)
- the period of "uncertainty" (1646–1670)
- the period of weak activity (1671–1710).

The period of discovery of sunspot activity and the astronomical telescope invention (1610–1645)

Historic

The use of telescope to see more clearly distant objects started in Italy between 1580 and 1586, and in the Netherlands in 1608 (Danjon and Couder 1935).

The first telescopes built at this epoch, with a convex lens and a concave ocular, were made by competent opticians or technicians. These tools were rare for a long time, as it was not easy to acquire a technology whose secrets were jealously kept. Nevertheless, this invention was too bright to be a long time in the sole hands of a few constructors. In a short time, this invention was known by many people and erudits were not the last one to be interested by it, in particular the famous Galileo from Florence.

It is through his Paris corresponding friend, Jacques Badouere that Galileo heard of this "Dutch lens", for the first time, in June 1609.

These tools had just appeared at merchants in the French capital (Danjon and Couder 1935). Pierre l'Estoile was stating these facts in his journal on Henri IV's reign for the year 1609: *"On April 30, Thursday, I went to an optician at pont Marchand and he showed to me and to several others a new type of glasses. These glasses consist of a one foot long tube with glasses at both ends which differ from each other; they are used to observe distant objects which can be seen else confusedly: you have to put the glasses near to one eye and to close the other; if you observe the object you want to see, it is much nearer and you see it very clearly so that you can identify him from a half mile: It was told me that these glasses were invented by an optician in Midelbourg in Zeland and he presented two of them last year to Prince Maurice with which distant objects could be seen very clearly, from three or four miles. This Prince sent them to the Council of the United Provinces which gave to the inventor as compensation three hundred écus with the condition that he should not tell anybody how to make similar ones."*

It is nearly sure that Galileo never had such a telescope in his hand, but, with the use of the refraction theory, he rapidly discovered the construction of the "approach lens".

After July 1609, he undertook the construction of a telescope by using a long pipe with two glasses at the extremities, one convex, the objective, the other concave, the ocular, and established that the observed objects were three times enlarged.

He built other telescopes: three of them are in the Museum of Science History at Florence, their characteristics are given in the following table:

diameter of the objective (cm)	effective diameter (cm)	focal distance (cm)	amplification power
5.1	2.6	132.7	14
3.7	1.6	92	20
3.8	objective partly destroyed	169	?

To obtain a better image definition, Galileo was stopping down the objectives. The resolution power of this lens was of 10 to 15″ with a field of view of approximative one half of the solar diameter.

With these tools Galileo discovered the Jovian satellites on January 7, 1610, the planets phase, the lunar mountains, the nebulas, the sunspots and the milky way stars. The telescope so described, was the sole operational during a long time. It had the inconvenience to have a narrow field of view.

Later Kepler showed in his "Dioptrique" published in 1611, that one can substitute the divergent ocular of the telescope by a convergent one; and, therefore he is the inventor of the astronomical telescope which was the first used by Father Scheiner, Jesuit at Ingolstadt.

If Galileo seems to be the first to observe the sunspot with a lens, it is Fabricius, young Dutch student in medicine, who was the first to publish a work named " Joh. Fabricii Physii de maculis in sole observatis et apparute earum cum sole conversione narratio, Wittenbergae, 1611 petit in 4° ". In this work of forty three pages, with dedicatory epistle, dated June 13, 1611, only eight pages concern sunspots.

Lalande (1778), translated all the Fabricius work on this subject as follows: *"Having invented the telescopes in Holland, one started to observe the Moon, then Jupiter and Saturnus and Galileo found there curious things: I myself have observed the Sun driven by the same curiosity, its borders seemed to me to have remarkable inequalities which have already been observed by my father David Fabricius as I learned from his letters he sent me during my study. I have observed (6 March, 1611) a black spot on the Sun with one side less dense and more pale, rather big as compared to the solar disk; at first I thought it is a cloud; but having observed it ten times with different glasses and calling my father to observe it, too we got sure that it is no cloud; as the Sun ascended more and more that we could not observe it any more, during more than two days the appearance of these objects changed: that is why I propose those who want to make similar observations, at first they should receive only a small part of the light to get accustomed the eye so that it could tolerate the light of the whole solar disk. We spent the rest of the day and the next night very impatiently and dreamt about what this spot could be: if it is on the Sun, I would see it again, if it is not on the Sun, its movement would make it for us invisible. Finally I have seen it again the next morning with great delight, but we have seen, too, that it changed its position a little bit what increased our incertitude: meanwhile we let the Sun's rays to come through a small hole into a black room on a white sheet of paper and we have seen this spot in the form of an elongated cloud; during the next three days, bad weather prevented us from observations. After this time, we have seen again the spot which proceeded obliquely toward West; we have seen another much smaller one at the border of the Sun which reached in a few days the centre; a third followed them, then the first one disappeared and the other followed it some days later. I wobbled between hope and fear not to see them again. Some ten days later, however, the first appeared to the East: I realized that it made a revolution and since the beginning of the year, I have been confirmed in this idea and I have showed them to other people who accepted them. Meanwhile I had a problem which prevented me from writing on them and which made me nearly repent the time which I spent with their observation: namely I found that the spots did not conserve the distance between them, their form and speed have changed, too, but my pleasure was the greater when I found the reason. Namely it is likely based on these observations that these spots are on the body of the Sun which is spheric and solid, therefore their movement is to be slower and smaller toward the borders: we invite enthusiasts of physical facts to profit from this sketch we presented them: they would surely suppose that the Sun rotates as said by Jordanus Bruno and recently by Kepler in his book on the movement of Mars as without it I don't know what we could do with these spots. I don't think that they are clouds; similarly I don't agree with those who put the comets into the Sun as emissaries which should return there soon: I prefer to keep silence than* *to guess; but I am tempted to consider this solar movement as the cause for other celestial movements following Aristotle who told in one of his poems "that the Sun is the father and author of movements"."*

The knowledge of the enthusiastic Fabricius on the sunspot must be certainly noted as a first step in this field. Galileo and Father Scheiner were able on many observations to give more precision to the nature and motions of these spots. Morever, both claimed the newness of the discovery.

Galileo in his "Discours sur la comète de 1618" in "Saggiatore o Trutinator" and in "Dialogues du systeme du monde" has always claimed that he was the first to see the sunspot in July-August 1616 and that Father Scheiner had began to observe them after he had seen Galileo's papers. This fact was refuted by Father Scheiner in his book "Rosa Ursina, 1630".

Whatever it might be, Galileo in his book "Istoria dimostragion interno alle macchine solari (Rome, 1613)", in which he presented for the first time the sunspots, showed that they are located on the solar surface, that sometimes there are a lot of spots and sometimes few or no spot, that the sun in its turning motion brings back them to our view, some of them last one or two days, others more than one solar rotation; they can merge or divide, they run on parallel circles; they never move more than thirty degrees from the equator.

Galileo also discussed the solar pole rotation, but he had not noticed the seven degree tilting of the solar equator to the ecliptic plan.

"*Since that time a set of detailed observations was just missing to the theory of the sunspots to confirm the rotation of the sun and to precise the location of its equator*" (Lalande 1778). This was done by Father Scheiner who published in "Rosa Ursina" the results of 2000 observations made during eighteen years.

"*If Galileo is the first one to understand the nature and motion of sunspots, Father Scheiner is the one who had most observed them and best studied their motions*" (Lalande 1778). For this purpose he added to his telescope a projection mechanism and an equatorial mounting following the invention of his colleagues (Father Gruenberger).

The English mathematician and astronomer Thomas Harriot (1560–1621), also observed sunspots from December 8, 1610 to January 18, 1613. In his manuscripts found at the end of the 18th century in the Castle of the Sussex earldom, there were hundred drawings of the solar surface whose copies were sent to Wolf by Carrington in 1857. Harriot is the sole astronomer of this epoch who determined with accuracy the duration of the synodic solar rotation, the mean value of his measurements is 27.154 days for two years of observations (Shirley 1974).

Finally, still for the beginning of the 17th century, we must notice the solar observations of Trade, Canon of Sarlat made during five consecutive years from 1615 to 1620, for all periods, for which the sky conditions allowed it. For this religious clergyman, sunspots can only be small planets with trajectories near the sun, and he explained this as follow: *We begin to tell here something ... mainly as we have seen several Italians or Germans who having seen these phenomena have told and maintained (be any the respective merits of philosophy and astrology) that these are spots on the body of the Sun imposing so on the father of the light as if the* *eye of the world would be sick of ophtalmy. However, it must not be supposed that the Sun be such dishonest and injured, we started its protection by trying to show here that they are neither spots nor cavities, any signs of other defects, nor clouds flittering before the Sun: but they are planets which have orbits around the Sun and which we called the astres of Borbon, after the very Christian king of France, Louis de Borbon ... I have been confused and kept secret for three years without permitting it to see the light considering it to be a difficult and dangerous topic which could produce in the public something new ...*"

Some of these observations have been published in 1622, and particularly those obtained on August 25, 1615 for which he observed thirty sunspots on the solar disk.

The sole other sunspots observations made towards the middle of the 17th century were those made by the Polish astronomer Hevelius in Danzig from 1642 to 1644. The apparatus used by him is presented on Fig. 2. These observations were published in his book Selenographia, in 1647.

Results

From these first sun observations made with the telescope at the beginning of the 17th century, it is possible to estimate the sunspot activity level which was existing during this epoch.

Galileo in his book, published in 1613 presented several drawings of the solar disk with the number of observed sunspots during June and July 1612 (Fig. 1).

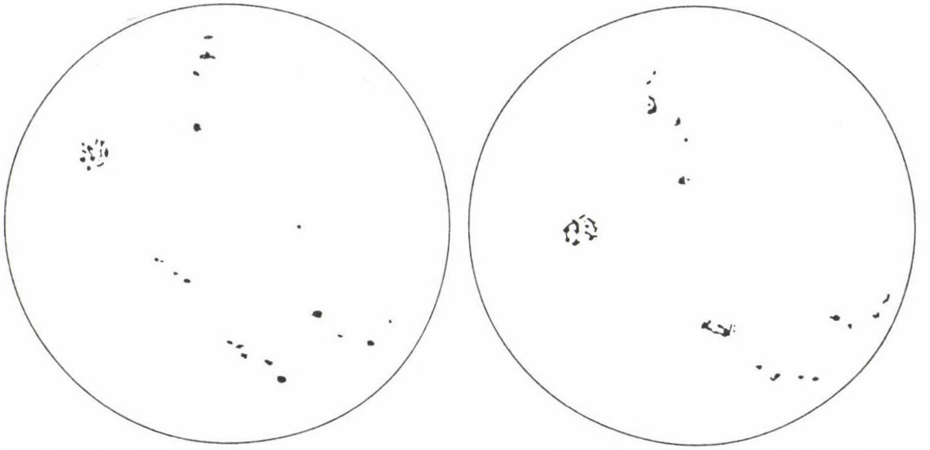


Fig. 1. Sunspot observed on July 5 (left) and July 6 1612 (right) by Galileo (Galileo 1613)

Sakurai (1980) has analysed these drawings and computed for each day of this period the relative Wolf number (R). He found daily values for R between 60 and 153. For thirty eight observational days, the R mean value is 99, this number corresponds to a relative high activity. This value is a minimum, if we take into account that some of the little spots must have escaped to Galileo. The analysis of the spot distribution (between +/ − 30° and +/ − 5°) as well as their structure indicate that Galileo made his observations during the solar cycle maximum.

In 1776, Lalande, in order to study precisely the duration of the solar rotation, analyzed again Father Scheiner's observations made between 1624 and 1627 and Hevelius's ones made from 1642 to 1644. So he could graphically estimate, by using the precise drawings made by these two astronomers (Fig. 2), the location of 176 sunspots and determine the date of their passage at the central meridian of the sun.

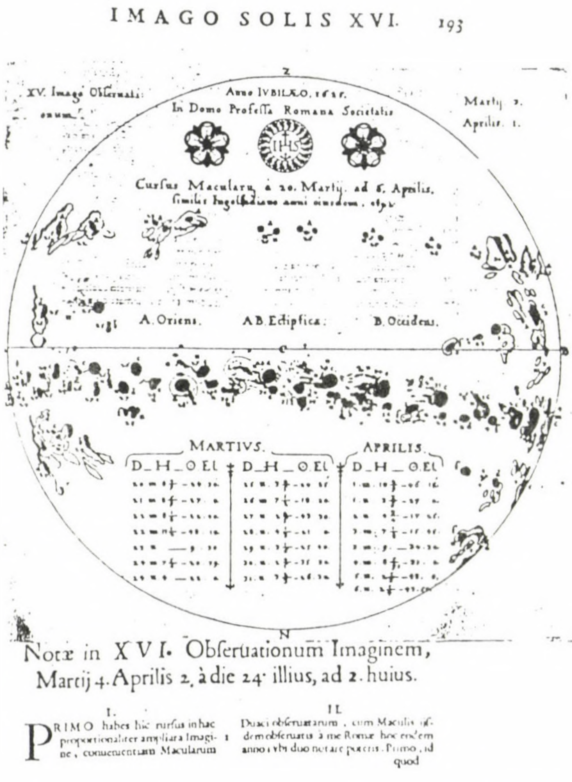


Fig. 2. Drawing of sunspot crossing from March 20 to April 6, by Father Scheiner (Rosa Ursina 1630)

Father Scheiner as Hevelius, have made mainly accurate drawings of the passage of the spot on the solar disk and rarely drawings of the solar disk aspect on a specific day. It is therefore difficult to estimate the Wolf number for these two periods.

For comparison with the sunspot activity which occurred at the end of the 17th century, we have computed the yearly mean value observed crossing spots. Between December 18, 1624 and November 28, 1627, Father Scheiner had observed 123 spot passages, *i.e.* 41 crossings for a year. The latitudinal distribution of these spots was between $+/-24°$ and $0°$. In Hevelius's book (1647), we found 49 spot passages between November 2, 1642 and October 9, 1649, *i.e.* 25 crossings for a year with a latitudinal distribution between $+/-18°$ and $0°$.

We actually know that the two distributions in latitude correspond to those found during the decreasing phase of the sunspot cycle.

By another way, we have looked if there were existing preferential solar longitudes for the spot appearance by computing for each of these 176 spot crossings the Carrington coordinates. Figure 11 shows the curve of the computed frequencies for the 10° classes of longitude. This curve have been smoothed by sliding average over three consecutive points.

One can remark two preferential areas, at the beginning (close to 45°) and at the end (close to 250°) of the Carrington rotation. This result has been also found by Trellis (1971) for the solar cycles 12 to 19. The too small number of data (3) for the analysed period prohibits to continue this kind of comparison but one can suppose that these preferential areas are permanent features.

The data analysis of this period 1610–1645 does not allow to present a homogeneous sunspot activity for all the observations made by the solar astronomy pioneers. On the other hand, one can conclude that the sunspot appearance was not a rare phenomenon and that activity would develop in the same fashion as for known solar cycles observed since 1850.

The period of "uncertainty" 1646–1670

For this period, it is difficult to found sunspot observations as the sun was not yet systematically observed by big observatories. It is necessary, as some authors recommended it (Eddy 1976), to conclude that the sunspot activity was very weak?

It is impossible to answer such a question without any previous knowledge on the frequency of sun observations, as we only dispose of two evidences:

— one from the French astronomer Picard who having seen a sunspot during August 1671 said: *"Since ten years I had not seen any of them in spite of efforts to look for them from time to time"*

— the other from Oldenburg, the Royal Society secretary, author of paper in Philosophical transaction, dated August 14, 1671. Oldenburg concerning the same sunspot, also seen by Cassini, noticed that no sunspot had been observed since the one seen with an excellent telescope by the English philosopher Boy, on April 27, 1660 at eight o'clock in the morning, in the presence of Hook.

If these two evidences plead in the favor of a weak solar activity, it is impossible to estimate sunspot activity. The few number of visible sunspots observed in China during this period (Zhen-Tao 1982), including some ambiguous events, do not change our conclusion.

The period of weak activity

Historic

It is the sole period of the 17th century where it is really possible to know the frequency of aurora appearance. Indeed, the construction of the Royal observatories, particularly those of Paris and Greenwich, shortly after the creation of academic assemblies (Sciences and Royal Society academies), contribute to the advance of astronomy and consequently to the development of systematic observations.

In 1669, in his program report before the Academy, Picard indicated: *"It would be necessary to continue with special attention the observations of the diameters of the Sun and of the Moon to discover some irregularities which occur mainly in the observations of the Sun, as during the last solstice e.g. the Sun appeared to be slightly greater than it used to be. And now it is smaller by 4 or 5 seconds as it didn't used to be for one year. There is no other way to detect the yearly path than to observe the diameter during a whole year and to deduce from it the distance of the Sun as being proportional with it, the path is very near to a circle but it can be distributed by small irregularities which can be found only by long series of observations. Such observations are to be made daily ..."*

In the same way, Cassini, in his recommendations to those who are working at the observatory, *i.e.* his students, noticed that: *"The Sun is to be observed each day at sunrise and sunset, to measure its apparent diameter, then to follow these observations after sunrise and before sunset till the corresponding heights."*

Piccard's observations, those of La Hire (who succeeded Picard, who died in 1682) and those of Cassini and his students are recorded in a series of registers kept in the Paris Observatory library. A copy of Cassini's observations, written during the 17th century, is also at the Institut de France library.

In Picard's and La Hire's registers, the meridian of the superior and inferior sun side, as well as the time of the passage of the oriental and occidental sides in the meridian plane are mentioned.

These observations made each time meteorologic conditions were appropriate, allowed to control the diurnal clock running, and to determine with precision the ecliptic parameters.

Several measurements were made on the sunspots which appeared on the solar disk. Since the beginning of Picard's program, the sun was the object of assiduous observation and the unusualness of sunspots was so noticed that each appearance was the object of a report at the Sciences Academy, or a note in the Journal des Savants.

After 1704, the increase of the sunspot activity was so important that the academy had published after that time only one annual report until 1716.

Besides, it is important to precise that sunspots observed from 1672 to 1704 appeared exclusively in the southern solar hemisphere and near the equator (rarely at a latitude greater than 10° south (Maraldi 1695)).

Later on September 27, 1707, when Cassini and Maraldi have observed a spot at the northern latitude of 10°, they noticed that they never remembered to have seen one spot in this hemisphere before this date, excepted for the event of April 1705 (Spoerer 1889, Maunder 1922).

In a same way, the existence of two distinct packs of sunspots on the solar disk was very scarce at this epoch, since, talking of the spot observed in October 1705, it rends in the history of Academy for this year: *"Since the observations by Father Scheiner some 80 years ago, one hadn't seen simultaneously two different groups of spots. We remarked in the History of 1700 how seldom this phenomenon had appeared, meanwhile it is the second time it has appeared within two years."*

It is necessary to say that the few number of observed spots during thirty years, and after 1702, the large number of spots, lead the astronomer of the beginning of the 18th century completely confused. We found the proof of that, in the history of Academy of 1713, where it is noticed: *"The time of the appearance of spots on the Sun is by no means regular. Ever since 1695 till 1700 one had not seen any of them. Afterwards, our Histories are full of them till 1710, when only one was observed and it seemed that they approximated their end. In 1711 and 1712, no spots were seen, and in 1713, only one in May ..."*

Results

The diagramms on Fig. 12 show the frequency of the daily sunspot observation on the solar disk (b). One can see that excepted for the time interval from 1677 to 1682, during which astronomers were on special mission for the king, few periods of thirteen days (duration of a spot crossing on the solar disk), are without any observations.

On the spot diagram (b), the Paris astronomers' data complete the European observations made by Kirch and his family (Landsberg 1980), Ettmüller, Liebknecht and Durham (Schröder 1990).

The analysis of this figure shows that few spots were appearing between 1671 and 1701, and that between 1690 and 1699, a single spot was seen on the sun in May 1695, the total number of days of spot observation by decade and the estimation of birth of crossing of spots on the solar disk can be summarized as follows:

years	days of observations	spot crossings/year
1671–1680	110	1.8
1681–1690	98	2.1
1691–1700	10	0.3
1701–1710	483	8.1

Compared to the values obtained by Father Scheiner (49 spot crossings/year) and Hevelius (25 spot crossings/year), the solar activity has been very weak at the end of the 17th century, particularly between 1691 and 1700.

Remarks on the Wolf number concerning the 17th century

In the empirical formula $R = K(10g + f)$ established by Wolf to determine the relative number of sunspots, the value of the K coefficient is a function of the apparatus used by the observer. Wolf has given a value of 1 for this coefficient for the telescope used by him which had an objective of 75 mm — diameter and a magnification of 64. For a more powerful tool, K is less than unit, it is greater in the opposite case. Wolf took also into account in the determination of K the regularity of the observer (Young 1883).

In these conditions, it appears to be difficult to use this formula as Wolf did it himself to compute the relative spot numbers before 1671 for the following reasons:

— the characteristics of the apparatus used at this epoch are not very well known and sometimes unknown

— one cannot know the annual days of sun observation before 1671, *i.e.* the assiduity of the observers which depends mainly on meteorological and also on personal factors.

The following table presents some relative annual numbers of sunspots computed by Wolf (Waldmeier 1961), compared to the spot crossings counted by Lalande (1776):

Year	R	Spot crossing	sun-observation days known
1625	41	88	?
1626	40	27	?
1627	22	7	?
1643	16	21	?
1644	15	26	?
1684	11	4	259
1642	6	2	?
1695	6	1	200

The examination of this table reveals the incoherency of these results in lack of any other complementary data which could well inform us on the value of the K coefficient given by Wolf.

For example there is a ratio of 7 between the R numbers for the year 1625 and the year 1695, while this ratio reaches 88 between the number of spot crossings.

It must be noticed that Sakurai's computations previously quoted for the determination of the sunspot activity of June and July 1612, seem to be more reliable as they are obtained from a data set of thirty eight days covering a period of two months. In this case, it is mainly the instrumental effect which can introduce uncertainty in the results.

Technical means and meteorological observation conditions at the end of the 17th century

The apparatus

The telescopes used at the beginning of the 17th century had a lack of performance due to the technical insufficiency of glassmakers and opticians, to which were added the natural aberration of the thin lens. Towards 1650 under the influence of Huygens, the astronomical optic performed evolution to a better quality.

Indeed, experiments had shown that one could improve the image from a telescope by increasing the focal distance of the objective: with constant diameter and enlargement, the sphericity aberration was reduced and simultaneously the angular diameter of the chromatic diffusion spot was diminished *"In a time when the efficiency of the objectives was so increased, the limits of the length of the telescope and of the discovery of the sky were not seen, and it was supposed that it would be possible to see animals on the moon"* (Auzout, cited by King 1979).

And that's why longer and longer telescopes were built, the focal distance of their objectives ($d = 24$ cm) could reach 68 m (Danjon and Couder 1935). These tools became very clumsy and difficult to handle. — With such an equipment Huygens discovered in 1655 a Saturnian satellite (Titan) and recognized the real structure of the ring of this planet, which had so intrigued Galileo. In 1664, Cassini observed, for the first time the shadow of the Joian satellites on the planet surface, in 1675 he perceived the distorsion of the Saturnian rings and between 1671 and 1684, he discovered four new satellites of the planet (Pingré 1901).

These different discoveries prove the quality of the tools used during the second part of the 17th century. To observe the sun, Picard and La Hire used a "quarter circle" and a wall quadrant including a telescope with a reticle whose focal distance was about five feet (1.60 m) (Bigourdan 1928). As Father Scheiner did, they probably used a thin plate colored glass to preserve eyes. In a same way, Cassini used a large quadrant having a telescope of 6 feet (1.95 m) focal distance, but he was sometime doing observations with more powerful tools in order to analyse the sunspot details or to see better small spots. Thus his observation note book includes some indications like:

— July 28, 1684: *"At 6 h according to the image received on the drum of the great telescope of 40 feet, may be due to the smallness and weakness of the image of the spot it could be influenced by the wind which moved the telescope, the spot was elongated six parts from the rim and passed at a distance of 31/2 (parts) from the centre."*

— July 29, 1684: *"the spot of the sun cannot be seen even by the telescope of 20 feet"*

— September 30, 1688: *"I observed two spots in the sun which did not appear neither in the morning, nor at noon."*

These particular observations refute Kopecky and Kuklin's remarks (1987). Kopecky and Kuklin state that long-lived visible sunspots were the sole observed, while the others were purely neglected.

This bright period of large tools, during which numerous discoveries have been made in astronomy-physics was followed by a long decline, because after Cassini's death, in 1722, the large telescope were forsaken on behalf of the quadrant.

"Fifty years later (1760), wrote Fabry, *the technique of these instruments was completely lost and the astronomers were unable to see several astres which had*

been discovered by Cassini; even the memory of these discoveries was nearly lost. The invention of the achromatic telescopes by Dollond (1758) in England did not modify this situation as the first instruments of this type being very perfect were too small to show what could be seen through the instruments of the 17th century: but these telescopes contained the germ of a progress which only brought his fruits much later. It is true that with the reflexion telescope which had been known for a long time but which were not used seriously, Herschel (1789) renewed the interrupted tradition of the great observers of the first years of the observatory (in Paris)." (Danjon and Couder 1935).

The solar observations did not escape this decline, as Lalande noticed at the beginning of his long memorandum on the subject in 1776: *"Since more then 60 years next to nothing has been made about the spots of the Sun about its rotation and about the situation of its equator and of its poles."*

It is obvious that the astronomical tools at the end of the 17th century were enough accurate so that the reliability of the sun observation could not be challanged and that the astronomers were observing with attention the evolution of the spot activity of the "star of the day".

The meteorological conditions

The 17th century, which had particularly variable meteorological conditions, corresponds partly to the period named "little ice-age" whose duration is mainly defined by the expansion of Alpine ice (see Appendix 1).

The cloudy covering during this epoch is sufficiently well known only by the forty years of meteorological records of Louis Morin (see Note 4 in Appendix 2). These records made at Paris, three times a day between 1675 and 1713, concerning temperature, pressure, direction of clouds and nebulosity (Legrand and Le Goff 1987) allow us to show the weak influence of climate on the sunspot observations, contrarily to the suppositions made by Kopecky and Kuklin (1987).

In his meteorological note book, Morin characterized the nebulosity by an index which was changing from 0, for a clear day, to 4 for a complete cloudy covered day.

For each day of solar observation at the Paris observatory, we know the indices for the cloudy covering. We have computed, for each class of the index, the mean percentage of the number of observation days as a function of the total number of days of the class from 1682 to 1709.

index 4: complete cloudy covering	8 % of observations for 44 days/year
index 3	34 % of observations for 55 days/year
indices 1 and 2	71 % of observations for 219 days/year
index 0: clear day	97 % of observations for 47 days/year

The general mean value for the whole set of 28 years is 61 %, it is comparable to the mean value obtained at Meudon for the years between 1930 and 1940 (Martres 1987).

Louis Morin's records show a clear increase of the cloudy covered days (with index 4) during the decade 1690–1699: 54 days/year with a maximum of 83 in

1692. The number of clear days remains constant. The increase of the number of cloudy covered days, induced by the frequent passage of frontal systems and the persistency of an anticyclone over the North sea (Legrand et al. 1990) had no repercussion on the number of sun observation.

One can conclude that during the last half part of the 17th century and until the end of the Sun King's reign, the sun was carefully observed, each time the weather allowed it. One can affirm that the cloudy cover of the coldest period of this epoch (1690–1700) is by no means the cause of the lack of observation of sunspots.

The auroral activity during the 17th century

The solar sources of the magnetospheric activity and the conditions of aurorae appearances

The solar sources

The analysis of the hundred and twenty years of geomagnetic data made by Legrand and Simon (1989, 1990) and Simon and Legrand (1989, 1990) shows that two solar sources of magnetospheric activity (*i.e.* auroral and ionospheric) are existing:

a) transient events associated to the sunspot cycle which generate shock waves

b) solar wind jet streams with a speed greater than 450 km/s.

The distribution of solar wind speed is determined by the intensity and the direction of the solar dipole whose cycle is related to the sunspot cycle, but with an opposite phase. Faster winds escape from the dipole poles and slower winds from the "neutral sheet" located at the equator of the dipole. Due to this fact the Earth is sweeped by speedy winds when the dipole reaches its maximum intensity and is a little tilted on the solar rotation axis, or when the dipole is turning during the maximum phase of sunspot cycle; this evolution is indicated on Fig. 3.

Nevertheless, for a weak amplitude of the solar cycle ($R_{\max}40$), at the maximum intensity of the dipole, the fast winds of polar origin never reach the earth and the "neutral sheet" is thick. During this phase, the geomagnetic and auroral activities are weak as the earth is swept by slow winds. The situation is not the same at the time of the dipole turning, *i.e.* during the multipolar phase, when relatively faster winds from a region contiguous to the neutral sheet can reach the Earth.

The conditions of aurora appearance at middle and low latitudes

To observe aurorae at middle and low latitudes, without any confusion, it is necessary that the boreal auroral oval extends towards the south of its mean position located around 66° geomagnetic latitude.

In general an aurora occurs in the altitude range from 100 to 400 km and the maximum intensity of the red line (6303 A°) is located around 300 km. Under the diagram of Fig. 13, the aurorae can be visible at 10° below the horizon at a distance

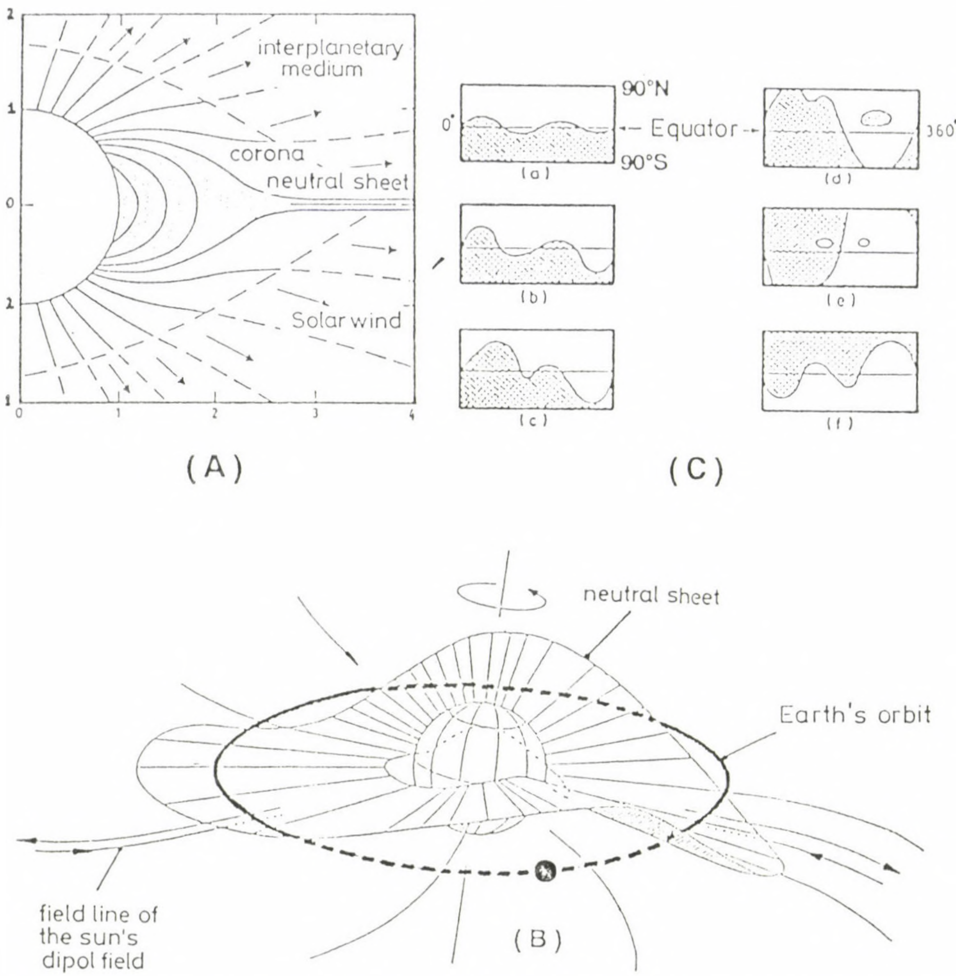


Fig. 3. "Structure and evolution of the solar dipole, the neutral sheet". (A) Theoretical model of the solar dipole field distorted by the solar wind flux: the discontinuous line schematizes the passage of the plasma from the solar crown to the interplanetary environment. The plasma trapped by the coronal field in the equatorial zone (grey surfaces) can only escape through the "neutral sheet" which is formed between the two magnetic sheets of opposed polarities. (B) Representation of the solar dipole "neutral sheet" at the level of the terrestrial orbit (T): in this diagram, the earth, during a solar rotation (2+3 days) crosses four times the "neutral sheet" source of the slow speed plasma. Between each of these crossings it is swept by solar wind of higher speed coming from the regions contiguous to the "neutral sheet". (C) Cycle of the solar dipole field established according to the photospheric data from 1963 to 1984; each rectangle corresponds to a different period of the eleven years solar cycle. The solar latitudes are in ordinate and the Carrington longitude in abscissae. The white and grey areas respectively correspond to North and South polarities. They are separated by a line showing "the neutral sheet" shape (a) and (b) show a phase during which the coronal field configuration is dominated by the solar dipole (c), (d) and (e) show the reversal stages of the "neutral sheet" occurring during the change of dipole polarity. The "neutral sheet" form is the signature of this multipolar phase. The "neutral sheet" goes up to high latitudes, and as a consequence fast wind sources are brought in the equatorial plane (f) what corresponds at the turn to a dipolar configuration but with opposite polarities

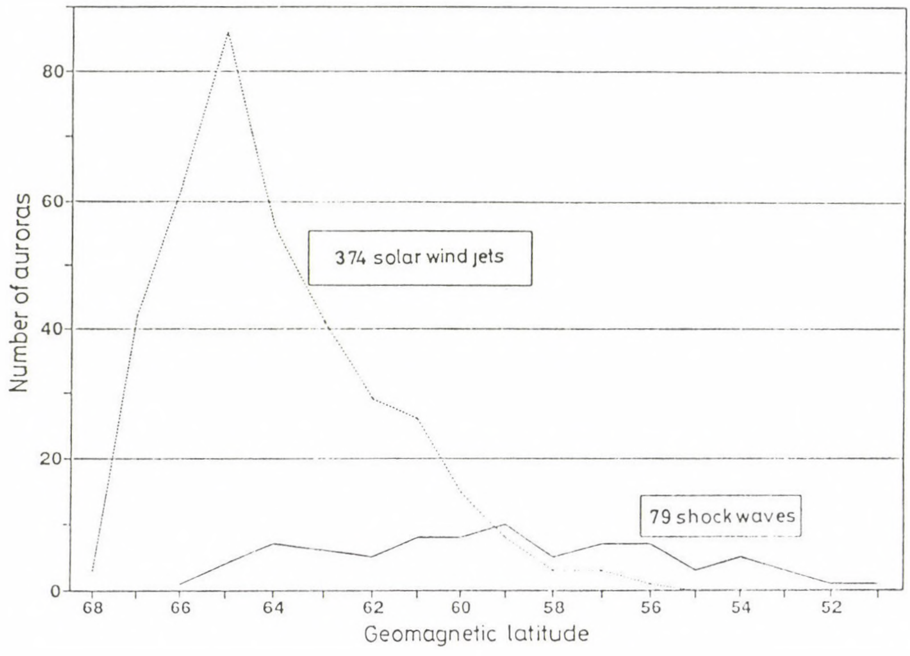


Fig. 4. Distribution of aurorae produced by solar wind jets and by shock waves from 1956 to 1965. These aurorae are observed at the zenith of British stations distributed from Calais to Faroe Island

of 1100 km, *i.e.* at a latitude 10° less than that where the phenomenon takes place. So, aurorae observed at the limit of the visibility at the geographic latitude of 50°, develop at a latitude of 60° (Legrand 1990).

Legrand and Simon (1988, 1989) studied the latitudinal distribution of 453 aurorae observed between 1956 and 1965 at the zenith of several British stations stretching from Calais to Faroe Island. Figure 4 shows the frequency of aurorae appearances as a function of geomagnetic latitude (F) according to their mechanism of generation: shock waves or solar wind jets.

This distribution has been derived from aurorae appeared during a particularly strong sunspot solar cycle (1956–1965). As Stringer and Belon (1967) showed that the location of the maximum of auroral activity does not change in the case of a much weaker sunspot activity, one can suppose that aurorae observed in the 17th century presented similar distribution.

It is necessary to know that the position of the geomagnetic pole can change during centuries, and that precisely in the 17th century, the location of the northern magnetic pole was 4° closer to Europe than is now (Champion 1980). One can also verify in Fig. 5 that aurorae produced by solar wind jets were more visible in occidental and central Europe. The geomagnetic latitudes were at this location 4° higher than now.

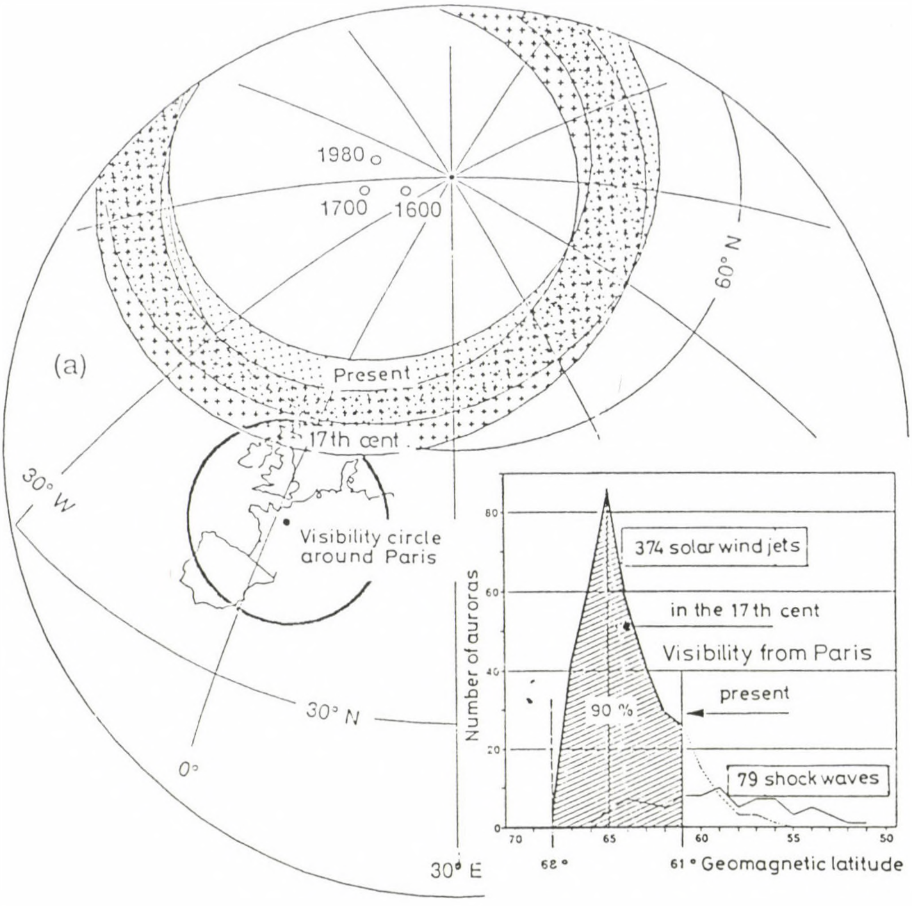


Fig. 5. Position of the auroral zone related to the solar wind jets, with regard to a station located at 50° latitude North during the 13th and 20th centuries

Aurorae observations during the 17th century

Since a very long time, large aurorae developing at middle and low latitudes were known. They have been described, especially in Europe and China where the civilization has early developed, by philosophers and erudits in chronicles which were published after invention of printing in the 15th century. The description always included fantastic expressions mixed to superstitious concepts as this phenomenon was considered by people of this epoch as an ominous sign which announced war, scarcity or other flail.

The first scientific description of an aurora was published by Conrad Gessner (1516–1563) after the aurora appearance on January 6, 1651. It contains details of the colour fluctuations, of brightness and of its structures: arcs, rays, diffuse spots, crowns.

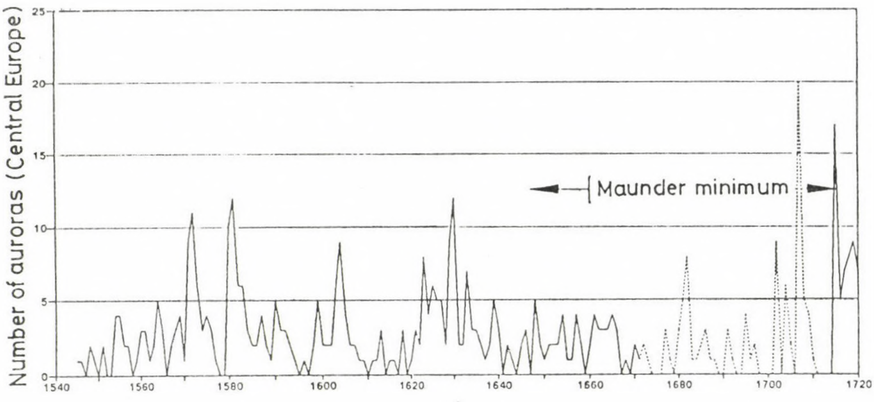


Fig. 6. Curves of annual frequency of aurora observations in Central Europe: the dotted part of the curve represents the period of the "Maunder minimum" during which the sunspot activity is known with greatest precision

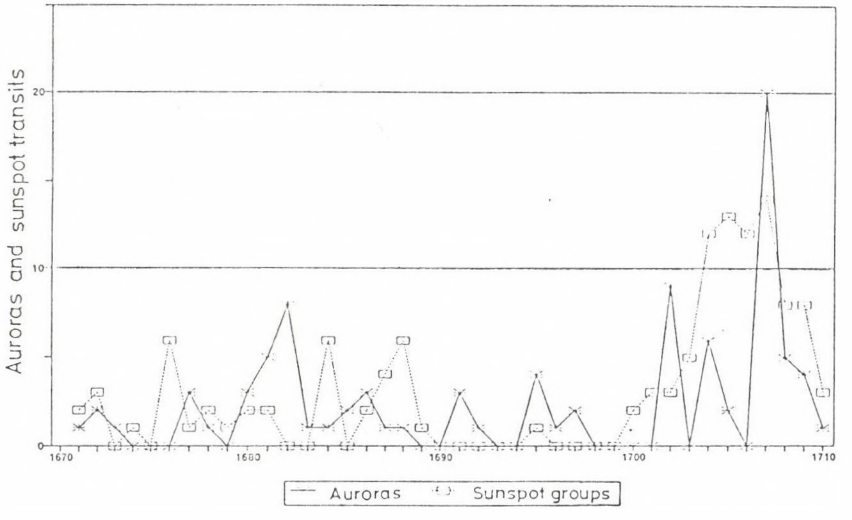


Fig. 7. Recapitulative curve of the auroral activity (total annual number) and of sunspot activity (total annual number of observations on 13 consecutive days) during forty years for which these two phenomena are precisely dated

During the 16th and 17th centuries, a substantial number of observations of this phenomenon has been reported in chronicles of this epoch. But wars, as the "Thirty years" war (1618–1648) or revolts brought their share of ruins in Central Europe and led to the distruction of precious documents. Successfully some works outlived, for example the journals of the astronomers David (1564–1616) and Johan (1587–1615) Fabricius and also of the Kirch family (between 1677 and 1774) contain a considerable number of auroral data among descriptions of the solar surface. More generally the ancient chronicles and the observation note books of erudits have been searched by several authors as: De Mairan, Frobesius, Schöning at the 18th century, Fritz at the 19th century, and Link, Keimatsu, Schröder among the more recent ones, to establish catalogs for which some data date back several centuries B.C. Some of these catalogs include not only the dates of aurorae appearances but also those of comets, meteors and atmospheric phenomena. The exploitation of these data request a considerable work as each description of these events must be examined carefully to identify the uncertainties.

Concerning Europe, the observations were generally made from countries located between 48° and 54° geographic latitudes: Germany, Austria, Spain, Hungary, Switzerland. Few auroral data are from England and Russia. The observation were generally irregular, but remarkable auroral events were always recorded except when the sky was completly cloudy.

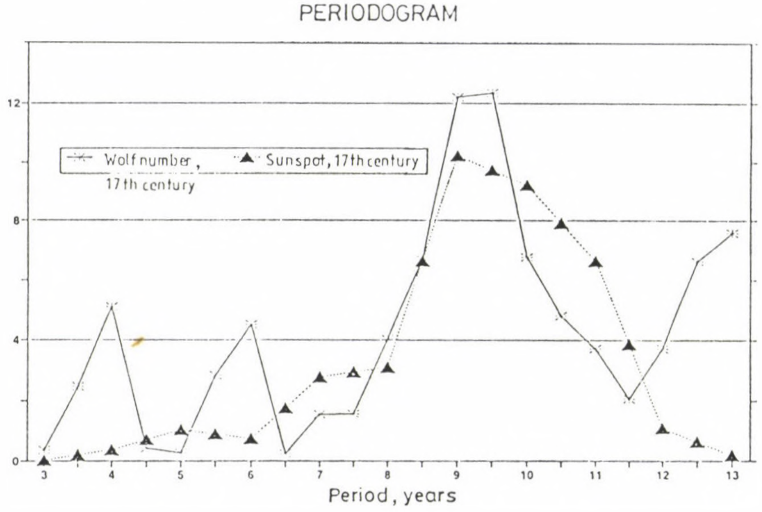


Fig. 8. Periodogram made with the series of spot crossings from the end of the 17th century (1671–1710, continuous line) and with the series of Wolf numbers during the 18th century (1744–1783, dashed line). The maximum occurs during a period of 9–10 years for these two series of same duration

Between 1550 and 1710, approximatively 430 aurorae visible from Europe have been recorded, 150 between 1645 and 1710, the period corresponding to the Maunder minimum.

To appreciate the homogeneity of the data collected during the 16th and 17th centuries, we have examined the seasonal variation of the reported aurorae which must follow the geomagnetic activity whose maxima appear at equinox and minima at solstices.

On Fig. 14 we have drawn this variation for the two distinct periods studied by Schröder:

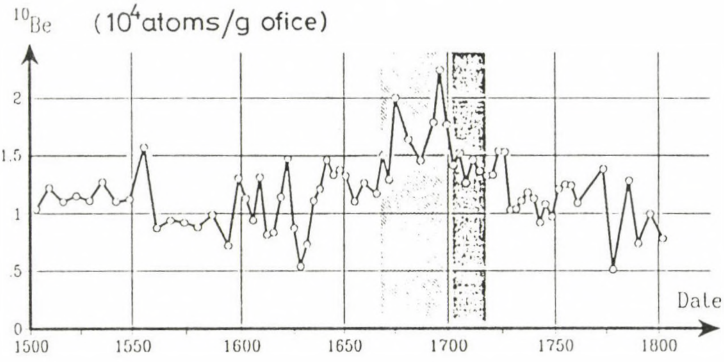


Fig. 9. Be10 concentration between 1500 and 1800 (Beer et al. 1983)

a) before the Maunder minimum, from 1545 and 1644 and

b) more recently from 1822 to 1956.

June and December minima clearly appear for the data set (b) while only the summer solstice minimum can be identified for the data set (a).

This summer minimum for the concerned latitudes, 48°N to 54°N, is mainly produced by the crepuscular effect due to the small height of the sun below the horizon. This summer minimum does not inform about the degree of homogeneity of ancient data. On the contrary, the absence of a winter solstice minimum and of pronounced equinox maxima, lead to the supposition that the available data are only an incomplete sample of the auroral activity.

The frequency of aurorae appearances between 1545 and 1728 presented on Fig. 6 (Schröder 1990) shows in spite of inhomogeneity of the data a deficit of aurorae between 1640 and 1702 characterized by the total absence of important maxima.

Periodicity of spots and aurorae appearances

The data set concerning the best documented sunspot activity period from 1671 to 1710, contains 91 days of aurorae observations and 561 days of sunspot observations; the low number of events representing a period somewhat shorter than four mean solar cycles yields uncertain the research of periodicity. Nevertheless we have computed the periodograms of the two series classified following half-yearly values. We have tested the stability of the periodograms by computations on the raw data, then corrected for the trend by second order regression, and then smoothed (or not) by a sliding mean over three ponts.

The two series are presented on Fig. 7 (annual values):

a) aurorae: attempts were unfruitful due to the instability of the results depending on the statistical data processing. Moreover, each updating of the data set added instability. This fact leads to confirm the inhomogeneity of the aurorae set previously noticed in the section on magnetospheric-solar activity.

b) the spots: For this data set, we have estimated the number of thirteen days periods (half solar rotation) for which a group of spots have been observed. By this mean we give the same importance to a spot observed during all the time of its passage on the solar disk as to an other spot which could not have benefit of the same time fo observation. The total number of groups is 123 with a half yearly maximum of 8. In this case the periodogram (see Fig. 8) is more stable and shows a maximum around 9 years. As a comparison we have plotted on the same figure the periodogram made with the Wolf numbers for a similarly 40 years period (1644–1683). With fifty more time data, the regularity is remarkable and we found in this case a series of four short cycles of 9 to 10 years.

One can conclude, prudently, that at the end of the 17th century, the weak sunspot activity appeared as several short cycles (9–10 years) for which it is impossible to precise the dates of maxima and that the auroral activity, as it is actually known, does not present any correlation with the sunspot activity.

The relation between the spot passages and the aurora appearances

Legrand et al. (1990) have studied the coincidence between the spot passages on the solar disk and the aurora appearances in Europe, in order to analyse the evolution of the solar activity during the 17th century and particularly during the period of weak sunspot activity (1671–1701).

One knows that a shock wave is generated when in an active region of the sun an important atmospheric eruption occurs associated to a jump in the X and radio radiations domains.

This wave reaches the Earth with a delay of fifty hours and produces a magnetospheric storm materialized by an aurora. This phenomenon is the more intense the eruption develops the nearer to the solar central meridian between 30°E and 30°W (Legrand 1990). The comparison of the spot crossing and the appearance of one or several aurorae in Europe between 50° and 55° latitudes, can therefore inform about the sunspot activity existing at this epoch on the solar disk.

Figure 15 shows that among the 44 aurorae observed between 1671 and 1701 originating from the catalogs of Link (1964), Schröder (1979) and Landsberg (1980) no one coincides with a spot passage. On the contrary during the period from 1702 to 1709, among the 47 observed aurorae, 28 coincide with a spot crossing: 12 aurorae present an exact coincidence and 16, an estimated one, if we are supposing that the observed sunspots are active during a half solar rotation (13 days). The following table resumes the whole set of sunspot and aurorae observations during the two periods: 1671–1701 and 1702–1710.

If the aurorae would appear at random, the probability of coincidence between one aurora (A: number of days with aurorae) and one spot (T: number of days with spots), Pr (AT/A), should be equal to the probability to observe one spot (Pr/T/N), N being the total number of days.

Period	1671–1701	1702–1710
Total number of days (N)	11315	3285
Days of sun observations (N1)	4098	1892
Days of spot observations	233	418
Estimated number of days with spots (T) (long-lived spot 13 days)	585	1014
Days of aurora observations (A)	44	47
Number of coincidences sunspot/aurora (AT) (including the propagation time)	0	28
Probability of sunspot activity Pr (T/N) population N, threshold = 95 %	4 to 6 %	28 to 32 %
Probability of coincidence Pr (AT/A) population A, threshold = 95 %	0 to 9 %	45 to 74 %

The computation of intervals with a probability of 95 % for T/N and AT/A (binomial law, Morice and Chartier 1954) shows that:

— between 1671 and 1701, there is no coincidence, therefore the hypotheses of independence between aurora and spot can be supported

— on the contrary, between 1702 and 1709, the probability of coincidence is significantly greater than the hazard and the hypothesis of a strong connection between spot crossing and an aurora appearance cannot be rejected, particularly after 1704.

It is also interesting to notice that the five aurorae of the year 1625, precisely dated (Link 1964), coincide with spot crossings observed by Father Scheiner. Knowing the date of these spot passages at the central meridian (Lalande 1778) and assuming a time for the shock wave transit of 48 hours, one can suppose that these five aurorae probably appear after chromospheric eruptions occurring between 0° and 30°W.

The absence of coincidence between spot crossing and aurora appearance from 1671 to 1701, leads to prove that only fast wind jets must exist during this period. This hypotheses in confimed by the analysis of the Be10 concentration of the polar ice (Beer et al. 1983).

The maximum of concentration appears between 1669 and 1700, this fact means that the intensity of the cosmical galactic radiation was large due to the absence of shock activity. Indeed, the shock activity induces a large very important modulation of the cosmic radiation compared to the modulation due to the high speed solar wind jets (Venkatesan et al. 1982, Legrand ans Simon 1985, 1989, Badrudinn et al. 1986).

The so-called "Maunder" minimum and the running of the solar machine

From the description of the sunspot activity and from the analysis previously presented we can remember the following facts:

a) during the period 1610–1645, the sunspot activity was relatively high and probably developed in cycles as those actually well known since their discovery by Schwabe in 1843. The curve of the Be10 concentration confirms it (Fig. 9) as well as the number of the eye-visible sunspots observed in China (Eddy 1983).

b) it is impossible to state anything for the period of "uncertainties" (1646–1670) due to the total absence of data on the number of days of sun observations. The comments of Picard and the paper published in Philosophical Transactions dated August 14, 1671 plead in favor of a very weak sunspot activity, but it is impossible to estimate it. On the other hand, the Be10 concentration for this period is similar to the Be10 concentration appeared from 1720 to 1730, which is relatively high.

c) all the spots which were observed between 1660 and 1704, were located in the southern hemisphere of the sun at a latitude close to the equator.

d) during the period 1671–1710, we have a data set which allows us to state that the sunspot activity was weak between 1671 and 1689 and almost inexistent betwen 1690 and 1700. Since 1702, active centres began again to appear on the solar disk.

e) between 1671 and 1710, a spectral analysis of the frequency of spot appearances seems to show the presence of a 9 year cycle as those for which the maximum occurred between 1769 and 1778.

f) We must recall that the speed of rotation of the sunspots between 1642 and 1710 presented variations (Ribes et al. 1987) and that the solar diameter seems to have a mean increase of 0.2 % with a periodic variation which also follows a 9 years cycle (Ribes et al. 1989).

All these results indicate that here was an anomaly in the running of the "solar machine" since a date which took place between 1646 and 1670, but it is impossible to precise it. This anomaly remained until 1702.

The auroral activity observed during this century cannot give any reliable information on the level of sunspot activity as the connection betwee these two phenomena, not very reliable, was turned off by aurorae due to solar wind jets. Indeed, at this time the position of the geomagnetic pole led to the visibility of such aurorae at Central-European latitudes.

On the other hand, the absence of correlation between sunspots and aurorae between 1671 and 1702 convinced us that the aurorae observed in Europe were only produced by solar wind jets and give information on the evolution of the solar

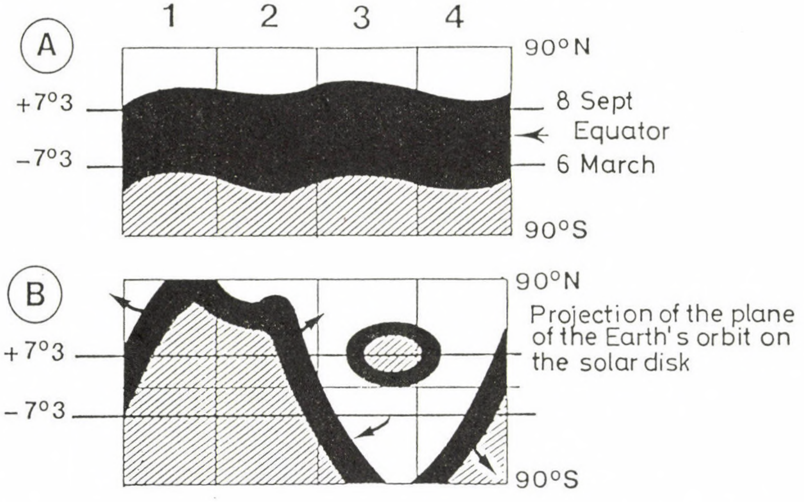


Fig. 10. Examples of the structure of the solar "neutral sheet" during the dipolar phase (A) and the multipolar phase (B). This example shows the extremum positions (March 6 and September 8) of the ecliptic plane with regard to the thick "neutral sheet"

magnetic fields: the toroidal component of the solar magnetic fields which represents the sunspot activity is closely related to the dipolar component which occurred 4 to 5 years earlier (Simon and Legrand 1989, Legrand and Simon 1990). One can conclude that the maximum of intensity of the dipolar field remained weak during the three solar cycles of the end of the 17th century. Therefore, the neutral sheet was thick (Fig. 10) and due to this fact the Earth was mainly sweeped by solar wind jets of low speed and not by fast jets originating from the poles as it is normally the case for this period of the solar cycle. The magnetic activity was probably very quiet and the auroral oval located at very high latitudes (around 74°) during several consecutive years for which no aurora was observed in Europe. This was the case from 1673 to 1676, from 1687 to 1690, and from 1698 to 1701.

It is at the time of the turning of the solar dipole polarity, *i.e.* during the phase of maximum of sunspots, and a little after that the Earth received solar wind jets of high speed coming from regions contiguous to the "neutral sheet"; at this period of the cycle this structure presents large undulations (Fig. 10). Probably these jets produced the aurorae observed at that time in Europe.

It seems perhaps audacious to propose such a scheme for the dipolar field evolution based only on the auroral data of middle latitudes (48° – 54°) which are available for this epoch. However the period of weak activity at the beginning of the 19th century, though it was a little more important than that occurring during the Sun King's epoch, allows us to maintain this hypothesis (see Appendix 2). As a matter of fact, between 1671 and 1700, three sunspot cycles with a periodicity probably near 9 years appeared successively. The structure of the spots which were only in the southern hemisphere, could only produce a very weak chromospheric activity.

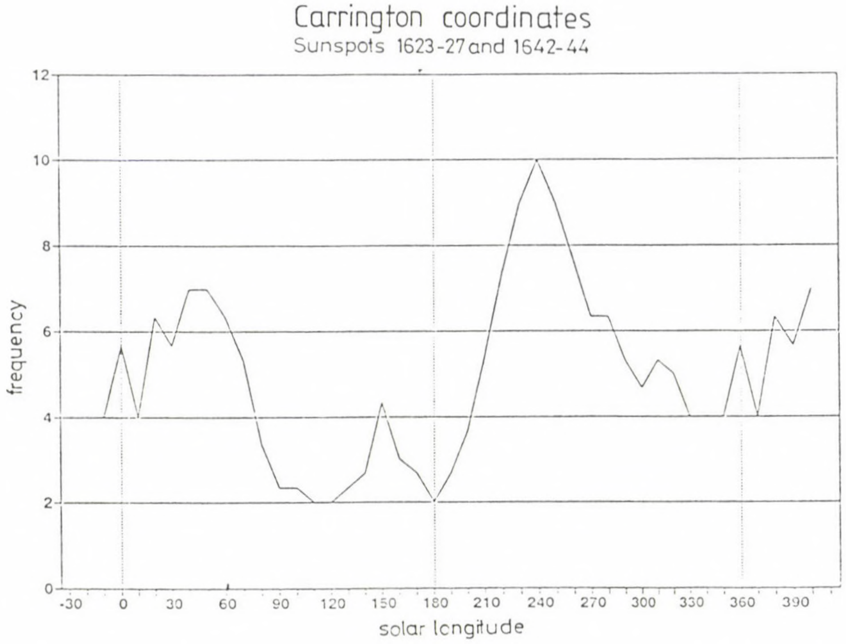


Fig. 11. Frequency of the spots as function of solar longitude in Carrington coordinates by Father Scheiner and Hevelius

After 1702, the spot activity, though weak, gave bright regions capable to emit shock waves and the "solar machine" took again progressively its normal rythm at the amazement of the astronomers of this epoch. The year 1701 underlined the end of this period of minimum of sunspot activity which ought to be named "Spoerer".

This period of low activity which lasted 30 years, perhaps 60 years, is the only one in the history of sun observations by telescope, and arises, precisely at the beginning of the development of the astronomy-physics. It is certain that this anomaly delayed the discovery of the spot cycle. So, even for the astronomers of the 17th century, working hard on numerous captivating studies (measure of the flattening of the earth at the poles, precise determination of the Sun-Earth distance, establishment of catalogs of stars, *etc.* ...) sunspots were a transitory phenomenon; their appearance does not have any rules, they were born and disappeared under the observer's eyes. These variations led the astronomers to confusion (Bailly 1782). "Don't stick to the spots, said Lalande to a young student, a phenomenon which has no law" (Deslandres 1906).

We know only of two other periods of weak sunspot activity, but with a cycle amplitude greater than the amplitude observed during the "Maunder minimum". One occurred from 1798 to 1823, and the other from 1878 to 1912. The astronomy was still limited to optical observations in the visible spectrum, and the variation of the earth magnetic field was the only one geophysical parameter recorded since 1847.

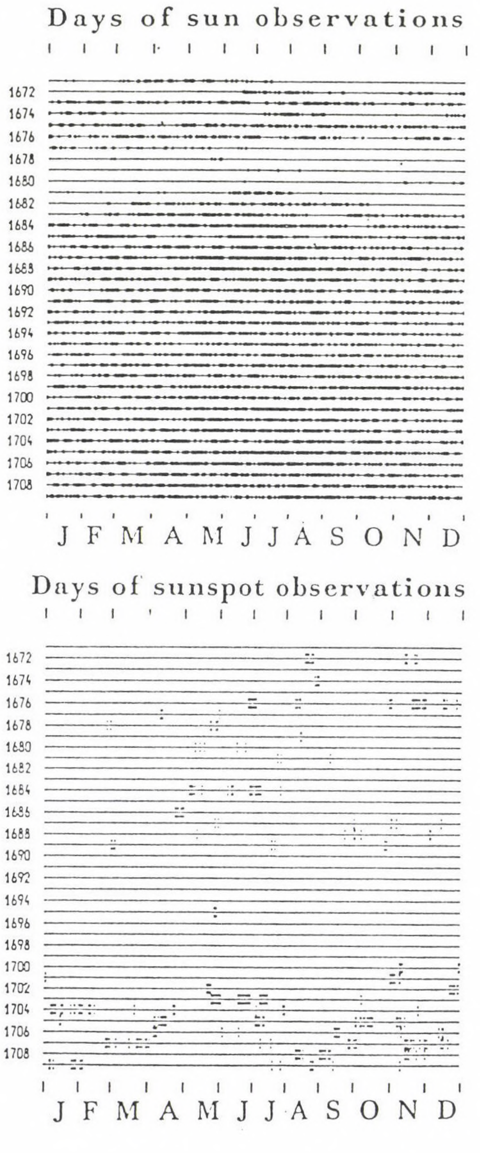


Fig. 12. Sun observations from 1671 to 1709: top, frequency of observation (the hollow dashes correspond to Cassini's observations); b) bottom, frequency of sunspots for all European observers together

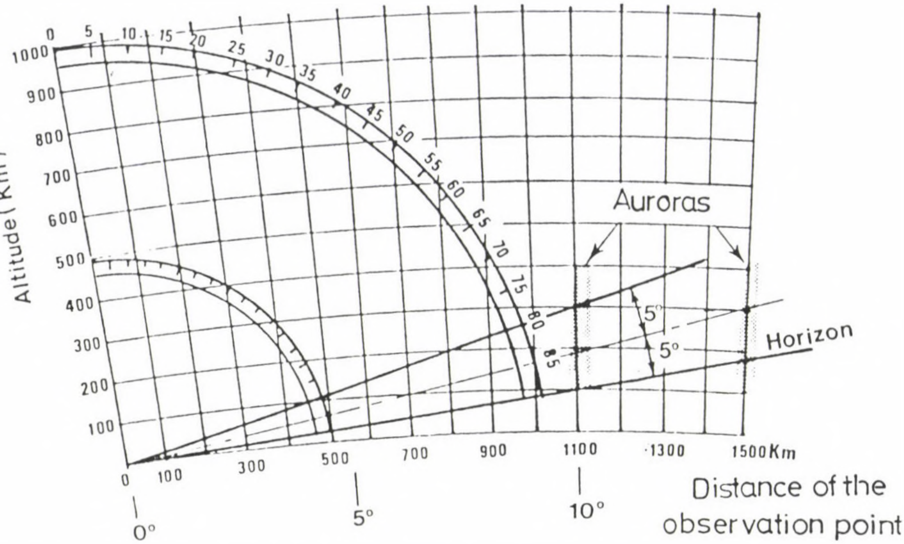


Fig. 13. Diagram of aurora visibility as function of the distance of appearance related to the station of observation

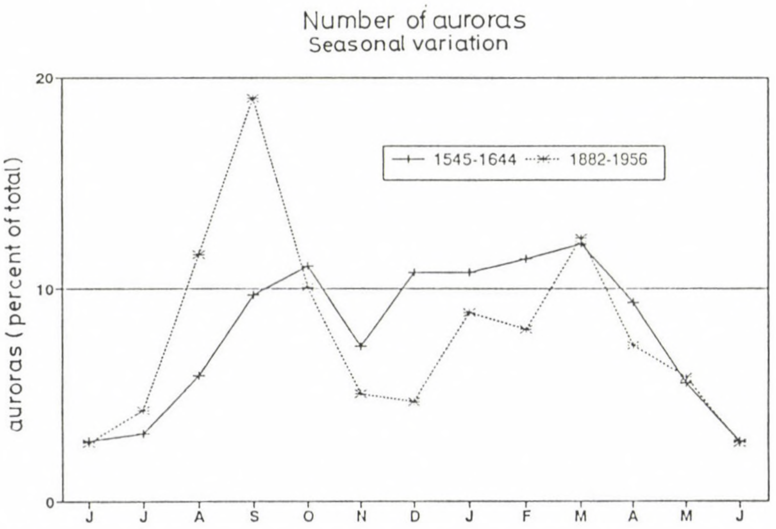


Fig. 14. Seasonal variation of aurorae from European observations between 1545–1644 and recently between 1882–1956

All the actual knowledge on the solar cycle and its consequences on the terrestrial environment were established since the International Geophysical Year and the advent of spatial era (1957), during the four last and large sunspot cycles ($R_{\max}$ between 106 and 190). For the elaboration of a reliable model of the running of the "solar machine", data are obviously missing on the development of small sunspot

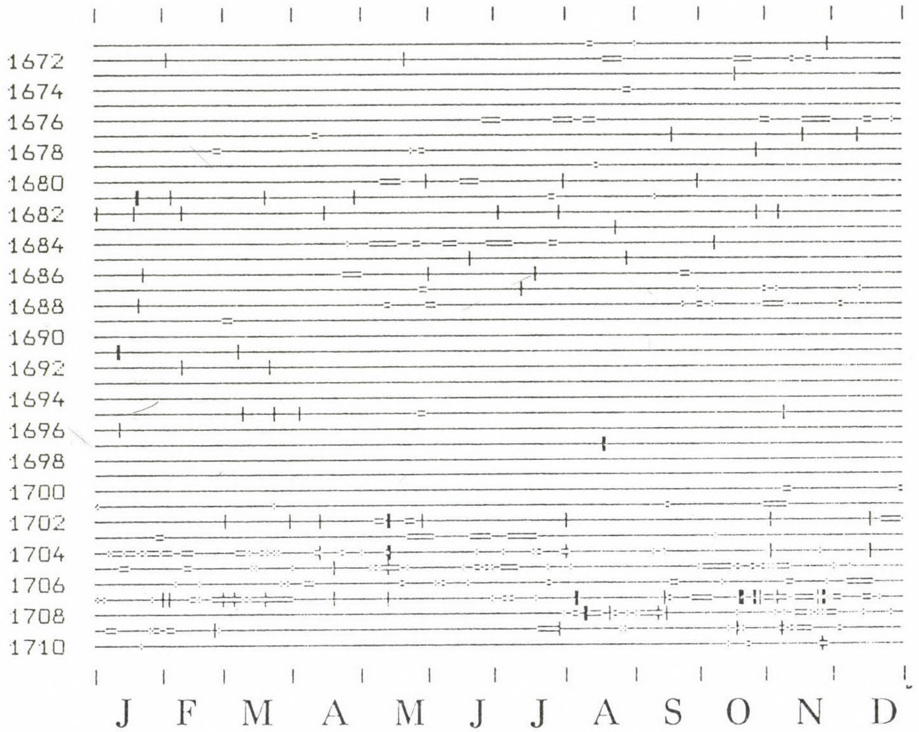


Fig. 15. Diagram representing the day when sunspots (small empty sign) and aurorae (fig. sign) were observed between 1671 and 1709

cycles. It is necessary to continue assiduous solar and geophysical observations, from the ground as well as from space, waiting the hypothetic return of a sun sine macula.

Appendix 1

The "Maunder minimum" and the climate variations during the 17th century

Several authors have established a relation between the minimum called "Maunder" and the "little ice-age" (Eddy 1976, Ribes et al. 1987) while other detailed studies on the climatic variations which occurred during the last five centuries made by Legrand (1979), Landsberg (1980) and Pfister (1980) do not reveal a permanent temperature reduction during this minimum from 1645 to 1715.

The ambiguity seems to result from the definition of the "little ice-age". This name is generally given to the cooling period which started around the middle of the 16th century and ended around 1700. It corresponds to the meteorological conditions: large temperature variations from one year to another and from one decade or several decades to the following. According to Lamb (1982), "it is reasonable" to consider the period from 1420, or even 1190, until 1850 or 1900, as part of the "little ice-age".

However that may be, the reconstitution of the climatic variations made by Legrand (1979) on the basis of the vintage dates in France and Switzerland and on hard winters which occurred in Occidental Europe shows that during the 16th and 17th centuries, the minimum of the "little ice-age" occurred between 1550 and 1630. It is during this period that "the historical maximum" of ice extension in Alps appeared (Le Roy Ladurie 1967, Pfister 1988). Another cooling, well pronounced, occurred between 1690 and 1700. It is confirmed by the temperature measurements made at Paris by Louis Morin (Legrand and Le Goff 1987). So, only this last brief and intense cooling occurs in the centre of the minimum named after Maunder. Besides, it coincides with the period of the weakest solar activity. But, it appears hazardeous to deduce a connection between the two phenomena since the maximum amplitude of the "little ice-age" took place during a period of relative sustained solar activity (Legrand et al. 1990).

Appendix 2

Comparison of the weak solar activity period of the end of the 17th century to the one of the beginning of the 19th century

The reconstitution of the spot cycles made by Wolf for the 17th century and the beginning of the 19th century, shows evidence of a weak activity period between 1798 and 1822 including two small cycles ($R_{\max} = 47.5$ in 1804; and $R_{\max} = 45.8$ in 1816) comparable to the cycle of 1703 ($R_{\max} = 53$).

During this period of a sunspot activity greater than the 17th century activity (1671–1700), the number of observed aurorae in Europe at a latitude equal or less than 55° (Fritz 1873), has been only 124 during 25 years, *i.e.* a mean value of 5 aurorae/year. This mean value is particularly weak in 1807 and 1813 since it is about 1 aurora/year. It is comparable to the auroral activity level from 1671 to 1700. Father Cotte (1808) had noticed this reduction:

On Fig. 24, we have plotted the curve of the frequency of aurora appearances at the geomagnetic latitude equal or less than 62°. This curve has been established by Legrand and Simon (1987), on the basis of homogeneous data. According to the previous remarks on the conditions of aurora observations, the maxima of this curve correspond to aurorae which were mainly produced by relatively high speed solar wind jets coming from the regions continuous to the "neutral sheet" and from polar coronal holes. The first ones sweep the Earth at the time of the solar dipole turning near the spot maximum (1804 and 1817) and the second ones when the dipole reaches its maximum intensity (1807 and 1820). The aurorae shown on a Bartels' diagram, present a clear 27 days recurrence during the years close to the maxima. This fact brings another confirmation concerning their origin.

The small number of aurorae which appeared in 1807 at the geomagnetic latitude 62° indicates that the maximum intensity of the dipolar field was weak announcing the development of the weak sunspot cycle 1816. The structure of such a field corresponded to a thick "neutral sheet" in which the Earth was during several consecutive years. Therefore, the magnetic activity remained very low from 1809 to 1815, and this leads to the quasi complete disappearance of aurorae at latitudes equal or lower to 55°.

On the contrary, in 1820, the auroral activity was high. It was corresponding to a more intense maximum of the dipolar field which announced the cycle of 1830 whose maximum reached 71 ($R_{\max} = 71$).

Notes

1. The "relative Wolf number" is an index of the sunspot activity introduced by Wolf in 1849. It can be computed by the expression $R = K(10g + f)$, where g is the number of spots groups and f of isolated spots, R the total number of spots, and K, a factor close to one, depending on the observer and his tool. This index which is globally proportional to the spotted surface of the solar disk is more significant than the simple number of spots.
2. To determine the location of spot appearance on the solar surface, Carrington defined the meridian of solar origin: it is the central meridian of the sun which pass by the ascendant node f the solar ecliptic equator on January 1st, 1854 at 12 UT. The sideral solar rotation period being 25.38 days, the synodic rotation is comprised between 22.20 and 27.33 days following the Earth position on its orbit. One can deduce the daily longitude of the central solar meridian.
3. The study made by Trellis concerns several ten thousand spot crossings during eight solar cycles.
4. Louis Morin (1635–1715) French physician, member of the Sciences Academy.

Активность солнца и полярного сияния в I7-ом веке

Ж П Легранд, М Ле Гофф, С Мазадие, В Шредер

Темные пятна на диске солнца, которые видны и невооруженным глазом, а также полярные сияния, которые иногда простираются до зоны экватора являются наиболее явными проявлениями активности солнца. Известно уже более I00 лет, что указанная активность имеет довольно правильные циклы с периодом в II лет, но по интенсивности этих циклов можно говорить о "малых" и "больших" циклах солнца.

В настоящее время целый ряд больших циклов следят друг за другом, но в конце I7-го века, во время возникновения инструментальной астрономии в Европе активность оказалась в течение десятилетий несравнимо слабее.

Этот период небольшой активности, который навлек на себя внимание германского астронома Шперера уже в I890 г., известен под названием "максимума Маундера".

Что мы знаем об этом периоде?

Были ли принятые приборы достаточно точными?

Могла ли настойчивость астрономов справиться с климатическими условиями, которые характеризовались повышенной облачностью?

Что мы можем узнать на основе данных, полученных еще в начале I7-го века, во время изобретения астрономической трубы?

Настоящий доклад занимается этими вопросами, и показывает оригинальные документы в свете настоящих сведений о нашей звезде.

References

Badrudinn R, Yadav S. Yadav N R 1986: *Sol. Phys.*, 105, 413–428.
Bailly J S 1779–1782: Histoire de l'astronomie depuis la fondation de l'école d'Alexandrie jusqu'à l'époque de 1782. Paris
Beer J, Andrée M, Oeschger H, Strauffer B, Bonani C, Stoller C, Stuter M, Wôlfi W, Finkel R C 1983: *Radiocarbon*, 25, 269–278.
Bigourdan G 1928: Observations inédites de taches solaires faites à l'Observatoire de Paris de 1683 à 1719 par Ph La Hire et G Ph La Hire. Mémoire Acad. Sci., 59, 1–44.
Champion D E 1980: *U.S. Geol. Surv.*, 314, 80–824.
Cotte L 1808: Manuscript 548, No 21–2, Bibliothèque municipale de Laon
Danjon A, Couder A 1935: Lunettes et Télescopes. Paris
Deslandres H 1906: Annuaire de Bureau des Longitudes pour 1907. Paris
Eddy J A 1976: *Science*, 192, 4245, 1189–1202.
Eddy J A 1983: *Sol. Phys.*, 89, 195–207.
Fritz H 1873: Verzeichnis beobachteter Polarlichter. Gerold, Wien
Gleissberg W, Damboldt Th 1979: *J. Brit. Astron. Assoc.*, 89, 5, 440–449.
Hevelius 1647: Selenographia. Dantzig
King H C 1979: The history of the telescope. London
Kopeczky M, Kuklin G V 1987: *Bull. Astron. Inst. Czechosl.*, 38, 193–200.
Lalande J 1776: Hist. Acad. Sci.
Lalande J 1778: Hist. Acad. Sci.
Lamb H 1982: Climate history and the modern world. London
Landsberg H E 1980: *Arch. Met. Geoph. Biok.*, Ser. B28, 181–191.
Legrand J P 1979: *La météorologie*, VI(16), 167–182.
Legrand J P 1979: *Revue Française d'aenologie*, 75, 23–50.
Legrand J P 1990: Les aurores de moyenne et basse latitudes, à paraître en 1991
Legrand J P, Le Goff M 1987: *La Vie des Sciences*, 3, 251–281.
Legrand J P, Le Goff M, Mazaudier C 1990a: *Ann. Geophys.*, 8, 637–644.
Legrand J P, Le Goff, , Mazaudier C 1990b: Document de travail CNET-CNRS, DT/CRPE/1195
Legrand J P, Simon P A 1985: *Astron. Astrophys.*, 152, 199–204.
Legrand J P, Simon P A 1987: *Ann. Geophys.*, 5A, 161–168.
Legrand J P, Simon P A 1988: *J. Brit. Astron. Assoc.*, 98, 6, 311–312.
Legrand J P, Simon P A 1989: *Ann. Geophys.*, 7A, 565–578.
Legrand J P, Simon P A 1990: *Solar Physics*, 131, 187–209.
Le Roy Ladurie E 1967: Histoire de climat depuis l'an mil. Flamarion, Paris
Link F 1964: *Geophysikalni Sbornik*, 212, 501–547.
Link F 1978: *Solar Physics*, 59, 175–178.
Maraldi J Ph 1695: Manuscript B4–8, Bibliothèque de l'Observatoire de Paris
Martres M J 1987: Pers. communication
Maunder E W 1922: *Brit. astrom. Assoc. J.*, 32, 140–145.
Morice E, Chartier F 1954: Méthode statistique. Paris
Pfister Ch 1980: *J. Interdiscipl. Hist.*, X:4, 665–696.
Pfister Ch 1988: *Hist. Mesure*, III-3, 313–358.
Pingré A G 1901: Les Annales Célestes du WVII siècle publiées par Bigourdan. Paris
Ribes E, Ribes J C, Barthalot R 1987: *Nature*, 325, 5.
Ribes E, Merlin Ph, Ribes J C, Barthalot R 1989: *Ann. Geophys.*, 7, 321–330.
Sakurai K 1980: *J. History Astron.*, 11, 164–173.
Scheiner P 1630: Rosa Ursina
Schröder W 1979: *J. Atm. Terr. Phys.*, 41, 445–446.
Schröder W 1988: *Meteorol. Atmos. Phys.*, 38, 246–251.
Schröder W 1989, 1990: Pers. Communications
Shirley J W 1974: Thomas Harriot. Renaissance Scientist. Oxford
Simon P A, Legrand J P 1989: *Ann. Geophys.*, 7A, 579–594.
Simon P A, Legrand J P 1990: *L'astronomie*, 363–372.
Spoerer F W 1889: Über die Periodizität der Sonnenflecken seit dem Jahre 1618. Royal Leopold Caroline Academy, Halle, Saale

Stringer W J, Belon A E 1967: *J. Geoph. Res.*, 72, 245–250.
Stringer W J, Belon A E, Akasofu S I 1965: *J. Atm. Terr. Phys.*, 27, 1039–1044.
Tarde J 1622: Les astres de Borbon et apologie pour le soleil, Monstrant et vérifiant que les apparences qui se voyent dans la face du soleil sont des planètes et non des taches, comme quelques italiens et allemands observateurs d'icelles luy ont imposés. Paris
Trellis M 1971: *C.R. Acad. Sci. Paris*, Ser. B, 1026–1028.
Venkatesan D, Shukla A K, Agrawal S P 1982: *Solar Phys.*, 81, 375–381.
Waldmeier M 1961: The sunspot activity in the years 1610–1960. Zürich
Wolf Ch 1902: Histoire de l'Observatoire de Paris, de sa fondation à 1793. Paris
Young C A 1883: Le Soleil. Paris
Zhen-Tao Xu 1982: Chinese ancient sunspot records and the variability of solar activity. Pers. communication

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 283-291 (1992)

CONTRIBUTIONS TO THE THEORY OF THE MOST FREQUENT VALUE

T FANCSIK[1]

[Manuscript received July 10, 1992]

The most frequent value characterizes in a most general sense the location of the most dense compaction of measured values thus the estimate denoted by M (Steiner 1973) is — as emphasized by Steiner, too — one of the many possibilities with the advantage of easy computation.

The present paper introduces a definition of the deviation based on the dihesion ε as on scale parameter. This deviation yields then formally in the same way the statistical norm Λ_p of the system of deviations belonging to all measured values, as the norm L_p is defined on the basis of the absolute values of the differences. In a special case the arithmetic mean appears here (as an algorithm making Λ_p a minimum) as well as M^* (its definition is given by Steiner 1988). As a limit case the algorithm yielding the usual most frequent value M is also included.

The present paper contributes to the theory of the most frequent value (conceived this estimation is full generality) by giving a treatment which differs from the usual one and enables thus to treat in just the same way several different known possibilities.

Keywords: distribution types; estimation; most frequent value

1. Introduction

The selection of the statistical method used for the processing of geophysical data systems is influenced by several factors. Such factors are the occurrence of outliers in the data systems which distort the results or the difficulty to predict something about the type of the distribution. The demands for such estimation methods were summarized by Steiner (1988).

In the following a norm will be defined which fulfils the mentioned criteria and which enables to deduce uniformly known estimation methods. A two-parameter supermodel is introduced using which a certain part of methods deduced from the mentioned norm are obtained on the basis of information theory, too.

2. Estimations based on a minimum of the distance from a data system

Let us consider a system of real data $x_, x_2, \ldots, x_n$. The point on the x-axis lying nearest to them is obtained using the so-called L_p norm

$$L_p = \left[\sum_{i=1}^{n} d_i^p\right]^{1/p}$$

[1]Pollack M. u. 39, H-3700 Kazincbarcika, Hungary

(where $p > 0$ and $d_i = |x_i - T|$) by minimizing: T minimizes the value L_p for the chosen value of p (and naturally for the given sample $(x_1, \ldots, x_n)$). For the case $p = 2$:

$$T = E = \frac{1}{n} \sum_{i=1}^{n} x_i$$

i.e. the arithmetic mean, and for $p = 1$:

$$T = \text{med}_n = \begin{cases} x_{(n+2)/2} & \text{for odd } n \\ \dfrac{x_{n/2} + x_{(n+2)/2}}{2} & \text{for even } n \end{cases}$$

the sample median. If the P-norm

$$P = \left[\prod_{i=1}^{n} \delta_i\right]^{1/2}, \quad k, \varepsilon \neq 0$$

$$\text{for} \quad \delta_i = \sqrt{(k\varepsilon)^2 + (x_i - T)^2} \tag{1}$$

is considered (ε and k are real numbers greater than zero), then the algorithm yielding the point with the minimum distance is according to the P-norm defined by the interatively meant equation

$$T = M_k = \frac{\sum\limits_{i=1}^{n} \dfrac{x_i}{(k\varepsilon)^2 + (x_i - T)^2}}{\sum\limits_{i=1}^{n} \dfrac{1}{(k\varepsilon)^2 + (x_i - T)^2}}$$

referring to the most frequent value (Steiner 1973, 1991). As statistical norms have also to fulfil the criterium for norms that by substituting x_i by $c \cdot x_i$, the norm should also change to its c-times value, ε has to fulfil the condition

$$\varepsilon^2 = \frac{3 \sum\limits_{i=1}^{n} \dfrac{(x_i - T)^2}{[\varepsilon^2 + (x_i - T)^2]^2}}{\sum\limits_{i=1}^{n} \dfrac{1}{[\varepsilon^2 + (x_i - T)^2]^2}} . \tag{2}$$

Let us define the following norm based on Eq. (1):

$$\Lambda_p = \left[\sum_{i=1}^{n} \delta_i^p\right]^{1/p}, \tag{3}$$

where ε fulfils the condition of Eq. (2). According to this norm, the T values characterized by the minimum distance from the data system are for different values of p the following:
for $p = 2$

$$T = E = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{4}$$

i.e. the arithmetic mean, for $p = 1$:

$$T = \frac{\sum_{i=1}^{n} \frac{x_i}{\sqrt{(k\varepsilon)^2 + (x_i - T)^2}}}{\sum_{i=1}^{n} \frac{1}{\sqrt{(k\varepsilon)^2 + (x_i - T)^2}}} \tag{5}$$

for $p \to 0$:

$$T = M_k = \frac{\sum_{i=1}^{n} \frac{x_i}{(k\varepsilon)^2 + (x_i - T)^2}}{\sum_{i=1}^{n} \frac{1}{(k\varepsilon)^2 + (x_i - T)^2}} \tag{6}$$

(for a proof, see Eq. 18). For $p = -1$ the minimum place is

$$T = \frac{\sum_{i=1}^{n} \frac{x_i}{[(k\varepsilon)^2 + (x_i - T)^2]^{3/2}}}{\sum_{i=1}^{n} \frac{1}{[(k\varepsilon)^2 + (x_i - T)^2]^{3/2}}} \tag{7}$$

and finally for $p = -2$:

$$T = M^* = \frac{\sum_{i=1}^{n} \frac{x_i}{[(k\varepsilon)^2 + (x_i - T)^2]^2}}{\sum_{i=1}^{n} \frac{1}{[(k\varepsilon)^2 + (x_i - T)^2]^2}} \tag{8}$$

(for the notation M^* see Steiner 1991b).

An infinite number of algorithms (*i.e.* estimations for the location parameter in the case of symmetrical distributions) can be deduced on the basis of Eq. (3), the above ones are the most important ones from different points of view. The case $0 < p \leq 1$ should specially mentioned. In this case, by substituting $p = 1/q$ $(q \geq 1)$, Eq. (3) yields the iterative formula

$$T_q = \frac{\sum_{i=1}^{n} \frac{x_i}{[(k\varepsilon)^2 + (x_i - T)^2]^{\frac{2q-1}{2q}}}}{\sum_{i=1}^{n} \frac{1}{[(k\varepsilon)^2 + (x_i - T)^2]^{\frac{2q-1}{2q}}}} . \tag{9}$$

This includes both the arithmetic means (Eq. 4) and the most frequent value (Eq. 6); in addition the estimation M^* (Eq. 8) can also be expressed as the determination of the minimum place of a norm. This is a new statement according to the knowledge of the present author: namely M^* is the value for which the distance to the data system is a minimum according to the norm Λ_{-2} (see Eq. 8).

The next section shows that Eq. (9) can be deduced from other mathematical principles, too.

3. The estimates M_k^q

If $f(x)$ and $g(x,T)$ are probability density functions of two error distributions, then the I-divergence of f referring to g is

$$I_g = \int_{-\infty}^{\infty} f(x) \log_2 \frac{f(x)}{g(x,T)}\, dx \,. \tag{10a}$$

This value is called also the distance of the two distributions; according to information theory, it measures the information loss (the integral is interpreted at $g(x,T) > 0$; Rényi 1962). If $f(x)$ characterizes the actual mother distribution, then $g(x,T)$ is called a substituting distribution. The estimate of the location parameter is obtained as the minimum of the information loss given by Eq. (10a). This fact results in the following expression:

$$\frac{dI_g}{dT} = \int_{-\infty}^{\infty} \frac{\partial g(x,T)}{\partial T} \frac{f(x)}{g(x,T)}\, dx = 0 \,. \tag{10b}$$

Let us consider the following family of two-parameter density functions (*i.e.* the following supermodel):

$$g(x,S,T) = n_q e^{-c\left[1+\left(\frac{x-t}{S}\right)^2\right]^{\frac{1}{2q}}} \tag{11}$$

with the location parameter T and the scale parameter S; $q \geq 1$ is a real number and c is to be chosen to make the integral

$$(n_q)^{-1} = \int_{-\infty}^{\infty} g(x,S,T)\, dx$$

finite. Figure 1 shows two examples symmetrical for the origin. If the substituting distribution is the supermodel given by Eq. (11), then the location parameter can be determined using Eq. (10b) as

$$T = \frac{\int_{-\infty}^{\infty} \frac{1}{\left[S^2 + (x-T)^2\right]^{\frac{2q-1}{2q}}} f(x)\, dx}{\int_{-\infty}^{\infty} \frac{1}{\left[S^2 + (x-T)^2\right]^{\frac{2q-1}{2q}}} f(x)\, dx} \,. \tag{12}$$

The best estimate of the location parameter of the density function modelling a data system $f(x_i,T)$, $i = 1,2,\ldots,n$ is in a maximum likelihood sense the solution of the equation

$$\sum_{i=1}^{n} -\ln f(x_i,T) = \min \,. \tag{13}$$

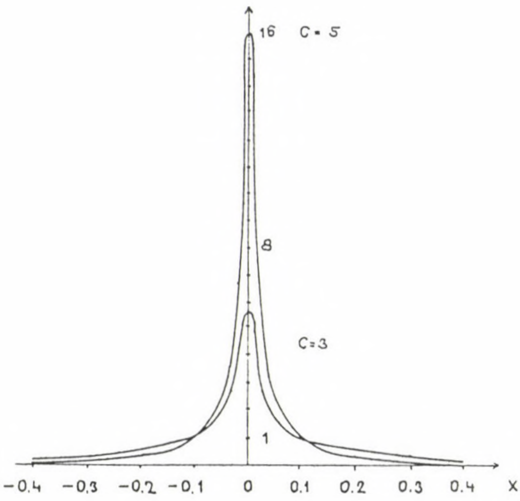


Fig. 1. Two probability density functions of the supermodel defined by Eq. (11), for the parameters $T=0, S=0.01, q=5, c=3$ and 5, respectively

Having carried out the operation and taken into account Eq. (11), one gets:

$$T = \frac{\sum\limits_{i=1}^{n} \frac{x_i}{\left[S^2+(x_i-T)^2\right]^{\frac{2q-1}{2q}}}}{\sum\limits_{i=1}^{n} \frac{1}{\left[S^2+(x_i-T)^2\right]^{\frac{2q-1}{2q}}}} \tag{14}$$

The integral form of Eq. (14) is evidently:

$$T = \frac{\int\limits_{-\infty}^{\infty} \frac{x}{\left[S^2+(x-T)^2\right]^{\frac{2q-1}{2q}}} f(x)\,dx}{\int\limits_{-\infty}^{\infty} \frac{1}{\left[S^2+(x-T)^2\right]^{\frac{2q-1}{2q}}} f(x)\,dx} . \tag{15}$$

Returning to Eq. (14), it is the same as Eq. (9), it is the expression for the data systems given by Eq. (12). That means that using supermodel of Eq. (9) the maximum likelihood principle, the criterium of the norm minimum and the information theory equally yield the same estimate of the location parameter. In connection with the generally meant most frequent value, the estimate obtained previously by three different ways is denoted by $M_k^{(q)}$. This estimate corresponds to the following characteristics of the probabilistic distribution having the character of a location parameter:

$$M_k^{(q)} = \frac{\int\limits_{-\infty}^{\infty} \frac{x}{\left[(k\varepsilon)^2 + \left(x - M_k^{(q)}\right)^2\right]^{\frac{2q-1}{2q}}} f(x)\,dx}{\int\limits_{-\infty}^{\infty} \frac{1}{\left[(k\varepsilon)^2 + \left(x - M_k^{(q)}\right)^2\right]^{\frac{2q-1}{2q}}} f(x)\,dx} . \tag{16}$$

In this case Eq. (2) is naturally to be substituted by the integral form of the dihesion:

$$\varepsilon^2 = \frac{3 \int\limits_{-\infty}^{\infty} \frac{\left(x - M_k^{(q)}\right)^2}{\left[(\varepsilon)^2 + \left(x - M_k^{(q)}\right)^2\right]^2} f(x)\,dx}{\int\limits_{-\infty}^{\infty} \frac{1}{\left[(\varepsilon)^2 + \left(x - M_k^{(q)}\right)^2\right]^2} f(x)\,dx} \tag{17}$$

Equations (16) and (17) yield further

$$\lim_q M_k^{(q)} = M_k , \tag{18}$$

or in another form

$$M_k^{(\infty)} = M_k$$

i.e. one gets again the generalized most frequent value (Steiner 1988). This proves the validity of Eq. (6), too, as $p \to 0$ is equivalent to $q \to \infty$, as $p = 1/q$.

4. The estimates $M_k^{(q)}$ for the generalized Student model family

According to investigations made *e.g.* by Jeffreys, Newcomb and Dutter and due to other practical experiences, too, the most acceptable distribution family for the characterization of the data systems in geosciences is the generalized Student distribution family:

$$f_a(x) = n(a) \frac{1}{\left(\sqrt{1+x^2}\right)^a} \qquad (a > 1) \tag{19}$$

$$n(a) = \frac{\Gamma\left(\frac{2}{a}\right)}{\sqrt{\pi}\Gamma\left(\frac{a-1}{2}\right)}$$

(for the standard case, *i.e.*, for $T = 0$ and $S = 1$). It contains the Gaussian, the Cauchy and the geostatistical distributions, too. The occurrence frequency of the mother distribution types (for data systems) in the corresponding publications can be approximated by a density function of the form

$$f(t) = 16te^{-4t} \tag{20}$$

where $t = 1/(a-1)$ (Steiner 1991a), therefore the present estimates will be checked for the model family $f_a(x)$.

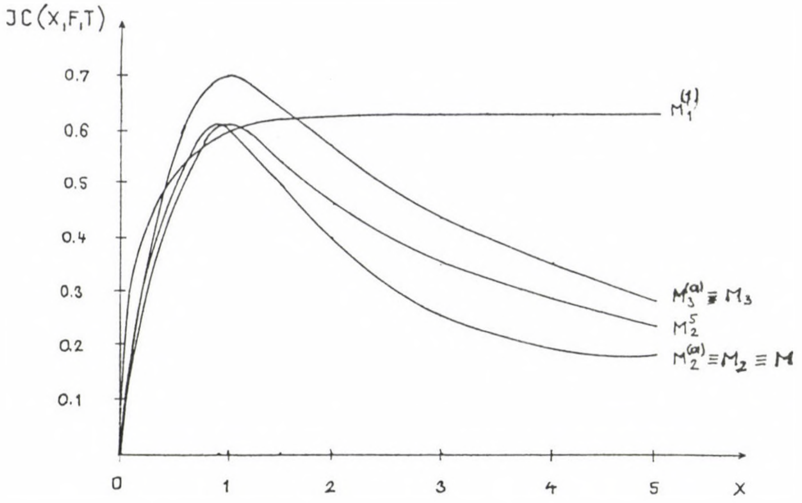


Fig. 2. IC-curves for some estimates $M_k^{(q)}$

The bias of the estimates can be studied by the so-called IC-functions. They are defined (Huber 1981) as:

$$IC(F,x,T) = \frac{S}{\int\limits_{-\infty}^{\infty} \Psi\left(\frac{y}{S}\right) f(y)\, dy} \Psi\left(\frac{x}{S}\right),$$

where the function Ψ characterizes the statistical algorithm in the following sense: the T location parameter can be determined as a solution of the equation

$$\int\limits_{-\infty}^{\infty} \Psi\left(\frac{x-T}{S}\right) f(x)dx = 0 .$$

This solution is in the present case

$$\Psi(x,S,T) = \frac{\frac{x-T}{S}}{\left[1 + \left(\frac{x-T}{S}\right)^2\right]^{\frac{2q-1}{2q}}} .$$

For the density function $f_8(x)$ Fig. 2 shows some IC curves. The basis of the estimate is directly obtained from the IC-function, as it can be shown that in the case of a sample with n elements an additional element x may change the estimate by $IC(x)/n$, if n is sufficiently large.

The function IC characterizes not only the bias but the statistical error, too, as the asymptotic variance of the estimates can be written in the following form (Huber 1981):

$$A^2 = \int\limits_{-\infty}^{\infty} IC^2(x) f(x)\, dx$$

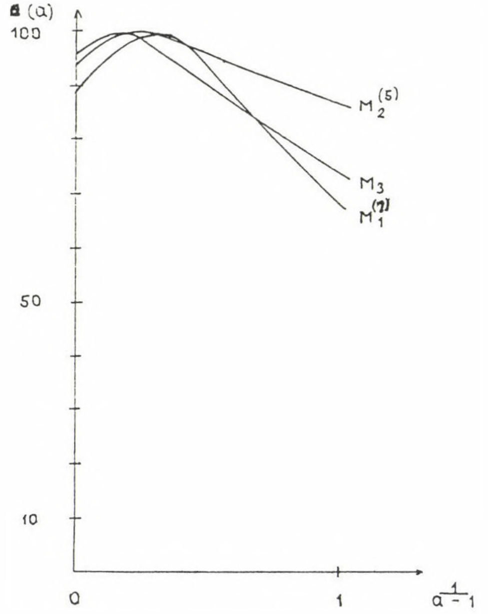


Fig. 3. Absolute efficiencies of the estimates $M_1^{(1)}$, M_3 and $M_2^{(5)}$

The minimum asymptotic variance divided by the asymptotic variance of the actual statistical algorithm yields the absolute efficiency of the estimate:

$$e = \frac{A_{\text{min}}^2}{A^2} .$$

For the supermodel of Eq. (19) one gets (Steiner 1990):

$$A_{\text{min}}^2 = \frac{a+2}{a(a-1)} .$$

It is a very important problem whether efficiences are sufficiently large in finite ranges of actually occurring probability distribution types. If this condition is fulfilled, the estimate is a robust one characterized for appliers also numerically by the average efficiency

$$\bar{e} = \int_0 f(t)e(t)\,dt$$

(Steiner 1991a). We have seen that in geophysics Eq. (20) is acceptable for $f(t)$. The mean efficiencies for some estimates:

E (see Eq. 4)	:	36 percent
sample median	:	80 percent
M_3	:	90 percent
$M_2^{(5)}$	:	95 percent
M_2	:	96 percent

A practically important demand for estimates is that asymptotic laws should be fulfilled in a good approximation for a low number n of elements in the sample. This behaviour for finite (especially for small) n-values can be studied most easily by plotting some error characteristics of the estimate *vs.* $1/\sqrt{n}$: this function is

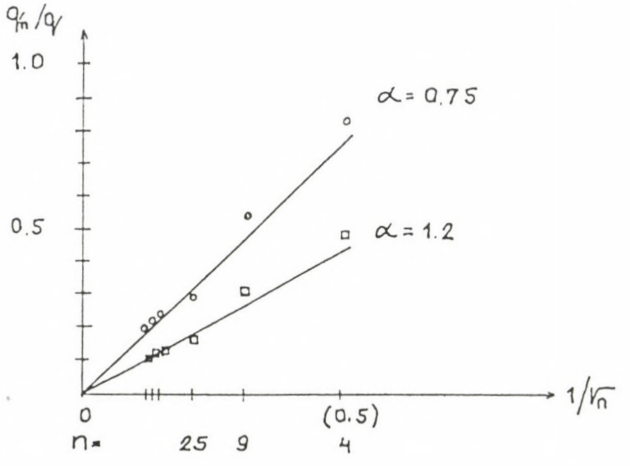


Fig. 4. The probable error q vs. $1/\sqrt{n}$ (n is the number of data in the sample) for the estimate $M_2^{(5)}$ in case of two distribution types of the supermodel $f_\alpha(x)$ ($\alpha = 0.75$ and $\alpha = 1.2$)

namely a straight line trough the origin if the asymptotic law is valid for all n values. Figure 4 shows on the ordinate axis the empirically determined expected error q for estimate $M_2^{(5)}$ in case of two symmetrical stable mother distributions, namely for the type parameters $\alpha = 0.75$ and $\alpha = 1.2$. (The supermodel $f(x)$ is the family of symmetrical stable distributions; analytical form of these density functions are given *e.g.* by Steiner 1990.) The test can be judged, indeed as satisfactory.

References

Huber P J 1981: Robust statistics. Wiley, New York

Rényi A 1962: Wahrscheinlichkeitrechnung. Berlin

Steiner F 1973: *Acta Geod. Geoph. Month. Hung.*, 8, 381–395.

Steiner F 1988: *Geophysical Transactions*, 34, No. 2–3.

Steiner F 1990: Introduction to Geostatistics (in Hungarian). Tankönyvkiadó, Budapest

Steiner F 1991a: *Acta Geod. Geoph. Month. Hung.*, 26, 135–151.

Steiner F 1991b: The most frequent value. Akadémiai Kiadó, Budapest

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 293–300 (1992)

"BUILT-IN" STATISTICS IN GEOPHYSICAL INSTRUMENTS

B HAJAGOS[1] and F STEINER[1]

[Manuscript received July 16, 1992]

One of the many possibilities of noise rejection is to eliminate data which differ by more than 4σ from the arithmetic mean of the data x_i ($i = 1,...,n$), and the remaining data are averaged (σ is here the standard deviation of the data x_i). The estimate $\overline{x}_{4\sigma}$ obtained so is called "4σ-trimmed mean". It is useful to investigate the efficiency of this estimate, *i.e.* to study it as a robust estimate similarly to the well-known α-trimmed mean — even if the primary aim of both estimation methods was to increase the resistance. A comparison with the most frequent value has shown some advantage of this estimate $\overline{x}_{4\sigma}$ in the immediate vicinity of the Gaussian distribution, whereas in the case of realistically heavy flanks proved Monte-Carlo investigations significant advantages of the most frequent value.

Keywords: most frequent value; robust estimation; trimmed mean

1. Introduction

In the first decades of geophysical exploration the data recorded by the instrument were correctly considered as primary ones, as all steps of the processing, including statistical ones, were carried out independently of the instrument. Let us remind to the principle "one measurement is no measurement", applied in certain fields of science (*e.g.* in surveying) earlier than in geophysics. The Eötvös torsion balance, the first geophysical instrument as produced in Hungary in the second quarter of the present century with automatic change of azimuth and with photographic recording led to four repetitions, — similarly the gravimeter (substituting it due to a much quicker field work) measurements were also carried out several times at the same station to achieve a higher accuracy in form of the average of the primary data. This operation, the averaging is the possible simplest statistical operation and it implies Gaussian distribution of the errors supposed to have a symmetrical distribution (in other cases averaging is not the best method). If some of the primary data differed significantly from the others then these were omitted: by supposing to have a blunder, only other data were averaged (Tárczy-Hornoch 1956).

Primary data do not mean necessarily a few numerical data which can be easily judged (visually) (*e.g.* to find outliers). The primary data system of digital seismic measurements contains plenty automatically recorded data which are separately, however, of no interest. The processing including many statistical methods deduces,

[1]Geophysical Institute, University Miskolc, H-3515 Miskolc, Egyetemváros, Hungary

however, from these data time sections which yield ample information both for geologists and geophysicists.

The vibroseis method results (due to the many repetitions) in a primary data quantity which cannot be stored even magnetically and therefore some preprocessing of the data should be made already during the field work to enable a decimation (by order(s) of magnitude) of the data quantity to be brought to the computer centre. Thus field instruments have as component a processor which produces significantly more reliable data of a much smaller quantity on ground of a less accurate (and much bigger) quantity of data.

It is, however, not necessarily a big amount of data which is advantageously preprocessed within the instrument. The firm Scintrex offers a new gravimeter on which the number of repetitions can be set. The primary data x_i ($i = 1,\dots,n$) are produced in each second by the electronics of the balancing system and they are transferred to the "statistical unit" which computes at first the average $\overline{x}$ from all data, then the standard deviation (empirical scatter, σ) and finally the average of the data within the interval

$$(\overline{x} - 4\sigma,\ \overline{x} + 4\sigma).$$

This value is then displayed and stored. This value is denoted here $\overline{x}_{4\sigma}$ and will be called 4σ-trimmed mean.

2. Statistical efficience of the 4σ-trimmed mean in the supermodel f_a

The traditional method to trim a data system is to omit from an ordered data system the highest and the lowest values, when the estimate is the average of the remaining set. If data are omitted in a proportion α at both ends of the distribution (thus the proportion of the remaining data is $1 - 2\alpha$), then the arithmetic mean of the remaining data used to be called α-trimmed mean and it is denoted by $\overline{x}_\alpha$ (*e.g.* Huber 1981). A value of $\alpha = 0.1$ is often applied as it ensures an acceptable protection against the distorting effect of outliers even if they occur in a relatively high proportion. The most frequent value is in itself a resistant quantity, but even this resistance can be increased by "trimming according to weight" for which two possibilities were presented by Hajagos and Steiner (1989). Such trimming methods can be treated analytically, too, *e.g.* analytical expressions can be given for the statistical efficiency, similarly to the α-trimmed means (for the see Huber 1981).

The same is not valid for the 4σ-trimmed mean as though the empirical scatter σ is finite for all samples x_i ($i = 1,\dots,n$), these values σ do not converge for every type of probability distribution to a finite value in the case of $n \to \infty$, *i.e.* the theoretical value of the scatter does not exist (with other words, it is infinite) what prevents the analytical study of this mean for an acceptable wide range of distributions. Thus computations are to be carried out by Monte Carlo-simulations by fixed value(s) of n. If three minutes is an acceptable time limit for a gravimetric measurement, then the value of n is chosen as round 200. In the following the sample contains in each case 200 data (independently if gravimetric or other measurements are considered), the random samples as well as computations of the different estimates (not only the 4σ-trimmed mean) are produced $N = 5000$ times in order to get realistic empirical

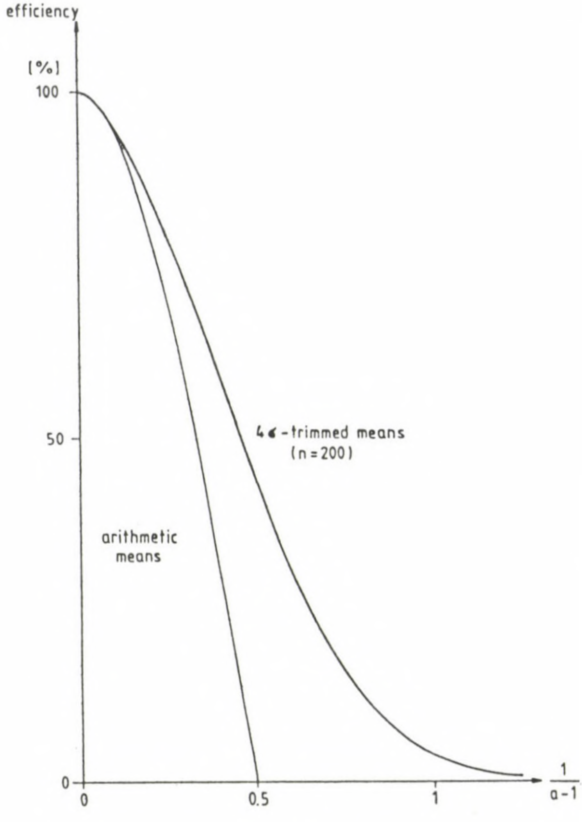


Fig. 1.

probable errors for the different estimates. As we were interested in the error of the error too, the whole process was carried out several times. (In Figs 2, 3 and 4 vertical bars connect the lowest and highest values received from nine repetitions for the distribution types investigated).

The computation of $\overline{x}_{4\sigma}$ aims mainly an increase of resistance (as referred to by the expression "noise rejection" in the description of the gravimeter), but the authors think it probable that the distribution of the primary data is in the presence of microseismic noise not Gaussian even without outliers (in the latter case of the probability for the occurrence of values being farther from the expected value than 4σ would be about 10^{-4}), and therefore constructors found the estimation $\overline{x}_{4\sigma}$ of higher efficiency for usual types of distribution than the simple arithmetic mean.

For any aim or method of the estimation, it is advantageous to study the efficiencies for different types of distribution using supermodels. Such a supermodel is supermodel f_a for which the density functions were defined in a standard form by Steiner (1991, p. 175, by Eqs 12-6 and 12-7), Csernyák and Steiner (1982) used first in statistical studies this supermodel defined for all real values a greater than 1

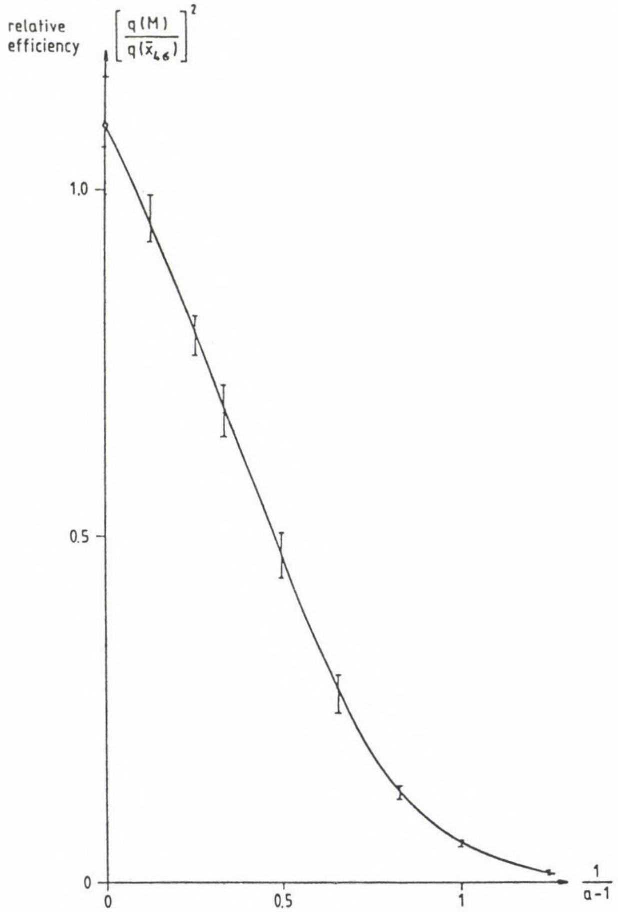


Fig. 2.

(types of the supermodel f_a are called correctly generalized Student-distributions, Hajagos 1985). Some better known distributions from this supermodel; $a = 2$: Cauchy; $a = 5$: geostatistical; $a = 9$: Jeffreys distribution, and the corresponding density functions, resp., in case of $a = \infty$ it converges to the Gaussian distribution. The variance (*i.e.* the square of the standard deviation is $1/(a-3)$ for $a > 3$; for $a \leq 3$ there is no scatter, or in other words: the scatter is infinite.

The square of the minimum asymptotic scatter, the so-called Cramér-Rao bound is computed for the supermodel f_a as follows:

$$A_{\min} = \frac{a+2}{a(a-1)}$$

(Steiner 1991, p. 302). By comparing this with the scatter, the efficiency of the

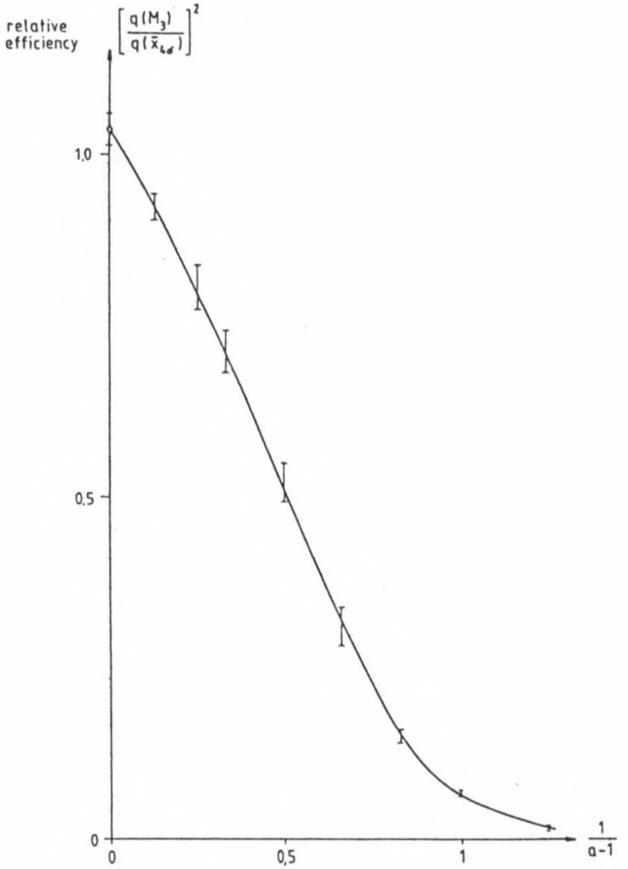


Fig. 3.

arithmetical means is expressed by a simple analytical formula:

$$e(\overline{x}) = \frac{A^2_{\min}}{\text{variance}} = \frac{(a+2)(a-3)}{a(a-1)},$$

if $a > 3$; the efficiency is naturally zero for $a \leq 3$. The corresponding curve $e(\overline{x})$ with a rapid decrease *vs.* $1/(a-1)$ is plotted on Fig. 1. The abscissa is here $1/(a-1)$ as this value characterizes adequately the type distance.

It is known (Huber 1981) that the distribution of the estimates — if they are "sufficiently good" estimates for a given error distribution — approximates asymptotically in the majority of cases the Gaussian distribution; this is supposed (as confirmed by some tests) for the estimates in this paper, too. As the semiinterquartile range q (or with other words, the probable error) and the scatter has a ratio of 0.6745 in the Gaussian case, the empirical semiinterquartile range (calculated on ground of $N = 5000$ estimates) multiplied by $n = 200$ and divided by 0.6745^2

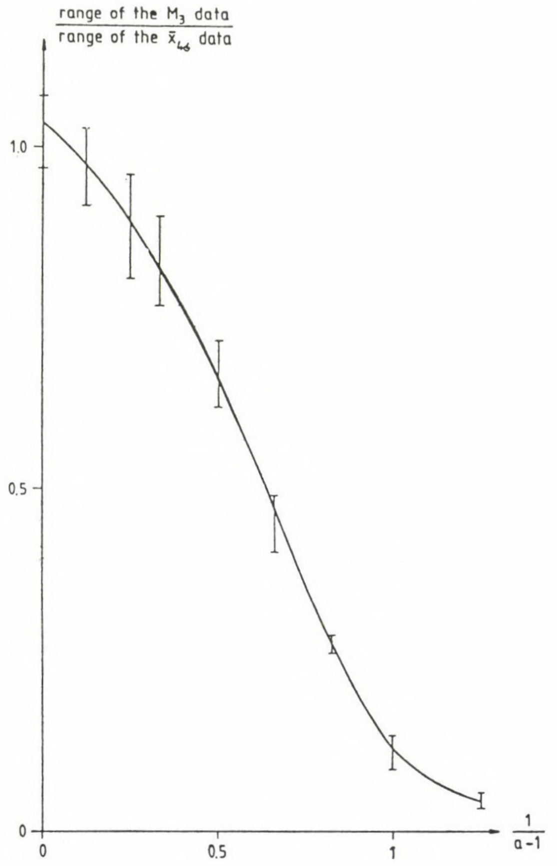


Fig. 4.

yields the empirical value of the asymptotic variance of the estimation method investigated. Thus the empirical values of the efficiency of the estimate $\overline{x}_{4\sigma}$ for the supermodel f_a are given by

$$e(\overline{x}_{4\sigma}) = \frac{(a+2)\cdot 0.6745^2}{a\cdot(a-1)\cdot 200\cdot q^2(\overline{x}_{4\sigma})}.$$

The procedure has been repeated 21 times for each type of distribution (with $N = 5000$) and the curve "4σ-trimmed mean" is plotted in Fig. 1 through the points corresponding to $a = 1.8$; 2; 2.2; 2.5; 3; 4; 5; 9 and ∞ obtained as medians of the estimates $e(\overline{x}_{4\sigma})$ (21 estimates in each case). In the vicinity of the Gaussian distribution, this curve fits nearly completely the $e(\overline{x})$ curve, at the geostatistical distribution, however, the efficiency of the estimate $\overline{x}_{4\sigma}$ has an efficiency higher by 10 percent than the average $\overline{x}$. The distribution of the primary data in the mentioned gravimeter may be of this distribution or even with lower a. This question is not dealt with in the present paper, the following refers to data systems independently of the gravimeter or any other type of instrument, and deals with $\overline{x}_{4\sigma}$ as with a general possibility of estimation.

3. Comparison of the estimate $\overline{x}_{4\sigma}$ with the most frequent value

The most frequent values M are obtained from the data x_i after Column 7 of the table at the end of Steiner's (1991) work. The ratio of the efficiencies of two estimates is measured by the relative efficiency. If the two estimates are denoted generally by $est1$ and $est2$, then the relative efficiency of $est1$ related to $est2$ is given by the ratio of the corresponding acymptotic variances A^2:

$$e_{\mathrm{rel}} = \frac{A^2(est2)}{A^2(est1)}.$$

As the ratio of the q-values (obtained from Monte-Carlo studies) and of the asymptotic scatters is for distributions near the Gaussian one a constant value, the empirical values of e_{rel} are simply determined from

$$e_{\mathrm{rel;emp}} = \frac{q^2(est2)}{q^2(est1)}.$$

The procedure has been repeated nine times both for $est1 = \overline{x}_{4\sigma}$ and $est1 = M$, and the maximum and minimum values are plotted in Fig. 2 versus the distribution type parameters characterized by $1/(a-1)$. The theoretical value is known for the Gaussian distribution, *i.e.* for $1/(a-1) = 0$, as the estimates $\overline{x}$ and $\overline{x}_{4\sigma}$ have the same characteristics for this type; this value is denoted by a small circle in Fig. 2. The relative efficiency of the standard most frequent value, M, as related to the estimate $\overline{x}_{4\sigma}$ is only in the immediate vicinity of the Gaussian distribution higher than 1, it is already at the Jeffreys-distribution ($a = 9$) less than 1, and the relative efficiency continues to decrease rapidly farther on.

Similarly nine repetitions of the procedure yielded the relative efficiencies of the estimate M_3 against $\overline{x}_{4\sigma}$ (the definition of M_3 is found at the same place as that of M; M is considered as the standard variant of the most frequent value, while M_3 has higher absolute efficiency for distributions with shorter flanks). The conclusions are the same as for Fig. 2, in spite of the fact that M_3 is an optimum estimate at a place significantly nearer to the Gaussian distribution, namely at $a = 9$, *i.e.* at $1/(a-1) = 0.125$, while the maximum efficiency of M is near to the geostatistical distribution characterized by $1/(a-1) = 0.25$.

In certain cases it may be desirable that no value should lie far from the correct value (even with a low probability). Therefore the full range of the $N = 5000$ estimates both for M_3 and for $\overline{x}_{4\sigma}$ were computed, and their ratio is plotted in Fig. 4 (similarly for nine repetitions of the procedure). Previous results concerning the more advantageous properties of the most frequent value as compared to $\overline{x}_{4\sigma}$ for the greatest part of the cases is confirmed by this figure, too.

4. Conclusions

Instruments are naturally produced from the best quality elements by the constructors. This concerns many different kinds of parts, beginning with mechanical parameters till quality parameters of electronic parts. As "built-in statistics" gets more and more frequent (not only in geophysical instruments), producers should be cautioned that high level is obligatory here, too. Resistance is anyway needed, *i.e.* if some data or data groups are evidently burdened by errors, then it is advisable to get free of these data in the first step (in the zeroth phase) of the processing, essentially according to the proposition made by Tárczy-Hornoch (1956). This elimination is naturally made today by the computer, too, and it is followed by a robust method of high efficiency for the given type of distribution. If the flanks of the distribution get heavier due to noise then the task is to be solved by the robust method; if outliers, however, appear separately, then it is possible that the best solution is not only to rely upon the robustness of the statistical method used. The efficiency of noise rejection gains if the robust statistics is used on a data set which if freed from evident outliers. To realize this aim, the essence of robustness and resistance should be clearly seen: both notions are closely connected but not identical, they can be distinguished and even they have to be distinguished. It is a pity that many authors do not have these clear ideas even nowadays and often consider as robustness of a method exclusively its ability to free data sets from outliers.

References

Csernyák L, Steiner F 1982: *Publications of the Technical University for Heavy Industry, Series A Mining*, 37, 47–64.

Hajagos B 1985: *Publications of the Technical University for Heavy Industry, Series A Mining*, 40, 225–238.

Hajagos B, Steiner F 1989: *Acta Geod. Geoph. Month. Hung.*, 24, 289–307.

Huber P J 1981: Robust Statistics, Wiley, New York

Steiner F ed. 1991: The Most Frequent Value. Akadémiai Kiadó, Budapest

Tárczy-Hornoch A 1956: Adjustment computations. Felsőoktatási Jegyzetellátó Vállalat, Budapest

Technical Description of the CG-3 and IGS-2/CG-4 AUTOGRAV Automated Gravity Meters. SCINTREX, 222 Snidercroft Road, Concord Ontario Canada, L4K 1B5

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 301–315 (1992)

GEOMAGNETIC INDUCTION HAZARD — A REVIEW

V WESZTERGOM[1]

[Manuscript received August 3, 1992]

With the time variation of the geomagnetic field electric fields of related amplitude and phase are associated. In some simple cases the electric field can be evaluated theoretically. This electric field drives currents (GIC) in earthed conductors often harming them. Many evidences of these harmful processes have been reported since the first submarine communication cables were created. The surface electric field is regarded as a superimposition of primary and secondary sources. The plane wave treatment is somewhat questionable in the auroral region. At midlatitudes the geomagnetic disturbances are less frequent and less intense. What concerns the secondary sources, low conductivity of the crust increases while high conductivity (like in Carpathian Basin) decreases the significance of the induction risk.

This paper attempted to cover the basic understanding of the geomagnetic induction and some engineering criteria for avoiding induction hazards. Although at midlatitudes the hazard is less dramatic, it may be worth to introduce the readers to the terminology and known results not so much for the "risk" itself but for further design of long conductors and for the scientific importance of contemporary observation of magnetic field and GIC in study of the Earth's interior.

Keywords: cables; electromagnetic noise; geomagnetic induction; induction hazard; pipeline; power transmission

Introduction

In recent years many papers have been published concerning environmental effects of the natural EM phenomena. Such effects are often observed especially at high latitudes. One of the most significant effects is the so called "induction risk" which can be defined as the effect of induced currents flowing in ground based (earthed) conductors due to the time varying geomagnetic field. There is no doubt that the phenomenon is more significant at high latitudes, in this paper special attention is devoted to midlatitude induction. It was Gauss who first showed by means of spherical harmonic analysis that a magnetic field of external origin is superimposed on the Earth's magnetic field. The commonly used representation of the field is :

$$B = B_e + B_M + B_a$$

where
- B_e represents the external field
- B_M main field
- B_a anomalous field.

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

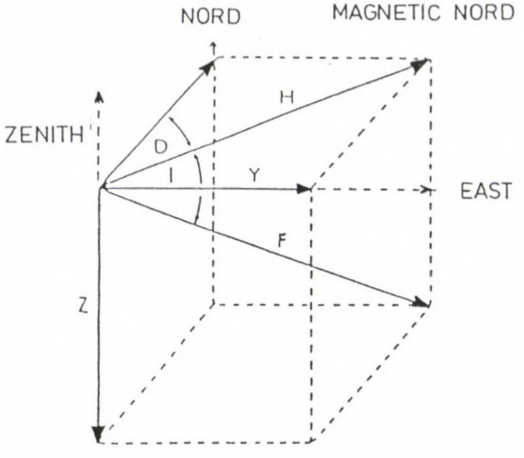
The time variation field has a considerably broad range of frequencies. Variations of the main field (B_M) like secular variation, magnetic field reversal and the apparent polar wandering with their period of $10^2 - 10^8$ years are out of the scope in this paper. The final source of the external magnetic field is the sun. A part of the sun's energy reaches the surface, another affects the space around the Earth causing observable magnetic variation on the surface. In the following the ionospheric, magnetospheric currents will be regarded as primary sources of $B(t)$. The term "magnetic activity" covers occurrence of fluctuations. These fluctuations show a great variety in frequencies, duration, some of them originate inside, others outside the magnetosphere. Although the intensity and the direction of the magnetic field vary not only in time but site by site, too, it is convenient sometimes to speak about it in terms of frequency spectrum . The spectrum has a minimum at about one cycle/second. Electromagnetic variations with frequencies greater then 1 Hz are out of our interest because of their relatively low energy. Pulsations are the lowest period significant variations in the Earth's magnetic field spectrum. Their frequency is $10^0 - 10^{-3}$ Hz, the amplitude at high latitudes may reach several hundred nT but at midlatitudes it seldom exceeds 1 nT. More or less regular variations are the daily variations. The typical amplitudes of these characteristic diurnal variations are several tens of nT-s. The most important transient magnetic events are sfe, substorm and storm. The first can be defined as the strengthening of the Sq current system, the second of the auroral electrojet. Magnetic storms are known as intervals of high activity which occur concurrently at all locations. Frequencies below 10^{-4} Hz, seasonal , semi- and biannual variations, geomagnetic jerks, impulses are out of our interest. In this part of the spectrum the induction effect is negligible. (The amplitude is low with respect to period.)

During great magnetic storms the amplitude of the variation field may exceed 1percent of the main field. (The components of the main field for 1991 *e.g.* in the Nagycenk Geophysical Observatory: D = 102′, H = 21 021 nT, Z = 42 813 nT.) One of the most intense magnetic storms recorded in Nagycenk Geophysical Observatory is seen on Fig. 2. Figure 1 shows the elements of the Earth's magnetic field.

Soon after the installation of the first telegraph systems the effect of geomagnetic fluctuation was observed and reported (Barlow 1849). During the great magnetic storm of March 1989 the following effects were reported by Joselyn (1989) :

- Power outage in Hydro-Quebec (Canada) (6 million users affected for more than 9 hours) and serious voltage loss in 6 Swedish 130 kW power transmission lines due to quasi DC Geomagnetically Induced Current (GIC).
- HF radio communications were all but absent during the storm (as a consequence of auroral electron precipitation and dynamic storm effect)
- The inflation of the neutral atmosphere (due to heating) led to increasing drag on low orbiting satellites.

The report of the WDCA for Solar-Terrestrial Physics of March 1989 cited additional evidences for technological effect during the above magnetic storm:



Elements of the Earth's magnetic field

Fig. 1.

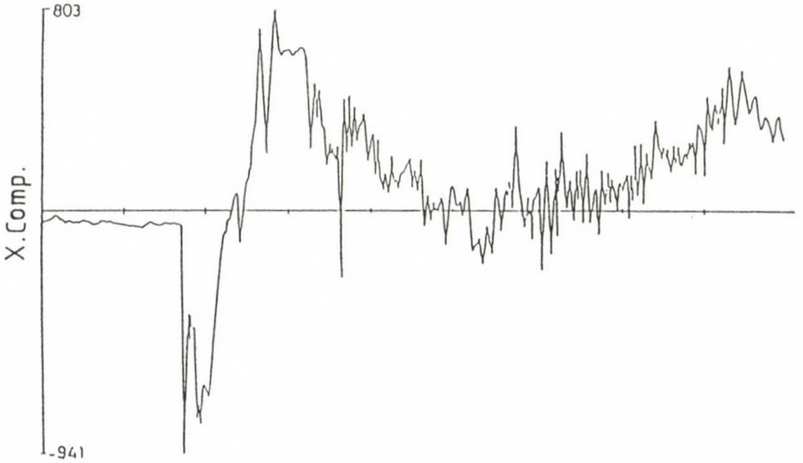


Fig. 2.

— Satellite communication circuit anomalies

— Minor power outage in New Mexico, New York and New Jersey

— Multiple microchip production facilities in NE US were out of operation

— Large voltage swings in undersea cables

— Geoelectrical surveys were disrupted

Geomagnetic induction in ground based conductors

As the cited reports showed probably the most serious problem associated with a geomagnetic storm is the geomagnetically induced current, GIC. Let us see now what is the physics behind this phenomenon.

Symbols and units used in this paper:

t: time(s)
ϕ: charge density (C/m^3)
E: electric field (V/m)
D: electric flux density (As/m^2)
ε_0: permittivity of the free space (8.854×10^{-12} As/Vm)
ε: relative permittivity
H: magnetic field strength (A/m)
B: magnetic induction (1Vs/m^2 = 1T)
μ_0: magnetic permeability ($4\pi 10^{-7}$ Vs/Am)
μ: relative magnetic permeability
c speed of light in free space (2.998×10^8 m/s)
σ: conductivity (Siemens=1/ohm)
j: electric current density (A/m^2)
η: surface electric charge density (C/m^2)
p: surface magnetic charge density (A/m).

The Maxwell equations in MKSA:

$$\operatorname{rot} H = j + \frac{\partial D}{\partial t} \tag{1}$$

$$\operatorname{rot} E = -\frac{\partial B}{\partial t} \tag{2}$$

$$\operatorname{div} D = \phi \tag{3}$$

$$\operatorname{div} B = 0 \tag{4}$$

$$B = \mu\mu_0 H$$

$$D = \varepsilon\varepsilon_0 E$$

Boundary conditions:

$$D_{n1} - D_{n2} = \eta$$

$$B_{n1} - B_{n2} = 0$$

$$E_{t1} - E_{t2} = 0$$

$$H_{t1} - H_{t2} = p$$

The second Maxwell equation, rot $E = dB/dt$ or in integral form ($\oint E\,dl = \int dB\,dF$) expresses Faraday's law of induction which implies F that an electric field is generated by the time variating magnetic field. The voltage between arbitrary points A and B is determined by $U = \int_A^B E s\, ds$. From Eqs (1–4) the following equations are derived by separating E and H

$$\Delta E - \varepsilon\mu \frac{\partial^2 E}{\partial t^2} - \mu\sigma \frac{\partial E}{\partial t} = 0 \tag{5}$$

$$\Delta B - \varepsilon\mu \frac{\partial^2 E}{\partial t^2} - \mu\sigma \frac{\partial B}{\partial t} = 0. \tag{6}$$

Pretty mathematical formulas are obtained by the following assumptions:

— the source field is a plane wave, harmonic in time and space (the source field — like pulsations, daily variation, transient events — appears nearly unified in character over rather big areas on the surface).

— the plane wave travels vertically (no vertical component)

— the Earth is assumed to be a homogeneous infinite half space with horizontal flat surface.

General solution of Eqs (5, 6) :

$$E(r,t) = E(z,t) = E' e^{i(\omega t - kz)} + E'' e^{i(\omega t + kz)} \tag{7}$$

$$B(r,t) = B(z,t) = B' e^{i(\omega t - kz)} + B'' e^{i(\omega t + kz)} \tag{8}$$

where $'$ and $''$ represent vertically downward and upward travelling waves and $k = \omega^2 \mu\varepsilon - i\omega\mu\sigma$ the complex wave number.

The electrical properties of media (0) and (1) are:

$$\begin{array}{lccc} 0 & & & \\ & \sigma_0 & \varepsilon_0 & \mu_0 \\ \hline & \sigma_1 & \varepsilon_1 & \mu_1 \\ 1 & & & \end{array} \quad \text{surface}\,.$$

For sake of simplicity the Cartesian coordinate system is chosen so that:

$$B = (0, B_y, 0) \quad \text{and} \quad E = (E_x, 0, 0)$$

Substituting the general solutions Eqs (7, 8) into Eq. (2), in medium (1):

$$E_{1x} = \frac{\omega}{k} B_{1y}\,.$$

Similarly in medium (0):

$$E'_{0x} = \frac{\omega}{k} B'_{0y} \qquad \text{and} \qquad E''_{0x} = \frac{\omega}{k} B''_{0y}\,.$$

The continuity equations have now the following form

$$E'_{0x} + E''_{0x} = E_{1x} \qquad \text{and} \qquad \frac{1}{\mu_0} B''_{0y} + \frac{1}{\mu_0} B'_{0y} = \frac{1}{\mu_1} B_{1y}\,.$$

The observed magnetic induction in medium (1) is obtained from the incident and the reflected waves: $B^0 = B'_{0y} + B''_{0y}$. Now the surface electric field ($E = E'_{0x} + E''_{0x}$) is expressed by B^0:

$$E = \frac{\omega\mu_0}{k\mu_0} B^0 .$$

It must be noted here that the surface electric field (and magnetic of course) is superimposed of primary and secondary sources. The latter is the effect of multiple reflections and attenuations which is often represented by the skin depth

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} .$$

More realistic calculations

In more realistic calculations generally a great deal of *a priori* information is needed. In case of nonharmonic variations assuming that

$$E_{My}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} E_{My}(\omega) e^{i\omega t}\, d\omega$$

and

$$B_{Mx}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} B_{My}(\omega) e^{i\omega t}\, d\omega$$

exists, Pirjola (1982) derived the following formula

$$E_{My}(t) = -\frac{1}{\sqrt{\pi\mu_0\sigma}} \int_{-\infty}^{t} du \frac{g(t-u)}{\sqrt{u}}$$

where

$$i\omega B_M(\omega) = q(\omega) \qquad \left(\frac{dB_M(t)}{dt} = g(t) \right) .$$

What concerns the secondary sources, the sphericity of the Earth can be ignored apart from global problems (Price 1973). The conductivity model of the Earth is a basic question in these calculations. Without lateral non-uniformities (homogeneous Earth in horizontal direction) at least within a region comparable in size with the penetration depth, a single transfer function expresses the relation E/H. Assuming horizontal layers or arbitrary changing properties in vertical direction, the above thinking can be followed. Detailed analysis was given *e.g.* Cagniard (1953), Tikhonov (1950), Albertson and Van Baelen (1970), Pirjola (1982). The computation of local induction anomalies is very laborous although a great progress has been made by numerical techniques. Lateral inhomogeneities like dikes may change dramatically the surface electric field computed by the above formulas. Figure 3 represents the distortion of the surface electric field above "dike like" conductivity anomaly models. (The computations were carried out by Steiner's (1990) numerical method.)

The plane wave treatment is somewhat questionable in the auroral region. Geomagnetic induction both in homogeneous and layered Earth in case of line current has been considered theoretically by Pirjola (1982): A sheet current is constructed summarising line currents. The general solution by Lehto (1984) was reduced by Hakkinen and Pirjola (1986) to the following ones:

$$E_y(x,\omega) = -\frac{i\omega\mu_0 J(\omega)}{2\pi}\int\limits_{-\infty}^{\infty} dbe^{ibx}\frac{Z(b)F(b)e^{-K_o h}}{i\omega\mu_0 + K_o Z(b)}$$

$$B_x(x,\omega) = -\frac{i\omega\mu_0^2 J(\omega)}{2\pi}\int\limits_{-\infty}^{\infty} dbe^{ibx}\frac{F(b)e^{-K_o h}}{i\omega\mu_0 + K_o Z(b)}$$

$$B_z(x,\omega) = -\frac{i\omega_0 J(\omega)}{2\pi}\int\limits_{-\infty}^{\infty} dbe^{ibx}\frac{bZ(b)F(b)e^{-K_o h}}{i\omega\mu_0 + K_o Z(b)}$$

where E_y, B_x, B_z represent the total field values at the surface, $Z(b,\omega)$ is the surface impedance $(-\mu_1 E_y(b)/B_x(b))$. $E_y(b)$ and $B_x(b)$ are Fourier transforms of $E_y(t)$ and $B_x(t)$ just below the surface and $F(b) = \int dxe^{-bx}f(x)$ is the Fourier transform of the sheet current's current distribution. For more details see also Wait (1981) and Albertson and Van Baelen (1970).

The geomagnetic induction cannot be treated theoretically in case of two dimensional conductivity anomalies and primary sources asymmetric with respect to the anomaly. The main effects are understood and it is quite clear that this asymmetry differs very much from the plane wave case. In the frame of a research programme between the Finnish Meteorological Institute and the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences (GGRI) in 1990, 1991 the auroral electrojet and different two dimensional conductivity anomalies were studied in the model laboratory of GGRI (see Fig. 4). Simple conductivity models were also created to get some comparison for theoretical calculations. (Theoretical treatment of the problem and comparison of analogue model results were given by Kauristie (1991) and Viljanen (1992).)

As stated earlier the voltage at the surface between arbitrary points A and B is determined as:

$$U = \int\limits_A^B E_s\, ds$$

This potential difference drives currents in the Earth and in earthed conductors like power transmission lines, cables, pipelines *etc.* These currents are called Geomagnetically Induced Currents (GIC).

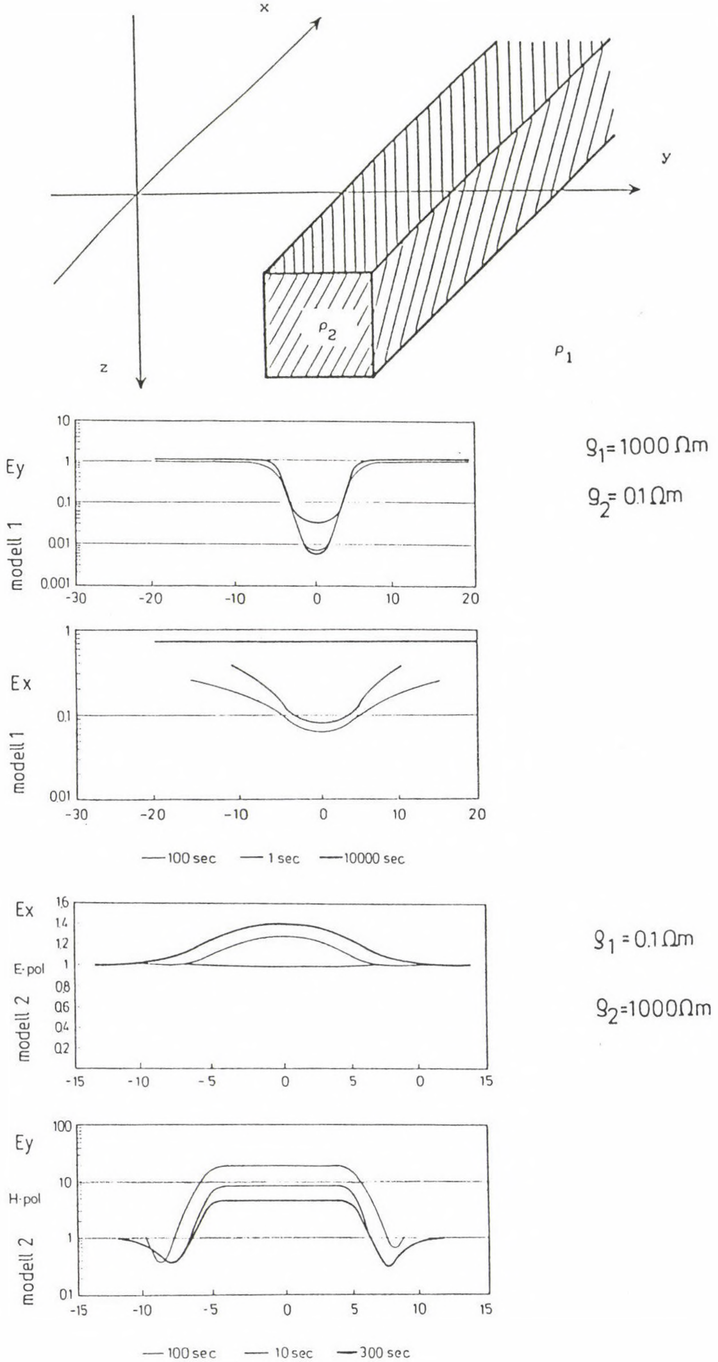

x
y
z
ρ2
ρ1
Ey
modell 1
ϱ1= 1000 Ωm
ϱ2= 0.1 Ωm
Ex
modell 1
100 sec
1 sec
10000 sec
Ex
E-pol
modell 2
ϱ1 = 0.1 Ωm
ϱ2=1000Ωm
Ey
H-pol
modell 2
100 sec
10 sec
300 sec

Fig. 3.

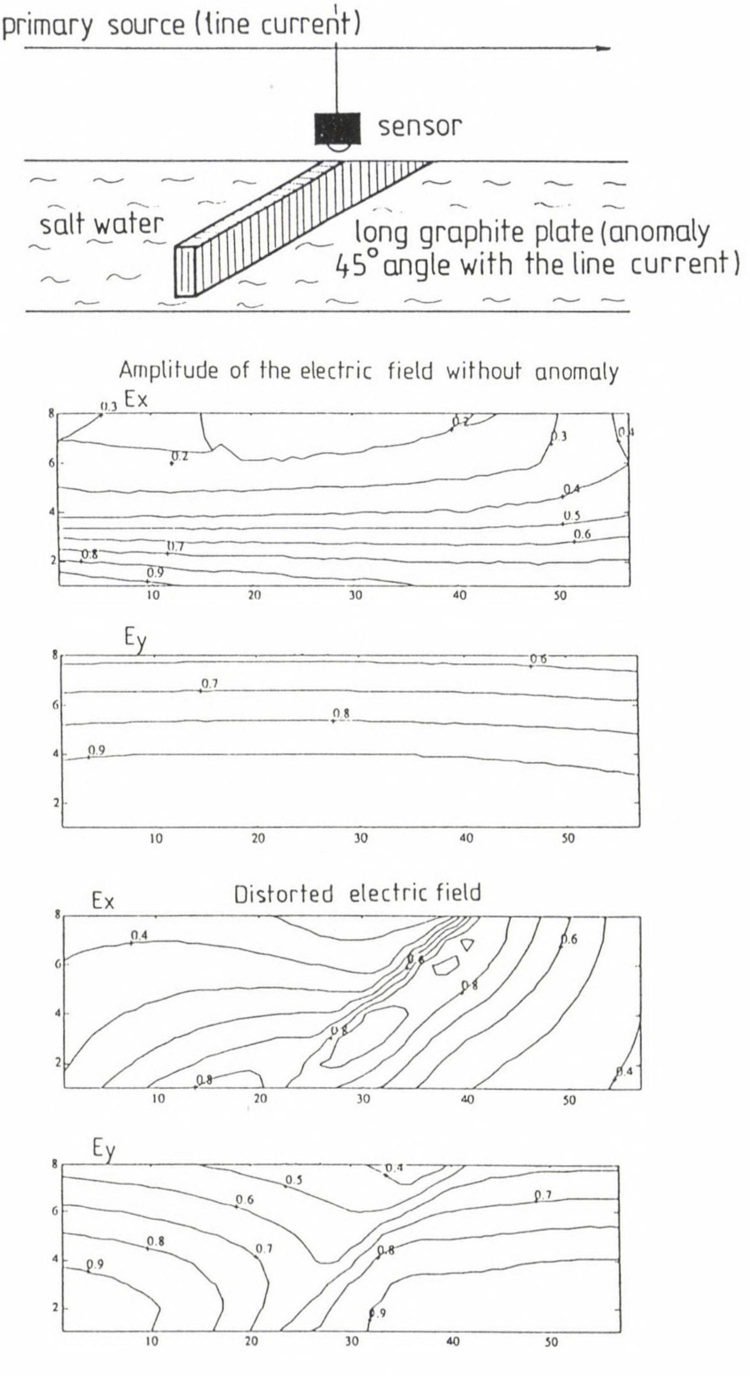

primary source (line current)
sensor
salt water
long graphite plate (anomaly 45° angle with the line current)
Amplitude of the electric field without anomaly
Ex
Ey
Distorted electric field
Ex
Ey

Fig. 4.

Numerical calculations and long series of observatory earth current data (*e.g.* in Nagycenk Geophysical Observatory) show that during quiet conditions the order of E is 10^{-2} V/km. The amplitude of 1V/km may occur during high magnetic activity quite frequently at high latitudes, an amplitude of 50 V/km is considered extremely high (see also Pirjola 1984).

In numerical calculations the electric field is calculated in absence of long conductors what implies that these objects are not visible for the electromagnetic waves in question. Campbell (1984) showed that the penetration depth in a steel pipeline would be 2 meters for 2 minutes periods and 30 meters for 6 hours periods. The few centimeters thick pipe wall hardly disturb these waves.

In lack of earthing points, for a conductor of 100 km length, considering 10 V/km electric field the voltage is from substituting into formulae 1.5, $U = 1$ V.

The force, separating the charges: $F = UQ$ and the Coulomb force:

$$F' = \frac{1}{4\pi\varepsilon_0}\frac{Q^+Q^-r}{r^2r}, \quad \text{equals in equilibrium :}$$

$$F = F' \Rightarrow Q = E4\pi\varepsilon_0 L^2 = 10^{-11}\ \text{As/Coulomb}$$

$I = Q/s \Leftarrow 10^{-10}$ A the current in case of the frequencies in question.

GIC in long conductors like pipelines, power transmission lines and communication cables

Pipelines

Concurrently with the construction of very long oil and gas transmission systems, research on the technological effect of GIC started. (*e.g.* Gideon et al. (1970), Hessler (1974), Campbell (1978, 1984, 1986), Peabody (1967), Viljanen (1989)).

The electric field defined in the previous chapter drives current in an earthed conductor like a pipeline. A possible electric model of a buried pipeline and the principle of corrosion protection are seen on Fig. 5.

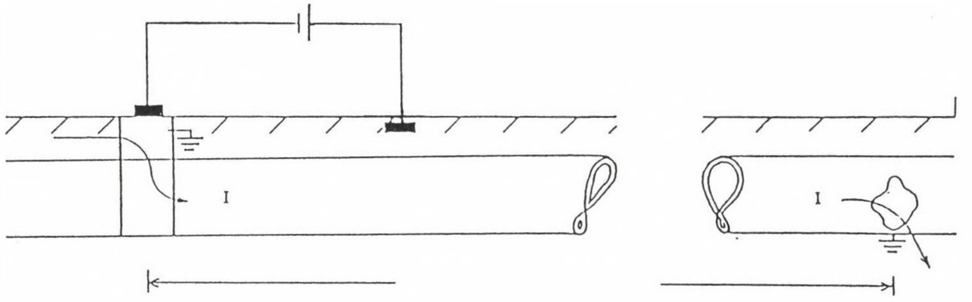


Fig. 5.

Although normally the pipe is electrically insulated from the earth the protective sheath is often damaged. These "holes" and other artificial groundings like pumping stations form paths for currents flowing between the soil and the pipeline. The current flowing between two earthing points can be calculated by Ohm's law ($I = U/I$). Campbell (in 1984) considered the Alaskan oil pipeline. It is approximately 10^3 km long, has a diameter of 1.22 m and the average wall thickness is 1.3 cm. The end to end resistance is 6 ohms (see also Table I). If the distance between two earthing points is long enough the GIC can be even as large as several hundreds of amperes. Table I shows the electrical properties of the pipelines.

Table I.

	Gas	Oil	Steel	Insulation	Water
ε	ε_0	2–5 ε_0	ε_0	2–9 ε_0^{**}	80 ε_0
μ	μ_0	μ_0	200 μ_0	2–3 μ_0	μ_0
σ	10^{-14*}	10^{-14}	$4 \times 10^{6**}$	10^{-8**}	$10^{-1} - 10^{2}$

* — Israel (1971, pp. 95, 248)
** — Viljanen and Pirjola (1989, pp. 405)

Currents flowing in and out of the pipeline through the earthing points according to the reversals of the electric field may cause electric action between the steel of the pipe and the H^+, OH^- ions of the groundwater. At the anode $Fe(OH)_3$, common rost is formed. While high frequency oscillation precludes the electric action, quasi DC current can remove a great amount of steel (Campbell 1984). According to Peabody (1967) one ampere of DC discharging into the usual soil electrolyte can remove approximately 10 kg of steel in one year. This is especially dangerous when the current density is large (narrow path (hole diameter)).

In practice this "hole to hole current" is less dramatic because of the relatively low amplitudes at low frequencies in the geomagnetic spectrum (see *e.g.* Campbell 1986). At mid latitudes the geomagnetically induced electric field is more a nuisance noise in pipeline protection than a direct corrosion problem. (Near the equator the Sq can contribute to the corrosion.)

Viljanen (1989) gave a detailed theoretical analysis of the Finnish natural gas pipeline. He discussed not only the parallel ("hole to hole") current but the radial current density as well. The radial current density (due to the finite resistivity of the insulation) is nearly always less than 1 $\mu A/m^2$ *i.e.* the applied protection current density. The current density along the pipeline ($I = \sigma EA$) is sometimes remarkable larger than 1 $\mu A/m^2$.

Power transmission systems

The effect of GIC on power transmission systems was widely discussed *e.g.* by Albertson and van Baelen (1970), Albertson and Thorson (1974), Lanzerotti (1983), Pirjola and Lehtinen (1985), Boteler and Cookson (1986) and recently a detailed analysis was given by Pirjola (1989), Pirjola and Viljanen (1989). Figure 6 shows a simplified model of a power transmission system earthed through transformers and the flow of GIC (after Albertson and van Baelen 1970).

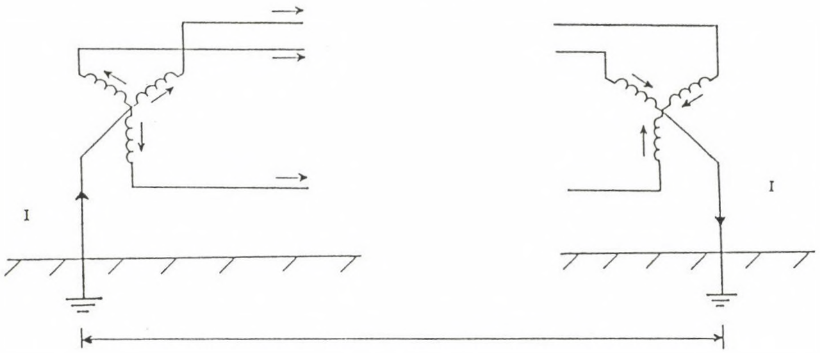


Fig. 6.

If a potential difference exists between two earthing points (in the present case the neutral points of the transformers) electric current flows in the network similarly to pipelines. As the most powerful range of the geomagnetic spectrum is between f $= 10^0$ Hz and 10^{-4} Hz, the widely used frequency in power transmission is 50-60 Hz, the relevant part of the GIC can be regarded as DC current. Different problems associated with GIC were discussed by Albertson and Thorson (1974). In presence of quasi DC currents half cycle saturation of current and power transformers occurs. The saturation leads to overheating of transformers, overloading of capacitors. Overheating of current transformers and overloaded capacitors cause malfunction of protective relays and judgemental problems for system operators while overheating of power transformers result in serious voltage losses, fluctuation or total power outage. Quite similarly, the induction phenomena can be considered in electric railways. Figure 7 represents a geoelectric model of a typical Hungarian electric railway circuit where L is the length of the circuit, R_1 the resistance of the rail, R_2 of the soil, R_3 of the wire. R_4 represents the resistance of the train and R_5 is the transformer which supplies the circuit. In absence of train the circuit is open at the end. Typical length of a circuit is about 50 km and $R_1 \simeq R_2$. From measurements:

$$\frac{R_3 + \frac{R_2}{2}}{L} = 0.1\ \Omega/\mathrm{km}\,.$$

Assuming that the resistance of the train is very small ($\Rightarrow 0$) and it is at the end of the circuit the GIC can be evaluated. Calculating by a (quite large) electric field of 1 V/km the geomagnetically induced current (I) is about 10 A. Although this current does not cause significant disturbances in power supply induced currents in railway security systems should be treated as potential hazard.

Figure 8 represents the occurence of GIC in a Finnish power transmission line as a function of the geomagnetic K-index and strength of GIC. (Data taken from Viljanen and Pirjola (1988).)

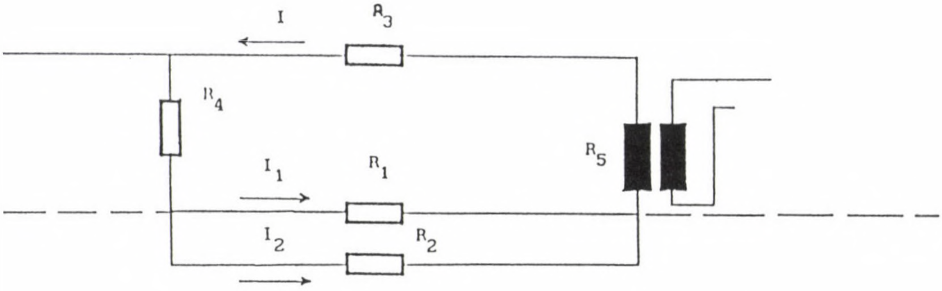


Fig. 7.

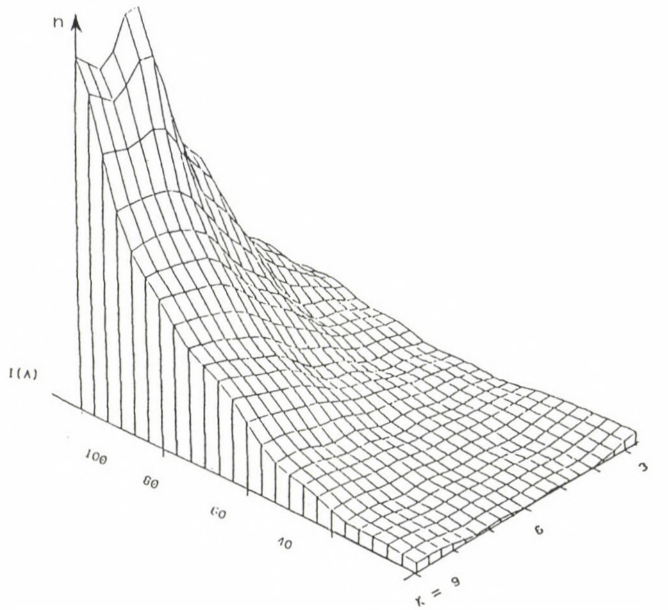


Fig. 8.

Cables

Just like any conductor earthed at both ends in some way cable systems are subjected to the geomagnetic induction. Detailed discussion of this subject was given by Medford et al. (1981), Lanzerotti (1983) and especially by Meloni et al. (1983). During geomagnetic storms the induced voltage may become so large that the powering systems shut down, in other cases it could be operated without external battery supply . Numerous examples of malfunction of cable systems are cited by Lanzerotti (1983). Medford et al. (1981) regarded the daily variation induced currents in submarine cables. They determined the induced voltage in the TAT 6 transatlantic cable for 5 geomagnetic quiet days as the spatial integral of the

Sq field component perpendicular to the loop formed by the cable and the earth (appr. 10^{12} m^2). Both the calculated and the observed voltage drop along the length of the cable was $\cong 5$ V.

The measured induced voltage in long cables in the frequency range $10^0 - 10^{-4}$ Hz is widely used for scientific purposes especially in studying the electrical properties of the crust in oceanic region.

(It must be noted that tide flow and geodynamo related induced fields can also contribute remarkably to the total field in submarine cables.)

References

Albertson V D, Van Baelen J A 1970: *IEEE Trans. on Power Apparatus and Systems*, PAS-89, 4, 578–584.

Albertson V D, Thorson J M 1974: *IEEE Trans. on Power Apparatus and Systems*, PAS-93, 1025–1030.

Barlow W H 1849: *Phil. Trans. Roy. Soc.*, 139, 61–72.

Boteler D H, Cookson M J 1986: Telluric Currents and their Effects on Pipelines in the Cook Strait Region of New Zealand. Materials Perform., 27–32.

Cagniard L 1953: *Geophysics*, 18, 605–635.

Campbell W H 1978: *Pure and Appl. Geophys.*, 116, 1143–1173.

Campbell W H 1984: *Surveys in Geophysics*, 8, 239–259.

Campbell W H 1986: *Surveys in Geophysics*, 8, 239–259.

Gideon D N, Hopper A T, Thompson R E 1970: Earth Current effects on Buried Pipelines. Analysis of Observation of Telluric Gradients and their Effects. American Gas Assn.; Cat. No.: L30500

Hakkinen L, Pirjola R 1986: *Geophysica*, 22, 31–44.

Hermance J F, Peltier W R 1970: *J. Geoph. Res.*, 75, 3351–3356.

Hessler V P 1974: *Corrosion*, 56, 1–18.

Hibbs R D, Jones F W 1976: *Geophys J. R. Astr. Soc.*, 46, 433–466.

Joselyn J A 1989: The Great Magnetic Storm of 13 March 1989. Geomagnetic Application Bulletin, International Association of Geomagnetism and Aeronomy, 3, 2.

Kauristie K 1991: *Geophysical Publications*, 24, 5–113.

Lanzerotti L J 1983: *Space Sci. Rev.*, 34, 347–356.

Lehto K 1984: *Geophysica*, 20, 105–121.

Medford L V, Meloni A, Lanzerotti L J, Gregori G P 1981: *Nature*, 290, 329.

Medford L V, Lanzerotti L J, Kraus J S, Maclennan C G 1989: *Geophys. Res. Letters*, 16, 1145–1148.

Meloni A, Lanzerotti L J, Gregori G P 1983: *Reviews of Geophysics and Space Physics*, 21, 4, 795–803.

Peabody A W 1967: Control of Pipeline Corrosion. Nat. Assn. Corrosion Eng., Houston

Pirjola R 1982: *Geophysica*, 18, 1–161.

Pirjola R 1984: *Geophysica*, 20, No. 2, 89–103.

Pirjola R, Lehtinen M 1985: *Annales Geophysicae*, 3, 485–491.

Pirjola R, Viljanen A 1989: *IEEE Transactions on Power Delivery*, 4, 1239–1245.

Pirjola R 1989: *Phys. Earth Planet. Int.*, 53, 214–220.

Pirjola R, Viljanen A 1989: *J. Geomag. Geoelectr.*, 41, 411–420.

Pirjola R, Viljanen A 1991: In: Environmental and Space Electromagnetics, Springer-Verlag, Hong Kong, 276–286.

Pirjola R, Hakkinen L 1991: In: Environmental and Space Electromagnetics, 288–298.

Price A 1973: *Phys. Earth Planet. Int.*, 7, 227–233.

Schmucker U 1970: *J. Geomag. Geoelectr.*, 22, 9–31.

Steiner T 1990: *Magyar Geofizika*, 30, No. 1, 9–17.

Tikhonov A N 1950: *Dok. Akad. Nauk. USSR*, 73, 295–297.

Viljanen A 1989: *Geophysica*, 25, 135–159.

Viljanen A 1992: *Finnish Meteorological Institute Contributions*, 7, 7–134.
Viljanen A, Pirjola R 1988: *J. Geomag. Geoelectr.*, 41, 411–420.
Wait J R 1962: Electromagnetic Waves in Stratified Media. The MacMillan Company, New York
Wait J R 1981: Wave Propagation Theory. Pergamon Press, New York
Williams D J 1979: In: Solar System Plasma Physics, Lanzerotti L J, Kennel C F, Parker E N eds. North-Holland, Amsterdam, 3, 327

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 317-329 (1992)

MODELLING OF VENTILATION NETWORKS TAKING INTO ACCOUNT THE THERMAL INTAKE AND DENSITY CHANGES OF THE AIR

Z Buócz[1] and J Janositz[1]

[Manuscript received February 5, 1990]

Ventilation network model is developed to include heat intake, density changes and changes in physical parameters of the air in air passage. Effects due to vapour intake are also discussed. Numerical examples give information about accuracy limits and effects to be expected.

Keywords: air passage; modelling; vapour intake; ventilation

Introduction

The aim of modelling of ventilation networks is generally to define the characteristics of air current within the network and the physical parameters of the circulating air.

In a well-known model based on known parameters of air passage resistances and those of the air moving equipments, presuming constant volume of the air, air quantities within the air passages can be determined. These calculations neglect influences due to heat and vapour intake of the air. It is obvious that in cases when the influence of heat and vapour intake and the changes of density is considerable on the flow conditions all these must be taken into account when modelling flow conditions in ventilation networks. This paper summarizes relationships and calculation methods.

Calculation of influence of heat intake and density changes on underground air flows

The influence of the natural depression due to heat intake and changes of density is known on flow conditions within ventilation networks (Kegel 1908, Heise and Drekopf 1927).

The calculation of this influence is based on the following:

— conservation of mass,

— conservation of energy,

[1]University Miskolc, H-3515 Miskolc, Egyetemváros, Hungary

— friction of flowing mediums and

— equations of state of air.

According to conservation of mass the mass flow is constant between the beginning and end of the air passage:

$$\dot{m} = F \cdot \varrho \cdot v \tag{1}$$

where
$\dot{m}$ — is the mass flow of the air (kg/s)
F — is the cross-section of the air passage (m^2)
ϱ — is the density of air (kg/m^3)
v — is the average speed of air in the cross-section F (m/s).

The differential form of the principle of conservation of energy in mining spaces without air moving equipment is as follows:

$$-\frac{dp}{\varrho} = f\psi - Ag \cdot dz + v \cdot dv \tag{2}$$

where
p — is the barometric pressure of air (Pa)
g — is the gravity acceleration (9.81 m/s^2)
z — is the height above sea level (m) (In case of inclined air passages $dz = ds \sin\alpha$, where α is the angle between the air passage and the horizontal and ds is differential length of the air passage.)
ψ — is the specific dissipated energy referred to mass (J/kg)
A — is equal to +1 (in case of downwards flow) and to −1 (in case of upwards flow).

Within the ventilation network, cross-section changes are dealt with separately, and the change of speed due to change of air density is negligible therefore within an air passage the term $v \cdot dv$ expressing the change of kinetic energy can be generally neglected.

The differential specific energy $d\psi$ along the differential length ds of the air passage which is dissipated due to friction can be calculated from the contour U of the air passage, from the cross-section F and from the friction coefficient λ as follows:

$$d\psi = \lambda \frac{U}{8F} \cdot v^2 \cdot ds. \tag{3}$$

The changes of the barometric pressure of the air within the air passages are described by the equation:

$$-\frac{dp}{\varrho} = \lambda \cdot \frac{U}{8F^3} \frac{m^2}{\varrho^2} ds - A \cdot g \cdot ds \cdot \sin\alpha. \tag{4}$$

Based on Eq. (4) the elementary Δp pressure change can be calculated for all Δs elementary sections in these sections, all other variables in Eq. (4) as U, F, λ, ϱ can be considered to be constant.

Such a calculation to model ventilation systems are restricted only by the computer-time and by the large quantity of data to be collected and entered. To eliminate the large amount of data and data storage it is investigated how computations based on Eq. (4) can be simplified.

Goal of these investigations is to simplify the calculation of the pressure-change within air passage lengths L characterized by constant geometric features (α, U, F) and by a given air-friction coefficient. This is possible when among parameters, p, ϱ and s being variables in Eq. (4) a relationship can be established. Thus functionality can be sometimes justified between p and ϱ, ϱ and s, further between T and s.

Supposed functionalities among physical parameters of the air within an air passage

McPherson and Robinson (1980) have shown that air conditions in intake and return shafts are politropic ones. The relationships

$$\begin{aligned} p \cdot v^n &= c \\ v &= \frac{1}{\varrho} \end{aligned} \tag{5}$$

yield from measurements the following values of n and c, calculated in McPherson and Robinson (1980) by regression:

— in the intake shaft $n = 1.303$ and $c = 79.3881$,

— in the return shaft $n = 1.3558$ and $c = 81.442$.

Computed values approximate surprisingly the measured ones. The correlation coefficients are $r = 0.99993$ and 0.99992 resp.

Substituting relationship (5) into the differential equation (4) the following formula is obtained:

$$c \cdot n\varrho^n \cdot d\varrho - A\varrho^2 \cdot gds \cdot \sin\alpha + \frac{R_n \cdot m^2}{L \cdot \varrho_n} \cdot ds = 0 \tag{6}$$

where

$R_n = \frac{\lambda \cdot \varrho_n}{8} \cdot \frac{U \cdot L}{F^3}$

ϱ_n — is the density of air in normal condition

($p_n = 101325$ Pa; $T_n = 273.15$ K; $\varrho_n = 1.292$ kg/m^3).

In Eq. (6) n is generally no integer (Heise and Drekopf 1927). Therefore Eq. (6) has no general solution thus Eq. (5) is not sufficient to reduce the air passage to elementary parts for the calculation of pressure changes where the change of ϱ would be negligible.

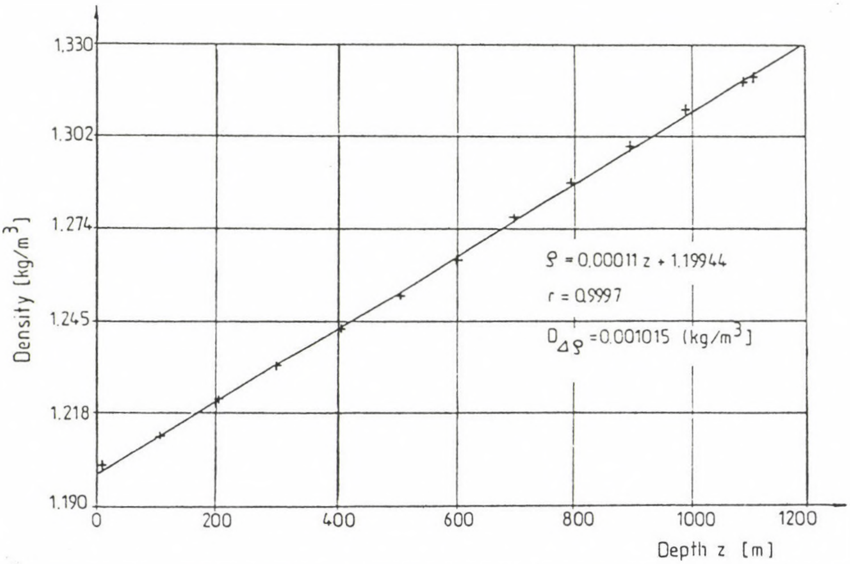


Fig. 1. Density figures in function of depth calculated from data of pressure and temperature measurements by McPherson and Robinson (1980) and the regression straight line defined from the depth-density figures (Intake shaft)

Linearly changing density along the air passage

Plotting the densities (after McPherson and Robinson 1980) in function of depth (see Figs 1 and 2) a close connection is found, thus

$$\varrho(s) = a \cdot s + \varrho_1 \tag{7}$$

where
ϱ_1 — is the density at the starting point ($s = 0$)
a — is a constant.
Substituting Eq. (7) into Eq. (4)

$$dp = A(as + \varrho_1) \cdot g \cdot \sin\alpha \cdot ds - \frac{R_n}{\varrho_n} \cdot m^2 \cdot \frac{ds}{as + \varrho_1} \tag{8}$$

from which the pressure difference between the two end points of the air passage is:

$$\Delta p = p_2 - p_1 = ag \cdot \ell \cdot \sin\alpha \cdot \frac{\varrho_1 + \varrho_2}{2} - \frac{R_n m^2}{\varrho_n} \cdot \frac{\ln \varrho_2 - \ln \varrho_1}{\varrho_2 - \varrho_1} \tag{9}$$

where $\varrho_2 = \varrho(L) = aL + \varrho_1$.

Should in Eq. (8) the term $v \cdot dv$ expressing the change of air speed due to the change in density also to be taken into account, then the term

$$+\frac{m^2}{F^2} \cdot \left(\frac{1}{\varrho_2} - \frac{1}{\varrho_2}\right) = 2 \cdot \Delta p_d \tag{10}$$

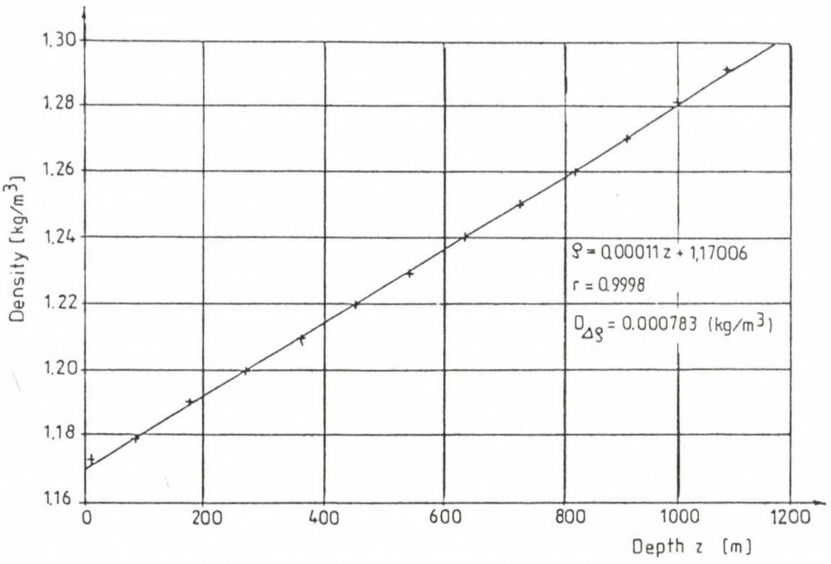


Fig. 2. Density figures in function of depth calculated from data of pressure and temperature measurements by McPherson and Robinson (1980) and the regression straight line defined from the depth-density figures (Return shaft)

referring to the dynamic change of pressure also appeare on the right side of Eq. (9).

Linearly changing temperature along the air passage

Plotting McPherson and Robinson's (1980) data on the air temperature in the shaft vs. depth (Figs 3 and 4) a linear relationship is observed, too:

$$t = bs + t_1 \tag{11}$$

where t_1 is the temperature of the staring point (C°).

Should the vapour content of the air be negligible then the air density as a function of pressure p and of the absolute temperature T can be calculated with adequate accuracy based on the equation of state of ideal gases:

$$\varrho = \frac{p}{RT} \qquad (\mathrm{kg/m^3}) \tag{12}$$

where

R — is the gas constant of dry air ($R = 0.287045$ (kJ/kg))

T — is the absolute temperature of the air ($T = 273.15 +$ (K)).

Applying Eqs (11) and (12) the differential equation (4) can be rewritten as:

$$\frac{dp}{dT} - B_1\frac{T}{p} + B_2\frac{p}{T} = 0 \tag{13}$$

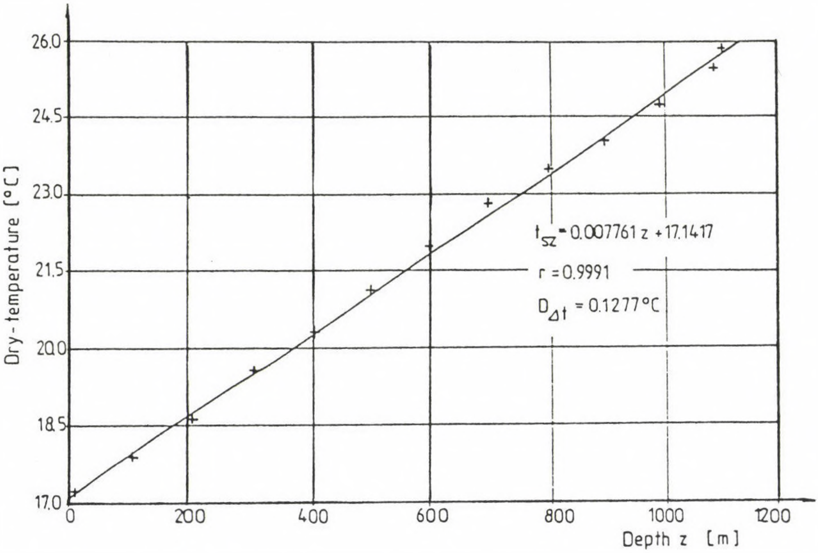


Fig. 3. Dry-temperature data measured in different depths (McPherson and Robinson 1980) and the regression straight line calculated from these data (Intake shaft)

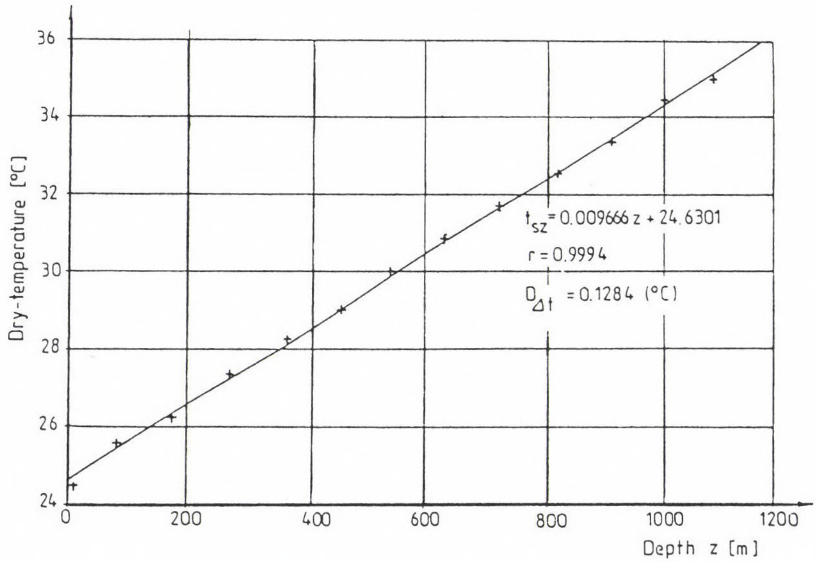


Fig. 4. Dry-temperature data measured in different depths (McPherson and Robinson 1980) and the regression straight line calculated from these data (Return shaft)

where $B_1 = \dfrac{R_n \cdot m^2 \cdot R}{\varrho_n \cdot b \cdot L}$; $\quad B_2 = -\dfrac{A \cdot g \cdot \sin\alpha}{b \cdot R}$.

Applying the values at the starting point (T_1 and p_1) and those at the end (T_2 and p_2) resp. in Eq. (13) for the pressure change the following relationships are found:

In the case of downward air flow ($A = 1$):

$$p = p_2 - p_1 = \frac{T_2}{B_3}\sqrt{1 - \left(\frac{T_1}{T_2}\right)^{B_4} (1 + w_1)^2} \tag{14}$$

where

$B_3 = \sqrt{-\frac{B_2 + 1}{B_1}}$; $b_4 = 2B_1 \cdot B_3^2$

$w_1 = \frac{B_3 \cdot p_1}{T_1}$ and $b = \frac{T_2 - T_1}{L}$.

In the case of upward air flow ($A = -1$) :

$$\Delta p = \frac{T_2}{B_3}\sqrt{\left(\frac{T_1}{T_2}\right)^{B_4} (1 + w_1^2) - 1.} \tag{15}$$

In the case of horizontal air flow ($A = 0$):

$$P_2^2 = p_1^2 + B_1(T_2^2 - T_1^2). \tag{16}$$

Substituting

$$p_k = \frac{p_2 + p_1}{2}; \quad T_k = \frac{T_1 + T}{2} \quad \text{and} \quad \varrho_k = \frac{p_k}{RT_k}$$

the last equation can be written as

$$\Delta p = p_2 - p_1 = \frac{R_n}{\varrho_n \cdot \varrho_k} \cdot \dot{m}^2. \tag{17}$$

Influence of vapour intake on pressure change

The above analyses have not dealt with the vapour content of the air and with its changes respectively, although except of salt mines the vapour content of the flowing air is in most mines considerable. Thus it is justified to investigate how far the vapour influences the atmospheric pressure-change in the flowing air.

Practically this requires simultaneous consideration of two impacts: the impact of the vapour content on density and that of vapour absorption on the mass flow. Should the vapour content of the flowing air be significant within the air passage but not changing considerably then the density of the wet air is to be taken into account. For example densities in data taken from McPherson and Robinson (1980) are calculated considering the moisture content (see Figs 1 and 2). Changes in vapour content (based on the above data) are plotted in Figs 5 and 6 which show that the vapour content is fluctuating but unsignificantly. Besides irregular changes a definite trend is also observed.

To decide, if this trend is to be taken into account it must be cleared how far the change of vapour content influences the air speed and pressure drop respectively in

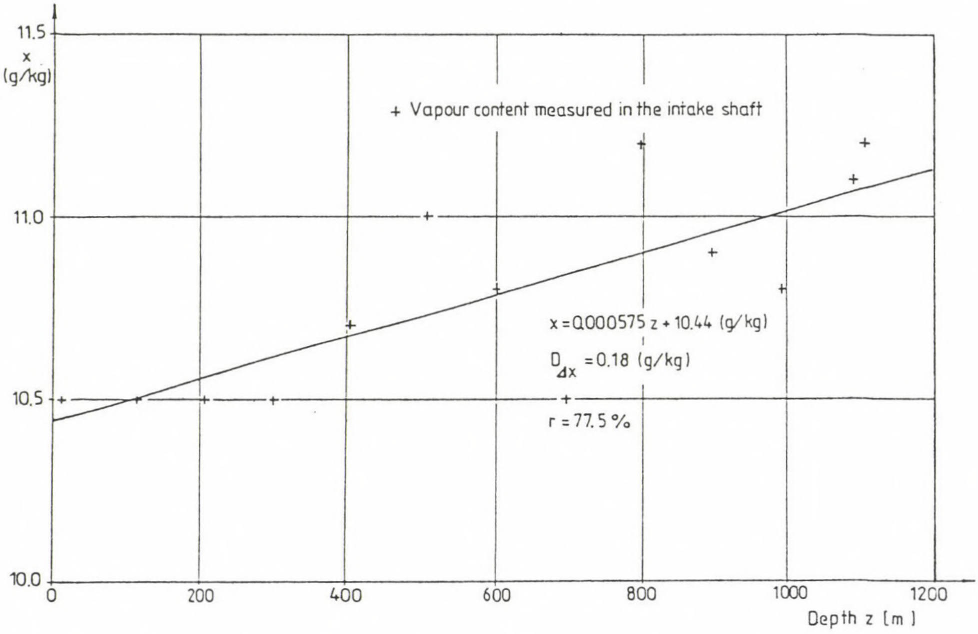


Fig. 5. Change of vapour content in an intake shaft (McPherson and Robinson 1980)

the air passage. For this decision the percentual volume increase is to be investigated when 1 g water vapour is absorbed by 1 kg dry air.

The relevant value is

$$0.001 \cdot \frac{m_{\mathrm{air}}}{m_{\mathrm{water}}} \cdot 100 = \frac{0.001 \cdot 28.96}{18} \cdot 100 \cong 0.16\% \tag{18}$$

where

m_{air} — is 28.96 (molecular weight of air)

m_{water} — is 18 (molecular weight of water-vapour).

Therefore the change of vapour content causes an essential difference in pressure differences required for air-moving when within the air passage the change of vapour content exceeds 20 to 30 g/kg. Such a vapour absorption or vapour discharge occurs in the praxis very seldom. In such cases the air passage is to be devided into parts, in which the increase of the vapour content can be considered as linear within the air passage.

$$x^* = x_0^* + b \cdot s \qquad (\mathrm{kg/kg}). \tag{19}$$

In the following the quantity of the flowing dry air in the mass flow of the air is denoted by $\dot{m}_0$. In this case the total amount of the flowing air in a distance s measured from the starting point of the air passage is as follows:

$$\dot{m} = \dot{m}_0(1 + x_0^* + bs). \tag{20}$$

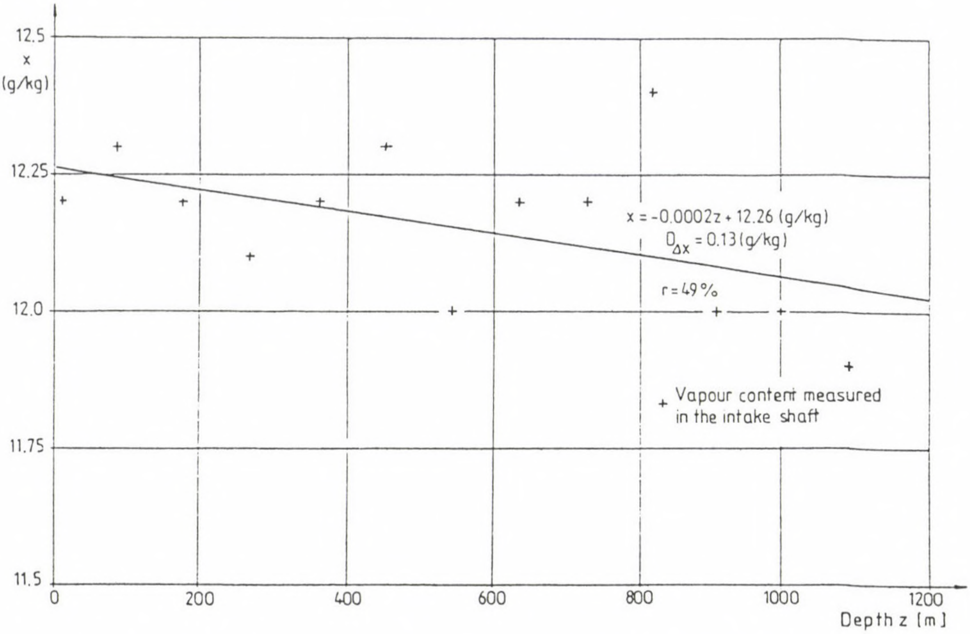


Fig. 6. Change of vapour content in a return shaft (McPherson and Robinson 1980)

Should also the density change be linear along the air passage which means that Eq. (7) is valid, then differential equation (2) (based on Eqs (1, 3, 7) and (20)) gets:

$$\begin{aligned} -\frac{dp}{\varrho_1 + as} &= \lambda \frac{U}{8F^3} \cdot \frac{\dot{m}_0^2 (1 + x_0^* + b_s)^2}{(\varrho_1 + as)^2} \cdot ds - A \cdot g \cdot \sin\alpha \cdot ds + \\ &+ \frac{\dot{m}_0^2}{F^2} \cdot \frac{1 + x_0^* + b_s}{(\varrho_1 + as)^3} \cdot (b\varrho_1 - a(1 + x_0^*) \cdot ds. \end{aligned} \tag{21}$$

Differential equation (21) yields the following relationship for Δp:

$$\begin{aligned} \Delta p &= p_2 - p_1 = A \cdot g \cdot L \cdot \sin\alpha \cdot \frac{\varrho_1 + \varrho_2}{2} - \\ &- \frac{R_n}{\varrho_n} \left(B + C \cdot m \frac{\varrho_2}{\varrho_1} \right) \frac{\dot{m}_0^2}{\varrho_1 - \varrho_n} + \frac{\dot{m}_0^2}{F^2} \cdot D \cdot E \end{aligned} \tag{22}$$

where

A — is the value considered in (2) and

$$B = \frac{\Delta x^{*2}}{2} + 2 \cdot \Delta x^*(1 + x_0^*) - \varrho_1 \frac{\Delta x^{*2}}{\varrho_2 - \varrho_1}$$

$$C = (1 + x_0^*)^2 - \frac{2 \cdot \Delta x^*}{\varrho_2 - \varrho_1}(1 + x_0^*) \cdot \varrho_1 + \frac{\Delta x^{*2}}{(\varrho_2 - \varrho_1)^2} \cdot \varrho_1^2$$

$$D = (1 + x_0^*) - \frac{\Delta x^*}{\varrho_2 - \varrho_1} \cdot \varrho_1$$

$$E = (1 + x_0^*) \cdot \left(\frac{1}{\varrho_1} - \frac{1}{\varrho_2}\right) - \frac{\Delta x^*}{\varrho_2} + \Delta x^* \frac{m\varrho_2 - m\varrho_1}{\varrho_2 - \varrho_1}.$$

Should the change of density of the flowing air be negligible within the air passage then the depression due to the air resistance of the air passage may be calculated as

$$h = \frac{r_n \cdot \dot{m}_0^2}{\varrho_n^2}. \tag{23}$$

In Eq. (22) k-times this depression appears where

$$k = \left(B + C \cdot \ln m \frac{\varrho_2}{\varrho_1}\right) \frac{\varrho_n}{\varrho_2 - \varrho_1}. \tag{24}$$

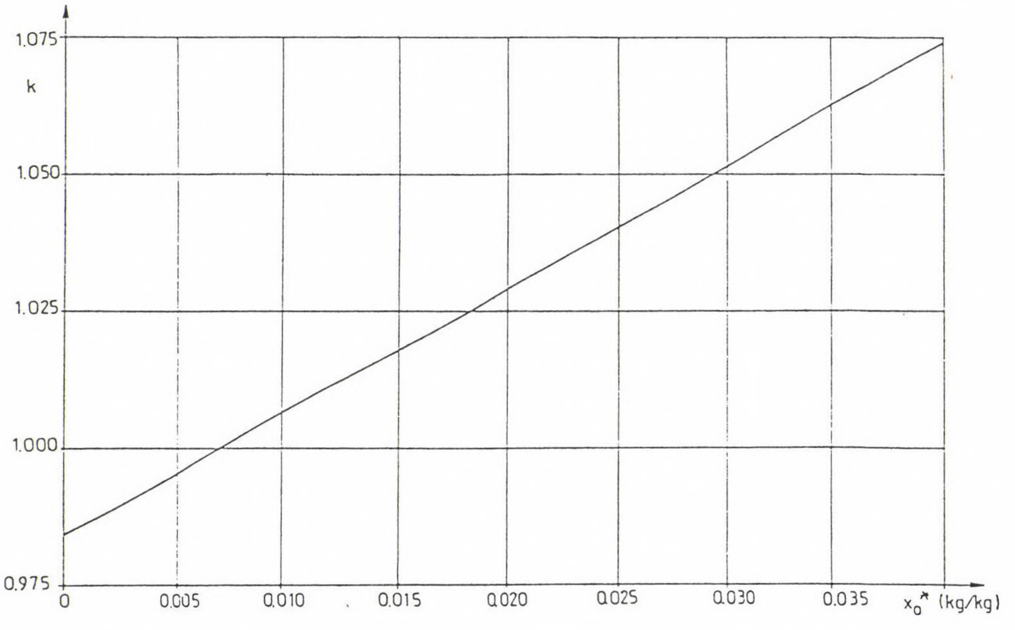


Fig. 7. Change of the value of k in function of x_0^* when $\varrho_1 = 1.25$ kg/m^3, $\varrho_2 = 1.3$ kg/m^3 and $\Delta x = 0.005$ kg/kg

This factor k (in knowledge of $\varrho_1, \varrho_2, x_0^*$ and Δx^*) shows how far the result of the calculations is influenced by the change of density, the vapour and by the vapour absorption, respectively. The more k differs from one, the more justified is that factors listed above should be taken into account when network calculations

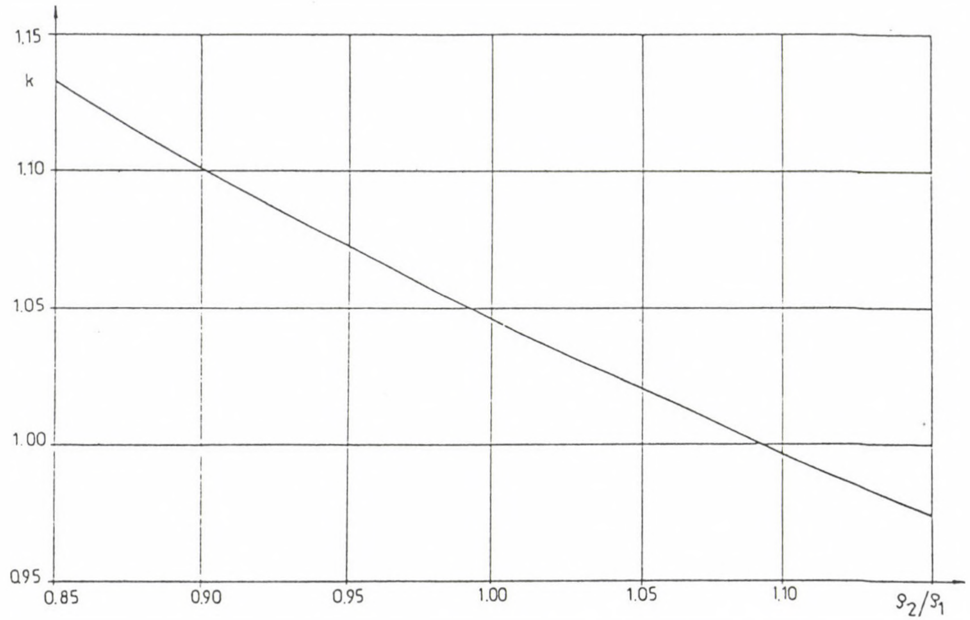


Fig. 8. Change of the value of k in function of ϱ_1/ϱ_2 when $\varrho_1 = 1.25$ kg/m^3, $x_0 = 0.025$ kg/kg and $\Delta x^* = 0.005$ kg/kg

are carried out. Figures 7, 8 and 9 are illustrate the dependence of k on the factors listed above. In all the three figures $\varrho_n = 1.25$ kg/m^3. In Fig. 7 the change of k in function of x_0^* has been illustrated. Here $\varrho_1 = 1.25$ kg/m^3, $\varrho_2 = 1.30$ kg/m^3 and $\Delta x^* = 0.005$ kg/kg.

Figure 8 shows the value of k *vs.* the quotient ϱ_2/ϱ_1. In Fig. 8 $\varrho_1 = 1.25$ kg/m^3, $x_0^* = 0.02$ kg/kg, $\Delta x^* = 0.005$ kg/kg. Figure 9 is illustrates the dependence of k on Δx^*, if $\varrho_1 = 1.25$ kg/m^3, $\varrho_2 = 1.3$ kg/m^3 and $x_0^* = 0.02$ kg/kg.

Based on these figures the value of depression caused by air friction calculated with consideration of the compression is differring max. 10 percent from that one disregarding the compression in case of parameters ϱ, x_0^* and Δx^* characterizing air flows in mines and due to their changes respectively. The influence of changes in speed due to vapour absorption on pressure difference is also perceptible, should these be compared with the depression in (23). This ratio k^* is

$$k^* = \frac{D \cdot E}{F^2 \cdot R_n}\,\varrho_n\,. \tag{25}$$

When calculating the value of k^* the data relating to the intake shaft are used as described by McPherson and Robinson (1980). These are as follows: $\varrho_n = 1.2$ kg/m^3, $\varrho_1 = 1.201$ kg/m^3, $\varrho_2 = 1.32$ kg/m^3, $R_n = 0.01407$ kg/m^7, $F = 23.9$ m^3, $x_0^* = 0.0105$ kg/kg, $\Delta x^* = 0.0007$ kg/kg. Substituting them in Eq. (25) $k^* = 0.054$. Thus the impact on the pressure difference of the speed change due to change of density and vapour absorption compared to the depression caused by air friction

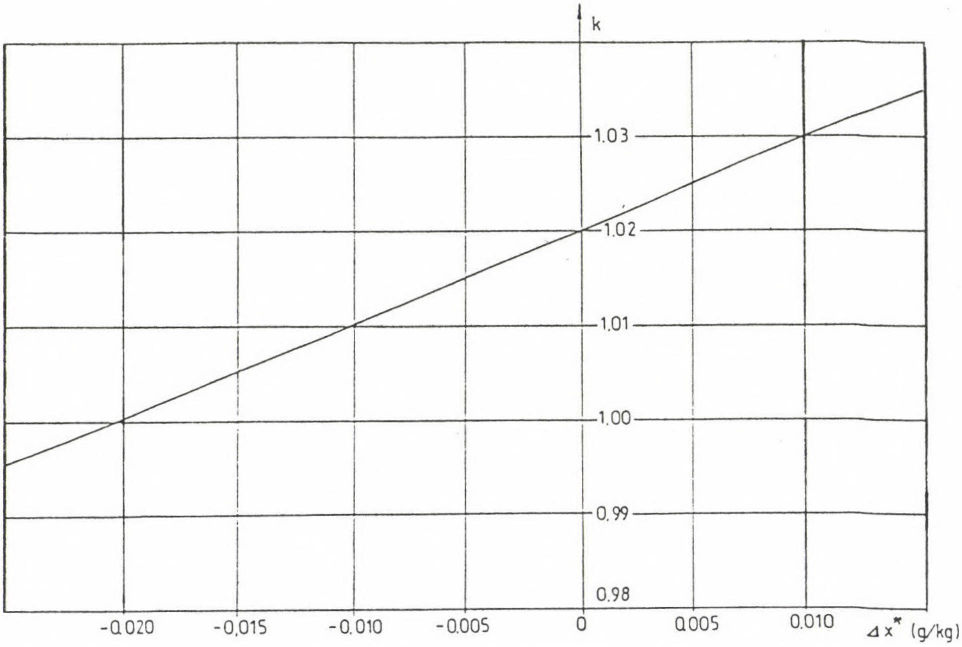


Fig. 9. Change of the value k in function of Δx^* when $\varrho_1 = 1.25$ kg/m^3, $\varrho_2 = 1.3$ kg/m^3 and $x_0 = 0.02$ kg/kg

amounts to 5.4 percent. Using the previous data when the change of vapour content is equal to 0 then $k^* = 5.44$ percent when $\Delta x^* = 0.02$ then $k^* = 4.87$ percent.

Based on the present investigations if the change of air density is considerable within the air network then not only the pressure changes in function of flowing air quantities, but also barometric pressures in the end points of air passages are to be calculated. The requirement on data of these calculations is less and calculations are simpler resp. when the change of vapour content within the air passage does not exceed 10 to 20 g/kg the air density or temperature within the air passage has a linear change in function of the length of the air passage.

In the first case when defining pressure-change it is enough if the average specific vapour content is known the change of vapour content may be neglected. In case of linearily changing densities and temperatures respectively in knowledge of the gradient of change the value of Δp can be calculated using Eqs (9, 10) and (14, 15, 17) resp. Should the condition of linearity not be valid for the total length of the air passage, then it should be devided to sections in which this condition is valid.

References

Düpre G, Hofbauer P J 1974: *Glückauf F.H.*, 35, 95–102.

Greuer R 1964a: *Bergbau-Archiv*, 25, No. 3, 39–59.

Greuer R 1964b: *Bergbauwiss.*, 11, 73–80.

Hall A E, Stoakes M A, Sangal M K 1982: CANMET'S thermodynamic ventilation network program. CIM Bulletin, 75, No. 848, 52–60.

Heise F, Drekopf K 1927: *Glückauf*, 63, 329–333.

Janositz J 1979: *BKL Bányászat*, 112

Kämpf G 1987: *Glückauf*, 123, 253–255.

Kegel 1908: *Glückauf*, 44, 118–124.

McPherson M J, Robinson G 1980: *Journal of the Mine Vent. Soc. of South Africa*, 145–164.

Pollak R 1985a: *Glückauf*, 121, 681–686.

Pollak R 1985b: *Glückauf F.H.*, 46, 87–92.

Schmidt W 1952: *Glückauf*, 88, 859–867.

Schmidt W 1955: *Glückauf*, 91, 549–568.

Wher R, Henningsen U 1970: *Glückauf*, 106, 641–648.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 331–340 (1992)

NEGATIVE PHENOMENA IN APPLYING CONCAVE SEMI-VARIOGRAMS

A FÜST[1]

[Manuscript received December 10, 1990; revised March 18, 1992]

According to experience concave semi-variograms often yield unreal, extreme estimated values in kriging. This paper intends to explain this phenomenon.

Keywords: concave semi-variogram; kriging

It is a well-known fact in geostatistics that unreal estimated values are usually yielded, if concave semi-variogram is used for kriging. The estimated value may differ with orders of magnitude from data registered in the neighbourhood. These extreme values seem to disregard any law. We try to resolve this enigma in the present paper.

1. May a concave semi-variogram exist?

The semi-variogram and the autocovariance function express mutually the environmental dependence of the examined parameter. In case of a concave semi-variogram (see Fig. 1) this dependence changes scarcely in the beginning, but as the distance increases, there is a sudden decrease, and it is equal to zero at the distance of influence.

On the other hand in case of common (convex) semi-variograms there is a suddenly decreasing dependence in the beginning, it gradually moderates and becomes equal to zero at the distance of influence (see Fig. 2). Functions without nugget-effect ($C_o = 0$) are represented in Figs 1 and 2 for the sake of simplicity. What is a parameter surface like that has a concave semi-variogram? Let us consider first the one-dimensional case *i.e.* the semi-variogram with respect to one direction. If the function $y = h^b$ ($b > 0$) describes the variability of a parameter in a given direction, the semi-variogram will be concave in every case. The greater the exponent b of the function is, the more intensive the variability of the semi-variogram will be. However, it can be observed that these semi-variograms are monotonously increasing, if $h \to \infty$, $\gamma(h) \to \infty$. It can be attributed to the trend effect. If trend is separated and the semi-variogram is calculated for the deviations from the trend, a convex function is to be obtained (in case of trend separation of corresponding degree). Nevertheless semi-variograms having more sills are often yielded (Füst 1984). The situation is similar to the two-dimensional cases. Figure 3 shows such examples

[1]Central Institute for Mining Development, POB 115, H-1300, Budapest, Hungary

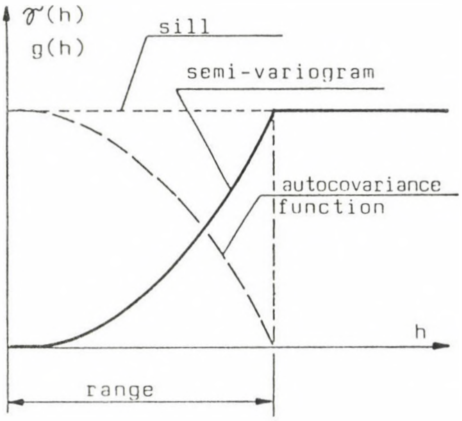


Fig. 1.

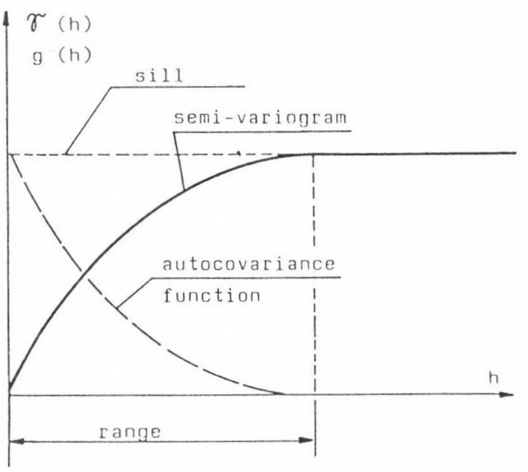


Fig. 2.

for caloric value, ash content and density of the coal mining area 'Balinka I'. Degree of the separated trend was 2 for all the three parameters (Füst et al. 1990). Practically, forms shown by Figs 4 and 5 or their combinations may occur.

Consequently, it can be established that a monotonously increasing concave semi-variogram is yielded in case of trend-like monotonously either increasing or decreasing parameter variabilities. Kriging should be applied only for deviations from the trend and the result should be superposed on the trend. The non-monotonously increasing concave semi-variograms reflect the common effect of a multisill structure, therefore interpretation and treatment of such semi-variograms require further considerations.

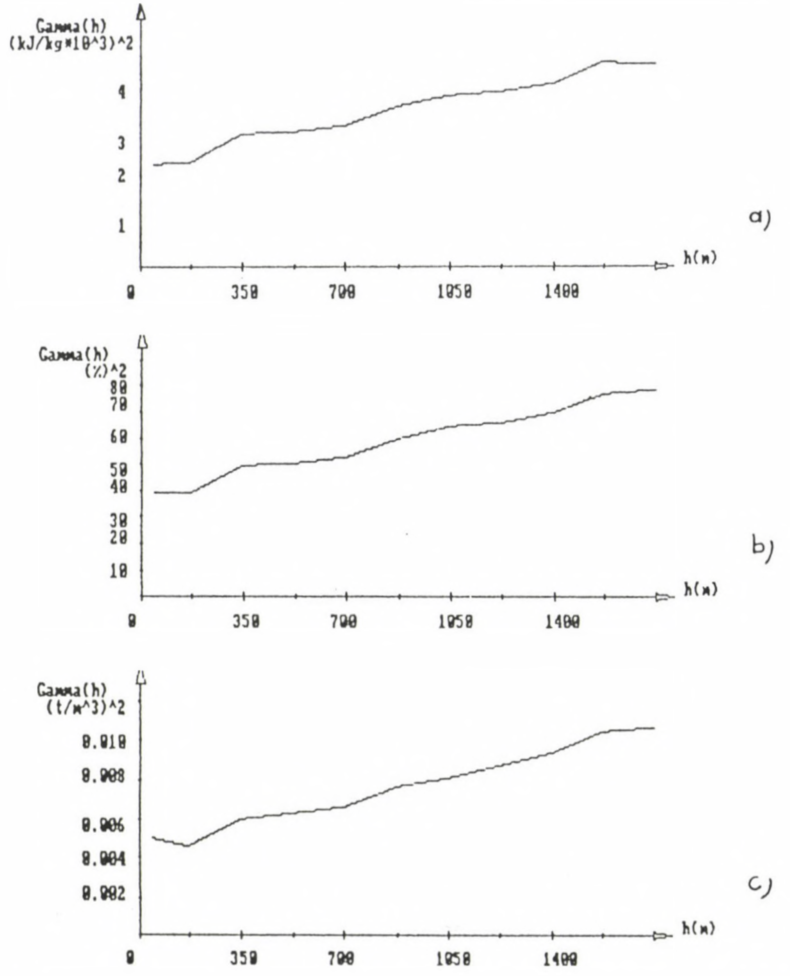
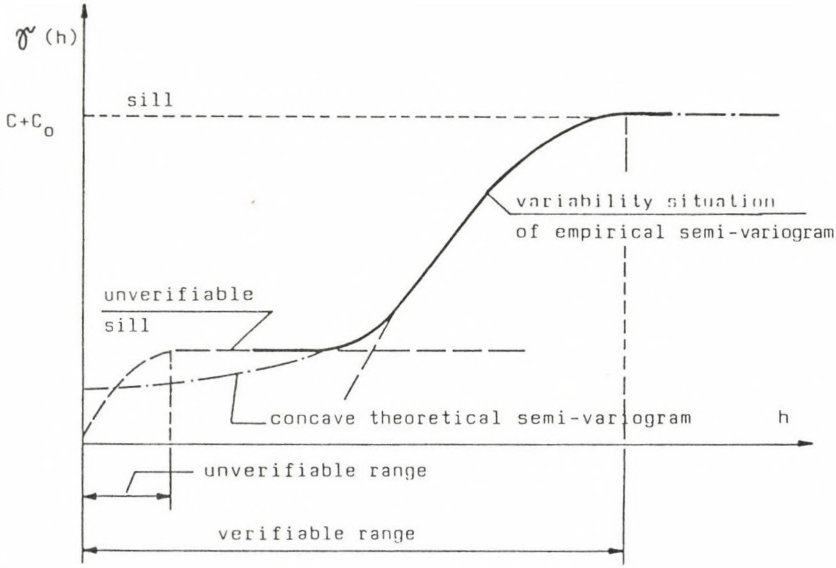


Fig. 3. Balinka-I. Empirical semi-variogram after trend separation (Matheron-type algorithm of empirical semi-variogram). a) Caloric value, b) ash content, c) volume weight

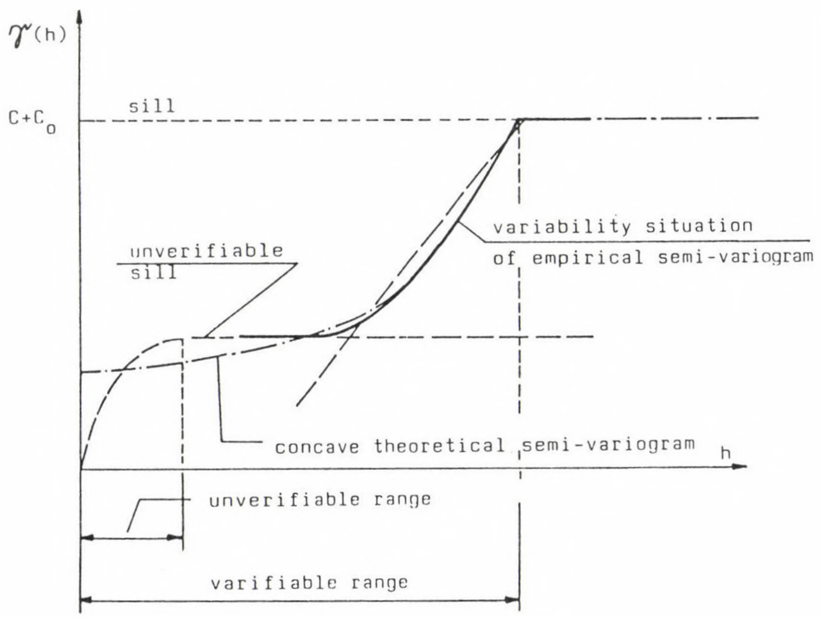
2. Treatment of concave semi-variograms having more sills

In case of two (or more) verifiable sills only the distance of influence of the greater sill (or those of the sills following the least one) is (are) known. Therefore the distance of influence belonging to the second sill should be used for kriging. But what kind of semi-variograms should be used? it seems to be evident to apply a concave model (Figs 4 and 5), but its effects are rather incalculable as we shall see later. If we approximated only the second sill, we would get a negative nugget effect with any type of convex semi-variograms. It is not possible to take into consideration the physical nature of the nugget effect. The common marginal case of concave and convex semi-variograms is the linear model. Consequently, in this case we propose



$\gamma^{\sim}(h)$
sill
$C+C_0$
variability situation
of empirical semi-variogram
unverifiable
sill
concave theoretical semi-variogram
h
unverifiable range
verifiable range

Fig. 4.



$\gamma^{\sim}(h)$
sill
$C+C_0$
variability situation
of empirical semi-variogram
unverifiable
sill
concave theoretical semi-variogram
h
unverifiable range
varifiable range

Fig. 5.

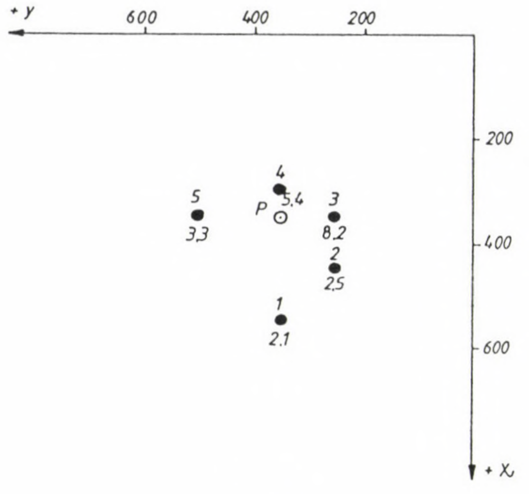


Fig. 6.

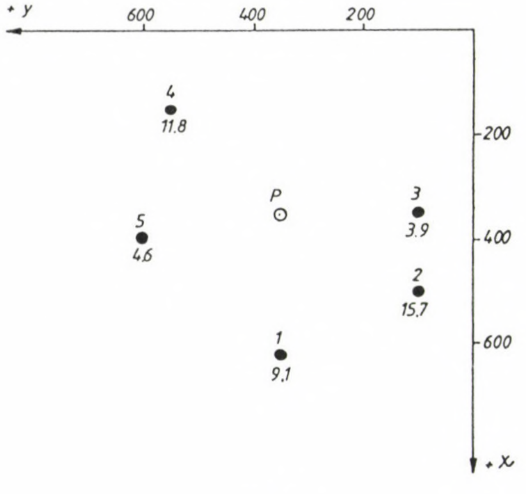


Fig. 7.

to apply such a linear semi-variogram whose nugget effect is equal to the technical measurement error of the parameter, and whose sill and range equals to the second sill and range (used for the kriging).

3. Effect of the concave semi-variogram on kriging

If the exponent increases when $b > 1$, concave semi-variograms yield kriging coefficients a_i in $(-\infty; \infty)$. They yield extreme estimation results. Negative coefficients

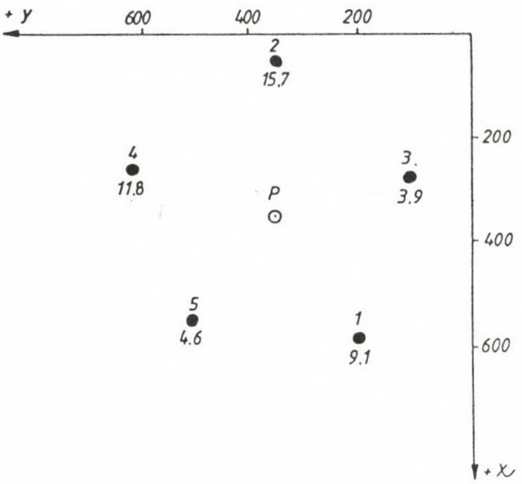


Fig. 8.

are especially dangerous. Strange shadow-effects (negative kriging coefficients) arise in the course of estimation which cannot be connected with the arrangement of the samples. We could prescribe as a further kriging condition that every weighting factor of the kriging should be positive (Molnár and Szidarovszky 1983) in order to eliminate these effects, but it is an artificial and rather biassing intervention.

According to our experiences an estimation is stable while the distances of samples are less than the range in every relation and while the exponent is not greater than 1. Nevertheless in some cases, results are acceptable if $b \leq 2$. Measure of instability (dispersion of estimation results) increases with the increasing number of matrix elements with 0 value and with the increasing exponent. The same phenomenon can be observed (though not so strongly), if the concave semi-variogram connects smoothly (with horizontal tangent) to the sill at the distance of influence (semi-variograms of Gaussian or sinusoidal type). Závoti (personal communication 1991) pointed out that not only concave semi-variograms could cause instability, it may occur because of the locations of the samples taken into account. The above statements are illustrated by three examples. Let us assume that in all the three examples the semi-variogram used for the estimation has the form:

$$\begin{aligned} \gamma(h) &= c \cdot h^b \quad 0 \leq h \leq 300 \quad (b > 0) \\ \gamma(h) &= 5.1 \qquad\quad h \geq 300 \end{aligned}$$

Table I contains the coefficients of the equation in ten variants.

Arrangements of samples in the first example are presented in Fig. 6. In this case it is true for each relation that the distances between the samples are less than the range. Arrangements of samples in the second and third example are presented in Figs 7 and 8, respectively. The sample values are the same in the second and the third examples, but in the third one we took such a location where almost every sample is out of range of any other samples, but all of them have influence to the

Table I.

Serial number of the equation (semi-variogram)	c	b
1	1.225 433 821	1.25
2	0.294 448 637	0.5
3	0.070 750 455	0.75
4	0.017	1
5	$5.6 \cdot 10^{-5}$	2
6	$1.8 \cdot 10^{-7}$	3
7	$6.296 \cdot 10^{-10}$	4
8	$2.098\ 765\ 4 \cdot 10^{-12}$	5
9	$6.995\ 884\ 8 \cdot 10^{-15}$	6
10	$2.331\ 961\ 6 \cdot 10^{-17}$	7

Table II.

Serial number of the semi-var.	Example 1		Example 2		Example 3	
	estimated value	standard deviation	estimated value	standard deviation	estimated value	standard deviation
1	4.77	2.079	8.60	2.437	8.71	2.019
2	4.99	1.704	8.09	2.394	8.41	2.015
3	5.07	1.371	7.47	2.346	8.12	2.009
4	5.06	1.084	6.71	2.290	7.86	2.002
5	7.45	0.026	1.25	1.936	6.92	1.961
6	3.99	0.527	−12.92	0.621	6.17	1.911
7	3.33	0.612	−63.96	3.382	5.58	1.860
8	2.75	0.647	1023.57	15.532	5.12	1.812
9	2.58	0.663	118.76	5.572	4.77	1.771
10	3.69	0.661	77.22	4.612	4.51	1.738

Table III.

Serial number of the semi-variogram	Kriging coefficients of example 2					
	a_1	a_2	a_3	a_4	a_5	μ
1	0.210	0.161	0.210	0.196	0.223	−0.974
2	0.218	0.118	0.232	0.189	0.243	−0.915
3	0.226	0.068	0.267	0.179	0.260	−0.848
4	0.235	0.006	0.315	0.167	0.277	−0.773
5	0.297	−0.441	0.716	0.099	0.329	−0.404
6	0.501	−1.655	1.824	−0.018	0.348	0.231
7	1.394	−6.170	5.881	−0.310	0.205	2.148
8	−19.663	91.679	−80.916	4.610	5.290	−37.229
9	−2.333	10.429	−8.733	0.386	1.251	−4.427
10	−1.626	6.788	−5.445	0.107	1.776	−2.902

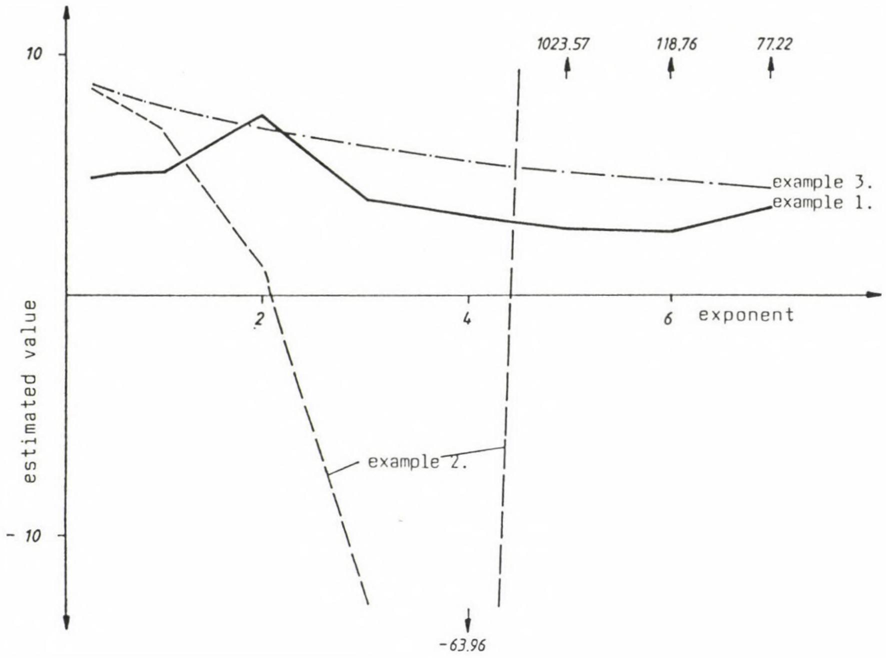


Fig. 9.

point of estimation. (In the second example this holds only for a few relations.)

Practically, the places of samples used for the estimation may be therefore modified here between certain limits without any effect on the estimated value. Table II contains the estimations and their standard deviation for the point P of the three examples. Results are graphically represented on Figs 9 and 10. These figures show that till $b \leq 1$, change of both the estimated value and its standard deviation are continuous. The following results — especially in the second example — are extreme. They can be followed by the weighting factors of kriging (Table III).

The first negative coefficient appears when $b = 2$ ($a_2 = -0.441$). When $b = 5$, extremely great negative ($a_1 = -19.663$, $a_3 = -80.916$) and positive weighting factors are yielded. The estimated values of the third example seem to be reasonable, although there is a negative weighting factor here too, if $b \geq 4$. However it is never less than -1 in the examined range (f.e. if $b = 4$, $a_2 = -0.014$ and if $b = 7$, $a_2 = -0.093$).

If extreme, unreasonable results are yielded, entries of the inverse of symmetric matrix **A** describing the effects of samples on each other can be greater with one or two magnitudes than in the normal case. In the second example if *e.g.* $b = 0.25$ (*i.e.* the variogram is convex), matrices **A** and $\mathbf{A}^{-1}$ are as follows (we give just the upper triangle because of symmetricity):

$$\mathbf{A} = \begin{bmatrix} 5.1 & 0.089 & 0 & 0 & 0 & 1 \\ & 5.1 & 0.811 & 0 & 0 & 1 \\ & & 5.1 & 0 & 0 & 1 \\ & & & 5.1 & 0.203 & 1 \\ & & & & 5.1 & 1 \\ & & & & & 0 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} 0.155 & -0.039 & -0.036 & -0.040 & -0.040 & 0.213 \\ & 0.171 & -0.063 & -0.034 & -0.034 & -0.183 \\ & & 0.169 & -0.035 & -0.035 & 0.187 \\ & & & 0.157 & -0.047 & 0.208 \\ & & & & 0.157 & 0.208 \\ & & & & & -1.104 \end{bmatrix}$$

If $b = 5$,

$$\mathbf{A} = \begin{bmatrix} 5.1 & 1.520 & 0 & 0 & 0 & 1 \\ & 5.1 & 4.941 & 0 & 0 & 1 \\ & & 5.1 & 0 & 0 & 1 \\ & & & 5.1 & 2.839 & 1 \\ & & & & 5.1 & 1 \\ & & & & & 0 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} -1.510 & 7.595 & -6.801 & 0.358 & 0.358 & -2.841 \\ & -35.039 & 31.099 & -1.828 & -1.828 & 14.509 \\ & & -27.466 & 1.584 & 1.584 & -12.573 \\ & & & 0.164 & -0.278 & 0.952 \\ & & & & 0.164 & 0.952 \\ & & & & & -7.561 \end{bmatrix}$$

4. Conclusions

On the basis of the above examples and theoretical considerations one can draw the following conclusions.

(a) Parameter changes having a trend character result in monotonously increasing concave semi-variograms.

(b) The concave semi-variograms experienced after the trend separation prove the presence of two (or more) structures, where the number of sills having calculable values is greater by one than the number of calculable distances of influence.

(c) This semi-variogram — if it is intended to be used for kriging — has to be substituted by a linear mode, whose nugget-effect is equal to the technical (measurement) error of the parameter and whose distance of influence and corresponding sill intended to be used for the estimation.

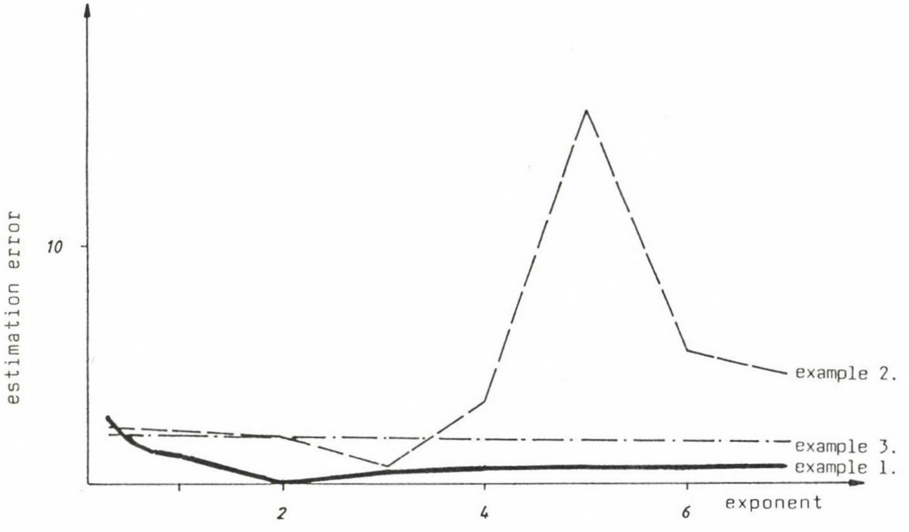


Fig. 10.

(d) During kriging we have sometimes to take into account the deforming effect coming from the location of samples. But this phenomenon occurs not only in case of concave semi-variograms.

References

Füst A 1984: Lecture Notes of the Technical University of Heavy Industry, Miskolc, Series I. Mining, 32, 93–112.

Füst A, Csóti T, Molnár S, Kun A, Lesták F, Makrai L, Börcsök L, Csóti T Jr., Markó L, Martinko M, Szabó J 1990: Geostatistical and risk analysis of seam parameters of the mining area 'Balinka I' (study)

Molnár S, Szidarovszky F 1983: *Journ. of Appl. Mathematics*, 9, 405–419.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 341-369 (1992)

ROCK MECHANICS OF VISCOELASTIC MATERIALS — A COOPERATION BETWEEN DEPARTMENT AND RESEARCH STATION[1]

B ROLLER[2]

[Manuscript received March 4, 1992]

This paper presents the work carried out in recent years at the Department of Civil Engineering Mechanics of the Technical University Budapest, concerning the altering state analysis of rock mass. A systematic overview is made about the explorations on the spot, laboratory experiments, tentative computation works carried out in planning bureaus, furthermore about the tasks of the research team and the computer center. An approximate calculation applying Laplace-transformation has been made in order to analyse structures mapped by simple bar systems, based on elementary considerations. A derivation of the compliance function of the elasto-viscoplastic material in uniaxial state of stress is given as well, also introducing the concept of a secondary elasticity modulus. Three interesting theorems are stated connecting the relationship of the model treated hereby, and the Poynting-Thomson viscoelastic model. A detailed example solved by PC is presented, too, the code of which has been developed at the Department. Finally some possibilities of the developments in the field of the research are outlined.

Keywords: laboratory experimentation; Poynting-Thomson model; rock mechanics; uniaxial stress; viscoelastic material

1. Introduction

The author and his junior partners performed a greater research on the strength of material analysis of rock continua in mines in the 70-es, on behalf of the Mining Research Station of the day, and in close connection with mining engineers. The main aim of this work was to clear up the effect of the loosening that is due to the abandoning of the drifts, on the compactness altering of the rock mass and on the displacements. Namely this effect influences not only the settlement of the terrain but also the water conducting capacity and alters the hydraulic relations in the mines as well.

The results were published at that time in several reports and papers (Roller and Kesserű 1976, Roller and Szentiványi 1976, Roller 1978).

Consequences of economical and social changes that have taken place since then have touched the researches in mining industry and science as well. Due to recessive

[1]Text of a speech held at the "1st Colloquium of Rock Mechanics" of the Central Research Station of Mining, Budapest, 1991

[2]Technical University Budapest, Dept. of Civil Engineering Mechanics, H-1521 Budapest, Műegyetem rkp. 3, Hungary

effects, the topic could have been investigated by a comparatively modest apparatus, so the formal cooperation stood still. The staff of the department dealt with detail questions according their own interest and turned rather to basic theoretical research (Roller 1980, 1989, 1990). Beside the problems of practical development work they have been busy with those problems of computer analysis and CAD of architecture, civil and structural engineering rather then questions arising in mining (Roller 1980, Kurutz and Kovács 1981). The research work has been extended even to the domain of geotechnics and biomechanics (Bojtár and Bagi 1989, Bojtár et al. 1990).

It seems that rearrangement of the research activity that takes place in a country-wide relation as well as the renovation of the Central Developing Institute of Mining presents an occasion to refresh a cooperation which was formerly quite successful. Therefore it is actual to look over the participants of both the research and the developing work, to summarize their aims and recent problems, furthermore present status of the mechanical analysis and further possibilities.

2. Participants of the rock mechanics investigation

The investigation must be based first of all on the real properties of the material constituting the rock mass, since withouth their knowledge any calculation is irrealistic. The structure of the strata and the relations of cracking are very important besides the strength properties. Therefore operations must start out from the spot of the mining work, by the exploration of the strata, respectively by tests of strength *in situ*. So the first potential participant is the *mining workshop*, that supplies this task expediently in cooperation with the research station.

The next phase is laboratory testing of materials, partly by short time loading tests, partly by permanent loading of the rock sample specimens. Besides it is necessary to carry out model tests which refer to the plane state investigation of complete cross section slices of the mass. This task belongs mostly to the *laboratory of the research station*, while researchers at the university may help here mainly in the completion of the theoretical methods of the numerical evaluation of the experimental results (Popper and Csizmás 1987).

Thereafter it is very practical to develop computation procedures that may possibly be based on serious mathematical concepts, still they could be carried out by simple derivations, applying short calculations using pocket calculators. The model is quite rough, the aim is an information at the level of a diagrammatical plan performed by a work that has to be arranged on the desk of the designer. This computation must detect the range of those data intervals that are necessary to be selected in order to make more extended computer simulated experiments, completing the laboratory information. The work is aided by offering skilled methods which can be applied according to instructions even by the *projecting engineers*, as well.

The next task of the *basis of computation and research active at the university* is to work out the algorithm and computer programs of those questions having already quite elaborated theories. Some systematizing work is necessary in this field, too, but the results of the former researches indicate acceptable success in this domain, as well. Parts of the investigation complex as proper problems are

i) Compilation of appropriate constitutive laws

ii) Establishing multi-degree of freedom computation models, with particular respect to the heterogeneous character of the rock mass, also to the presence of oblique strata, to the appearance of foldings *etc.*

iii) Detection of building, fastening, locking and even collapsing periods by computations that simulate experiments

iv) Development of particular computation techniques with special respect to the enormous amount of unknown quantitites, also to the permanent altering of the investigation network.

Another group of the research tasks is connected to problems which need a lengthy investigation from theoretical as well as from practical aspect. These belong partly rather to the field of physics than to engineering, partly they cannot be considered as solved questions, mainly because of the immense number of unknowns that must be introduced into the analysis during the practical application. Such problems are

i) Seepage analysis at the presence of threshold gradients (Roller 1979)

ii) Influence of the altering of water conducting capacity caused by modification of the state of stresses and deformations, due to loosening and collapse of the rock mass, thereafter catastrophic inruption of the water into abandoned mine holes. The hydrophysical and hydraulical aspects of this problem were exhaustively and successfully dealt with in the Central Developing Institute of Mining (Roller and Kesserű 1976).

iii) Altering state analysis of coupled systems. A problem belonging to this complex is the clearing up of the stress, strain and current relations of continua, consisting of solid skeleton and pore water (incidentally gas phasis, too), suffering primary as well as secondary consolidation. Also the heat conduction is a familiar problem here and interesting questions arise of taking into account occasionally the change of the physical conduciton during the seepage occurring in the porous rock (Simoni and Schrefler 1989).

iv) Therorems, first of all variational statements have to be declared according to the materials that have complex behaviour, complicated structure or are built up by giant moleculae, for example viscoelastic material, strain hardening yielding plastic material, loosing, cracking, and locking material, maybe with openings and closing joints (Roller 1984, Kurutz and Kovács 1981). Elaborating of new mathematical concepts, eliminating the drawbacks of the difficulties of the customary time stepping and finite element techniques.

Some of the questions mentioned in i–iv are in fairly crude stage, so it is not too reasonable to plan their solution in the frame of either a research team work or a department project. Instead we have to count for the interest, endeavour and unselfishness of the *individual scientists.*

Finally an up to date data preparing for the computers, also the lucidly arranged presentation of the results is highly important. The organization and the mechanization of appropriate input-output systems like this is the task of software developing engineers in the computer centers, still the *Department that is prepared to the computing* problems can play the role of virilist in this activity. Developing of interactive computation and planning methods, then the organizing of the work at the monitors of the personal computers by selecting automatic menues, also by preparing electronic drawings by means of a "mouse" or a light pencil are particular problems to be dealt with here. So the up-to-date design is the constructing of the structural elements and setups, as well as network arrangements by the computers themselves, rather then by traditional methods. The results can be presented instead of the former tedious and baffling tables either in the form of coloured contour maps or as perspective figures which can be rotated up to the demand of the user, also being visualised by forms and sights suited to several points of view and looking angles (Kirchner 1989).

3. Presentation of elastic, plastic, viscoelastic and complex phenomena in the Freudenthal-space

The coordinate system at the same time the space spanned by the axis t, σ, ε, proposed forst by Freudenthal is fairly appropriate in order to systemise and summarize

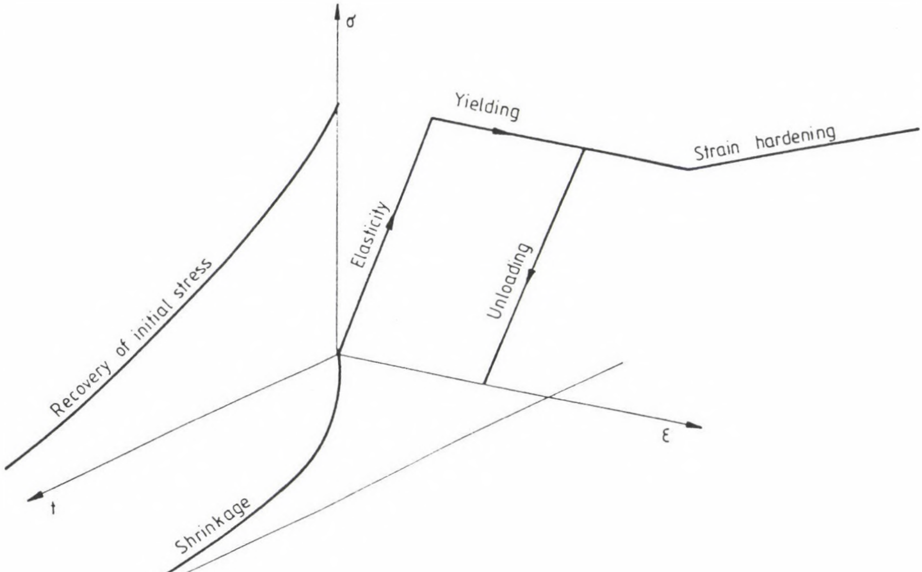


Fig. 1.

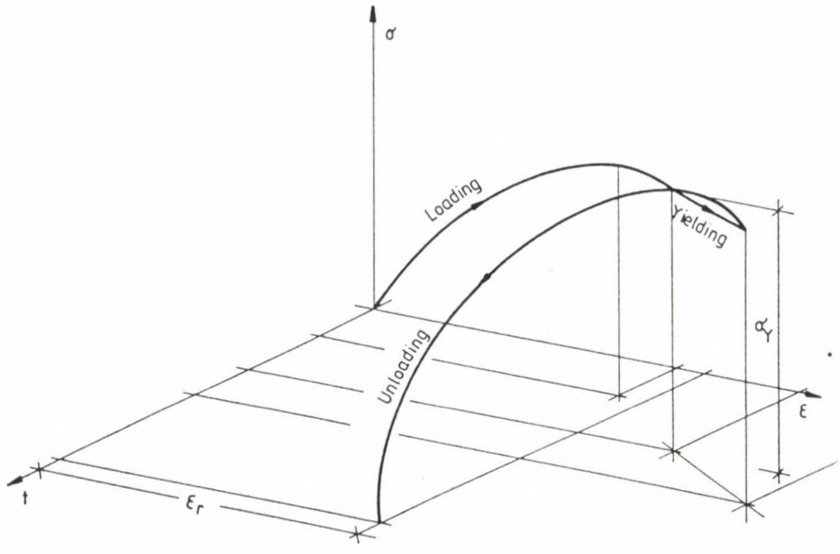


Fig. 2.

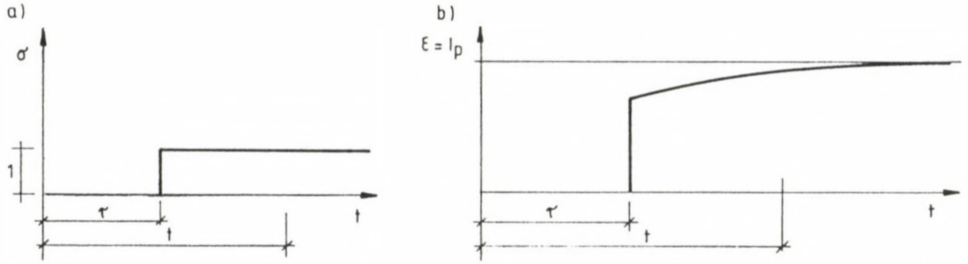


Fig. 3.

all of the problems connected with the phenomenological behaviour of the material. Detailed information about this space is published by Roller (1990).

Applying this representation we can clearly understand that the equilibrium path develops generally by simultaneous occurrence of particular phenomena, still the most basically important events of engineering practice, *e.g.* the elastic, the ideally elasto-plastic, the fully plastic behaviour as well as the strainless reduction of material stresses, the unhindered shrinkage *etc.* can be described by special functions of just a single variable (Fig. 1). If the state is a compound one, superposition holds only to the increments of the variables, so in this respect the Prandtl-Reuss

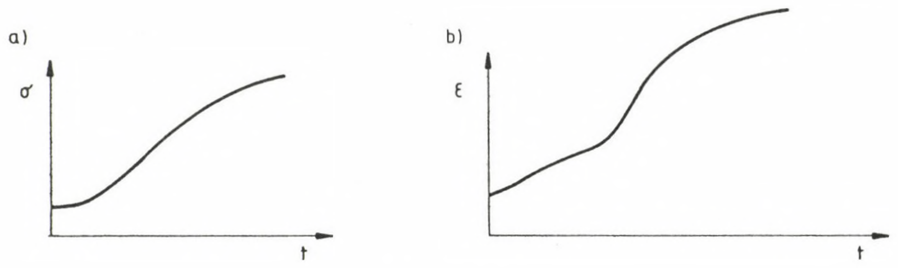


Fig. 4.

theory of the plastic deformation states just the same sentences as the theory of functions with several variables (Kaliszky 1989).

Time is a non-negative independent variable in this Freudenthal-space. The time process of the plastic deformation and the viscosity are in connection according to meaning, still the former phenomenon is characterized by the rearrangement of the stresses that are a basically permanent level, rather than the increasing of the stresses. Unloading leads back to the coordinate plane t, ε belonging to the unstressed situation, still it can be definitely seen in the diagrams that some finite time is needed to the final rise of the remanent strain (Fig. 2).

Planes of the space being particularly arranged, also the families of curves that can be drawn by selecting the values $t = t_1, t_2, \ldots, t_n = \text{const}$ — are not shown in the figures, still they can illustrate the aging of the material.

4. Tasks connected with the testing of materials. Investigation of material samples, parts of continua and complete rock mass

The first aim of the laboratory test is to discover whether the material of the rock is visco-elastic, or visco-elasto-plastic, moreover whether the viscoelastic part of the altering state is linear or nonlinear, finally whether the aging, or — in case of a fabricated building material — the effect of resting of the sample is of considerable importance. After performing the experiments, the constitutive equation has to be determined by making use of the test diagrams. The experiments refer either to uniaxial or to triaxial compression tests. Their traditional performing method is permanent loading, when the specimen is subjected to the effect of a system of forces, having a permanent distribution in time and also a distribution infinite in principle. Practically this experiment lasts till a time when either the delayed deformation stops at a permanent level or the strain becomes rapid, for instance in consequence of a tertiary creep which can be explained by the fatigue of the material.

In case of permanent loading, if the creep turns out to be linear, the creep compliance function will be determined starting out of an appropriate level of the stress, selected as a unit (Fig. 3). Generally this function has a single variable besides a parameter. The variable is time, while the parameter is the instant of the loading.

Describing an input stress

$$\sigma(t,\tau) = \Delta(t-\tau) = \begin{matrix} 0 \\ 1 \end{matrix} \quad \text{if} \quad \begin{matrix} t < \tau \\ t \geq \tau \end{matrix} \tag{1}$$

the strain respectively creep compliance

$$\varepsilon(t,\tau) \equiv J_p(t,\tau) \tag{2}$$

is measured. Having done this, we can describe the strain in case of a stress process which is shown in Fig. 4 and has an arbitrary distribution, by the superposition integral

$$\varepsilon(t) = J_p(t,t)\,\sigma(t) + \int_0^t \frac{dJ_p(t,\tau)}{d(t-\tau)}\sigma(\tau)\,d\tau. \tag{3}$$

The first term of the right side stands here for the actual effect, while the second for the after effect of the loading history.

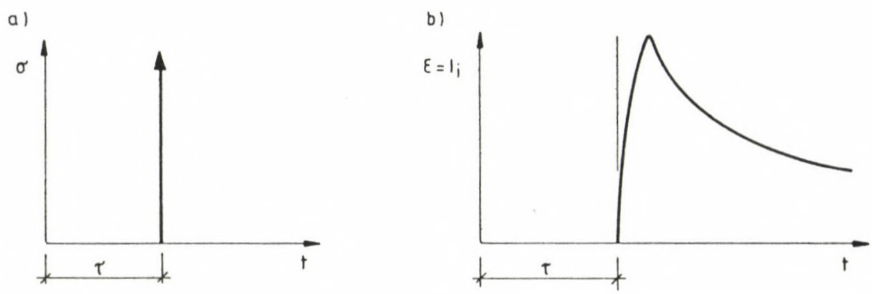


Fig. 5.

Provided the ageing is not considerable and the material can be supposed to behave as time invariant, one can write particularly

$$\varepsilon(t) = J_p(0)\sigma(t) + \int_0^t \frac{dJ_p(t-\tau)}{d(t-\tau)}\sigma(\tau)\,d\tau\,. \tag{4}$$

Impulsive loading means that the specimen carries a short time and quite intensive load. The distribution of this effect can be described as

$$\sigma(t,\tau) = \delta(t-\tau) \tag{5a}$$

where

$$\delta(t-\tau) = 0 \quad \text{if} \quad t \;\; \tau, \quad \text{besides} \quad \int_{-\infty}^{\infty} \delta(t-\tau)\,d\tau = 1, \tag{5b}$$

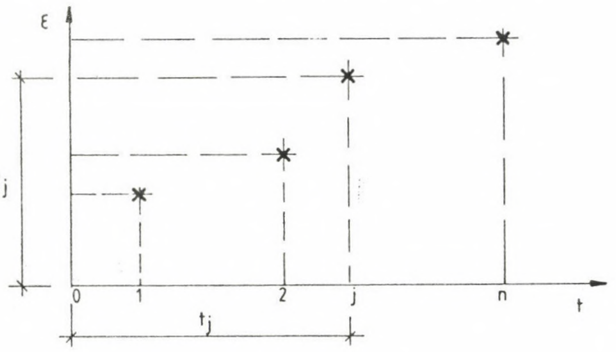


Fig. 6.

representing the Dirac-delta.

The strain due to a loading process like this is denoted by

$$\varepsilon(t,\tau)=J_i(t,\tau) \tag{6}$$

(Fig. 5). Now the deformation shown in Fig. 4, which is a consequence of a stress process of arbitrary distribution, can be described by applying Eq. (6), too, especially in the form of a superposition integral. Hence if we know the diagram presented in Fig. 3b, respectively the equation of the curve,

$$\varepsilon(t)=\int_0^t J_i(t,\tau)\sigma(\tau)\,d\tau \tag{7}$$

can be calculated, as well.

Equations (4) and (7) can be converted into each other since the input as a cause and the resulted outputs agree with each other in both cases. The advantage of impulsive loading is that in this case the test carried out is a more rapid one, also the mortgage of the laboratory is smaller, thus each calculation can be built up onto much more measurement data than in case of the permanent loading. Its drawback is, on the other hand, that it needs a more refined, also a more expensive experimental device than the conventional method.

The results of the laboratory tests are automatically presented by monitors of digital devices as coordinates of measurement points, while the analogous automatic device produces continuous characteristic curves. The first case is illustrated in Fig. 6, presenting the results of a single test. To fit to the points a compliance curve which is optimal in a proper sense, also to determine the analytical equation of a curve like this is a task that is customary and belongs rather to the team of the theoretical researchers as well as to the computer center than to the laboratory. Up-to-date methods dealing this topic belong into the domain of parametric identification. One of the appropriate possibilities is the formal application of the exponential series proposed by Dirichlet in such a way as to calculate the coefficients by nonlinear adjustment. Then the shape of the function looked for is

$$\varepsilon(t) = A + \sum_{i=1}^{m} B_i \exp(-C_i t) \tag{8}$$

$A, B_1, B_2, \ldots, B_m,\ C_1, C_2, \ldots, C_m$ denote unknown constants. The graph of the function $\varepsilon(t)$ must pass through the points

$$t_1, \varepsilon_1; \quad t_2, \varepsilon_2; \ldots t_j, \varepsilon_j; \ldots t_n, \varepsilon_n \tag{9}$$

thus the set of equations

$$A + \sum_{i=1}^{m} B_i \exp(-C_i t_j) = \varepsilon_j \qquad j = 1, 2, \ldots, n \tag{10}$$

which is nonlinear with respect to its unknowns, must be satisfied in case of all the n measurements. As a simple example, we select $m = 2$ and $n = 5$ wiriting

$$\begin{aligned}
A + B_1 \exp(-C_1 t_1) + B_2 \exp(-C_2 t_1) &= \varepsilon_1 \\
A + B_1 \exp(-C_1 t_2) + B_2 \exp(-C_2 t_2) &= \varepsilon_2 \\
A + B_1 \exp(-C_1 t_3) + B_2 \exp(-C_2 t_3) &= \varepsilon_3 \\
A + B_1 \exp(-C_1 t_4) + B_2 \exp(-C_2 t_4) &= \varepsilon_4 \\
A + B_1 \exp(-C_1 t_5) + B_2 \exp(-C_2 t_5) &= \varepsilon_5 .
\end{aligned}$$

Since $n > m = 1$, the system of equations is overdetermined thus it has no definite solution. An approximate set of resulting elements is to be looked for by the methods of adjustment.

The Department carried out some research about the development of this topic (Popper and Csizmás 1987). The essence of the results has been that the usual method of solving a determinate, nonlinear system of equations can be extended by a proper rearrangement of the algorithm. At the same time $m + 1$ approximate solution vectors are selected, then from them another, $(m+2)^{\text{th}}$ one is combined, by the help of the equation, applying appropriate weights. Starting the next step we delete the approximate solution vector loaded by the greatest error of solution, this latter fact is indicated by some suitable norm. Then the calculation of the weights and the updated approximation vector have to be repeated and so on.

The next task of the laboratory is the investigation of the rock mass being in plain strain state by applying a model that is able to imitate the presence of the cave and also the effective loading, supplied at the same time basically with plain strain relations.

Formerly experiments like this were performed by photoelastic techniques, still it seems more expedient to apply materials that are closer to the real rock mass, at the same time to carry out a direct investigation by proper means of mechanics. Latest results in this field were published recently by Gajári (1990). The displacements have been indicated in that text by the altering of the shape of a rectangular raster which can be seen by unaided eyes. Moreover Gajári investigated the stiffening effect of the anchor drilled into the rock, too. The hindrance of the technique is that the calculation of real stresses is much more complicated than in case of the photoelastic test.

The most appropriate and realistic, still very expensive laboratory method is to carry out part scale spatial models. This concept is represented in Middle Europe most excellently by the Institute and Research Station ISMES, *Bergamo*, Italy. In such cases there is even possible to imitate the rock mass according to the request of the contractor by a set of stiff blocks. It is a pity that Hungary is not prepared to jobs like this at the time being.

5. Analysis made at the level of diagrammatical planning

The fundamental idea of simple approximating analysis has been proposed by v. Mises, related to the investigation of bars made of ideally elastic — fully plastic material, at the early 20th century (Kaliszky 1989). Basically, the behaviour of a symmetrical truss model consisting of three bars can be quite simply described during the process of loading, yielding and unloading, at the same time information is obtained about the properties of more complicated structures or bodies. The Mises-model will be extended hereby to structures shown in Fig. 7 but being of

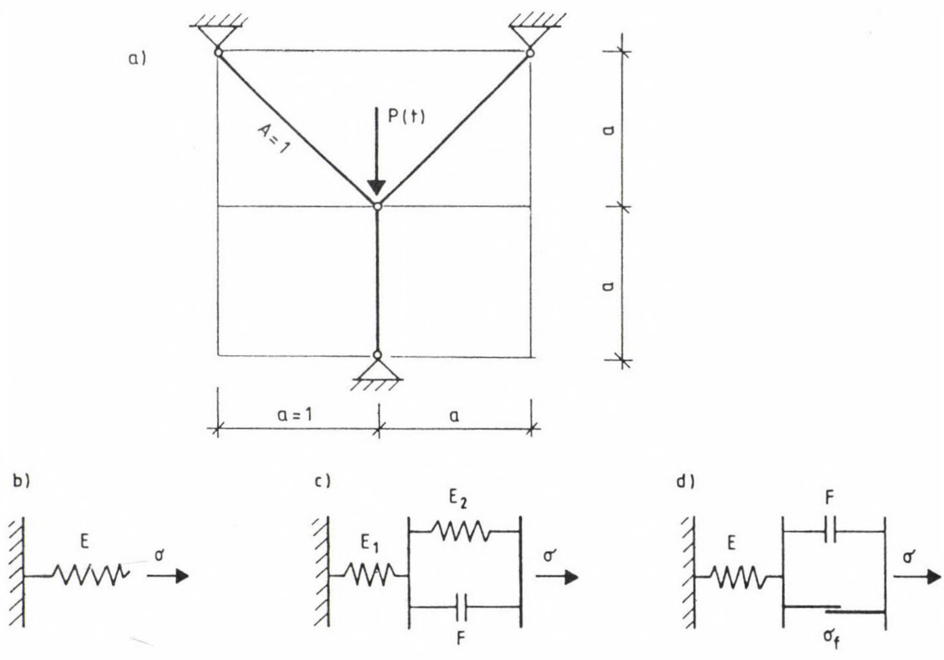


Fig. 7.

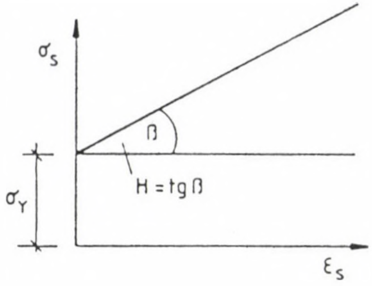


Fig. 8.

viscoelastic material. Length and cross section measures have been selected as unity for sake of simplicity.

The problem is the solution of the structure sketched in Part *a* of the figure, in case of particular viscoelastic constitutive properties illustrated in Part *b*.

The equilibrium condition reads as

$$2S_1 - S_2 - P = 0 \qquad P = P_0\Delta(t)\,. \tag{11a}$$

The geometrical equation is

$$\Delta s_1 = -\frac{1}{2}\Delta s_2\,. \tag{11b}$$

The constitutive law valid for the skew and the vertical bars is stated in the form of integral equations

$$\begin{aligned} \Delta s_1 &= (A_1 - B_1)S_1(t) + \int_0^t \exp[-C_1(t-\tau)]S_1(\tau)\,d\tau \\ \Delta s_2 &= (A_2 - B_2)S_2(t) + \int_0^t \exp[-C_2(t-\tau)]S_2(\tau)\,d\tau\,. \end{aligned} \tag{11c}$$

Applying the solution technique of the displacement method, by reducing the above formulas, the deflection of the node is obtained as

$$v(t) = A + B\exp(s_1 t) + C\exp(s_2 t) \tag{12}$$

A, B and C being constants to be determined by a lengthy calculation as

$$A = \frac{1}{s_1 s_2} \quad B = \frac{1 + q_1 s_1 + q_2 s_2^2}{s_1(s_1 - s_2)} \quad C = \frac{1 + q_1 s + q_2 s^2}{s_1(s_1 - s_2)} \tag{13}$$

$$s_1 < 0, \qquad s_2 < 0$$

where s_1, s_2 are characteristic roots while q_1 and q_2 are material constants of the Poynting-Thomson model.

The bar force of the strut in compression is

$$S_2(t) = -(E_2 - F_2)\,v(t) + \frac{H_2B}{G_2+s_1}[\exp(s_1t)-1] + \frac{H_2C}{G_2+s_2}[\exp(s_1t)-1] \qquad (14)$$

The expression of the force $S_1(t)$ is exponential as well. Constants A_1 to H_2 are functions of the parameters q_1, q_2.

If the load is not stationary, the process-functions of the state variables have to be determined by superposition as follows

$$\begin{aligned} v(t) &= \int_0^t v_I(t-\tau)\,dP(\tau) = \int_0^t v_I(t-\tau)\,\dot{P}(\tau)\,d\tau \\ S_1(t) &= \int_0^t S_{1,I}(t-\tau)\,dP(\tau) = \int_0^t S_{1,I}(t-\tau)\,\dot{P}(\tau)\,d\tau \end{aligned} \qquad (15)$$

$()^{\cdot}$ stands for differentiation with respect to τ.

The use of the above formulas is quite simple. Selecting a broad spectrum of input data, many informative results are obtained concerning the relations of order of magnitude among the material parameters and the altering state variables.

6. Problems concerning the constitutive equation

The state of stress as well as the state of strain of the rock mass are triaxial states. Still the determination of the constitutive laws is based on the relations that are expired in the uniaxial state, besides the experiments to be performed in this latter case are far more simple than those necessary in the triaxial one. Thus both the laboratory and the theoretical efforts are supported by the analysis of the uniaxial state of stress. The constitutive equations of this discipline have been

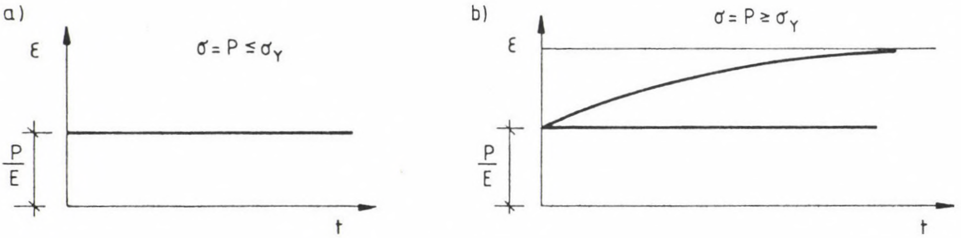


Fig. 9.

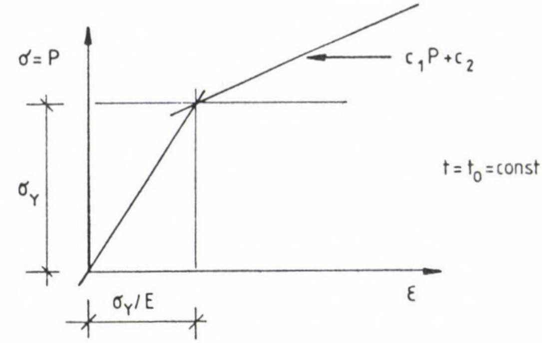


Fig. 10.

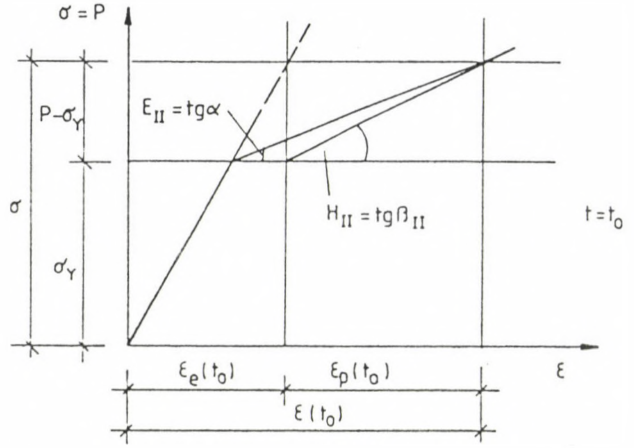


Fig. 11.

established by Eqs (2) and (6) as well, so statements are to be completed just by the particular material laws of the time invariant materials. Now the formula

$$J(t,\tau)=J(t-\tau) \tag{16}$$

is valid, hence every compliance function can be generated by a parallel shifting of the function $J(t-0)=J(t)$ along the time-axis t. Time invariant materials are described by spring and dashpot models, the elasticity moduli and viscosity coefficients of which, respectively those of their basic elements are constants. These models can be represented beside of the creep compliance functions also by ordinary differential equations with constant coefficients, too. As a typical example the Poynting-Thomson model which is shown in part c of Fig. 7 is mentioned.

The differential equation of this model is

$$\sigma+p_1\dot{\sigma}=q_0\varepsilon+q_1\dot{\varepsilon} \tag{17}$$

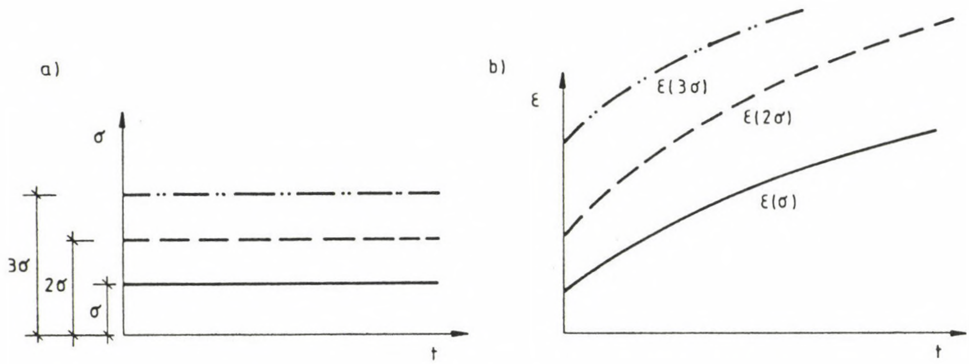


Fig. 12.

having the parameters p_1, q_0 and q_1 (Flügge 1975). This model is very familiar in the mechanics of mining.

The time invariant form of the superposition material law reads as a special case of Eq. (3) as

$$\varepsilon(t) = J_p(0)\sigma(t) + \int_0^t \frac{dJ\,(t-\tau)}{d(t-\tau)}\sigma(\tau)\,d\tau \tag{18a}$$

or instead of Eq. (7)

$$\varepsilon(t) = \int_0^t J_i(t-\tau)\,\sigma(\tau)\,d\tau. \tag{18b}$$

In case of triaxial state of stress we have to declare separate constitutive laws in order to describe the effect of the spherical as well as the deviatoric part of the stress tensor. Particularly, for instance having incompressible material, a single constitutive equation is enough, in such a case

$$\varepsilon_x = \frac{2}{3}\mathcal{P}\sigma_x \qquad \varepsilon_y = \frac{1}{3}\mathcal{P}\sigma_x \tag{19}$$

with

$$\mathcal{P} = p_0() + p_1()^{\cdot} + p_2()^{\cdot\cdot} + \ldots + p_n()^{(n)} \tag{20}$$

$()^{(n)}$ stands for the n-th derivative with respect to the time. Equation (20) is an n-th order differential operator with constant coefficients, containing $n+1$ parameters (Flügge 1975).

Considering triaxial state of stress, expecially applying an integral equation for the material law, a single relationship is sufficient, as well. This is described by a second order, two subscript stress tensor and a fourth order compliance (respectively kernel) tensor, in the following form

$$\sigma_{ij}(t)J_{ghij}(0) + \int_0^t \sigma_{ij}(\tau)\frac{dJ_{ghij}(t-\tau)}{d(t-\tau)}\,d\tau = \varepsilon_{gh}(t). \tag{21}$$

A very important, interesting and particular case of elasto-viscoplastic material is represented by a model combined of a spring, a dashpot and a slider shown in Fig. 7d. Next we deal in detail the constitutive equation of this model, for its analysis leads to quite special and characteristic statements.

The aim of the following investigation is to establish a creep compliance law to the case of the model shown in Fig. 7d (Owen and Hinton 1980). Due to the presence of the slider this relationship is nonlinear, since one constitutive law is valid at low stress level and another at a higher one.

The analysis begins with stating the equilibrium condition. Mind first a section through the spring in the figure. Clearly

$$\sigma = \sigma_e \tag{22}$$

(σ_e for the stress in the elastic spring).

Select now a section through the dashpot and the slider

$$\sigma = \sigma_d + \sigma_s \tag{23}$$

(σ_d — stress of the dashpot, σ_s — stress of the slider).

Also

$$\sigma_e = \sigma_d + \sigma_s. \tag{24}$$

Concerning geometrical compatibility, one has as complete strain

$$\varepsilon = \varepsilon_e + \varepsilon_{vp} \tag{25}$$

(ε_e — strain of the spring, ε_{vp} — strain of the delaying element), since the two parts of the model are coupled in series. In particular, the viscoplastic part contains a dashpot and a slider coupled parallel, therefore these latter elements suffer a common strain:

$$\varepsilon_{vp} = \varepsilon_d = \varepsilon_s \tag{26}$$

(ε_d ... strain of the dashpot, ε_s ... strain of the slider).

Now consider the constitutive relationships of the particular elements:

Material law of the spring is

$$\sigma_e = E\varepsilon_e. \tag{27}$$

That of the dashpot is

$$\sigma_d = F\frac{d\varepsilon_d}{dt}. \tag{28}$$

The rule concerning the slider is

$$\begin{aligned} \varepsilon &= 0 && \text{if} \quad \sigma_s \le \sigma_y \\ \varepsilon_s &= \frac{\sigma_s - \sigma_y}{H} && \text{else} \end{aligned} \tag{29}$$

σ_y denotes the yield stress of the represented material, H stands for a hardening modulus (see Fig. 8).

The first statement about the constitutive equation is that in case of low stress level the behaviour is elastic:

$$\sigma \leq \sigma_y \qquad \sigma = E\varepsilon_e = E\varepsilon \,. \tag{30}$$

At the same time

$$\varepsilon_d = \varepsilon_s = 0 \,.$$

The determination of the behaviour due to a higher stress level is far more complicated. Consider the complete strain:

$$\varepsilon = \varepsilon_e + \varepsilon_{vp} = \frac{\sigma_e}{E} + \frac{\sigma_s - \sigma_y}{H} = \frac{\sigma}{E} + \frac{\sigma_s}{H} - \frac{\sigma_y}{H} \,. \tag{31}$$

As

$$\sigma_s = \sigma - \sigma_d \,, \tag{32}$$

one gets

$$\varepsilon = \left[\frac{1}{E} + \frac{1}{H}\right]\sigma - \frac{\sigma_d}{H} - \frac{\sigma_y}{H} \,. \tag{33}$$

Replacing the constitutive equation of the single dashpot:

$$\left[\frac{1}{E} + \frac{1}{H}\right]\sigma - \frac{\sigma_y}{H} = \varepsilon + \frac{F}{H}\frac{d\varepsilon_d}{dt} \,. \tag{34}$$

Now ε_d is eliminated by applying Eqs (25) and (26)

$$\varepsilon_d = \varepsilon_{vp} = \varepsilon - \varepsilon_e \,. \tag{35}$$

Also

$$\varepsilon_d = \varepsilon - \frac{\sigma}{E} \,. \tag{36}$$

Therefore, denoting ordinary derivatives with respect to the time by a dot like $(\,)^{\dot{}}$,

$$\dot{\varepsilon}_d = \dot{\varepsilon} - \frac{1}{E}\dot{\sigma} \,. \tag{37}$$

Inserting this result into Eq. (34) we obtain after rearranging

$$\left[\frac{1}{E} + \frac{1}{H}\right]\sigma + \frac{F}{EH}\dot{\sigma} - \frac{\dot{\sigma}_y}{H} = \varepsilon + \frac{F}{H}\dot{\varepsilon} \,. \tag{38}$$

This is the constitutive equation of the four parameter elasto-viscoplastic model.

Differential Eq. (38) can be written in a more comprehensive canonical form by applying the notations listed below

$$\frac{F}{E+H} = p_1 \quad \frac{E}{E+H} = k \quad \frac{EH}{E+H} = q_0 \quad \frac{EF}{E+H} = q_1 \,. \tag{39}$$

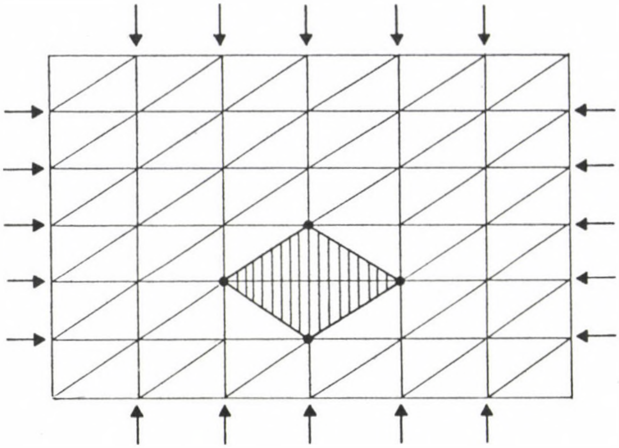


Fig. 13.

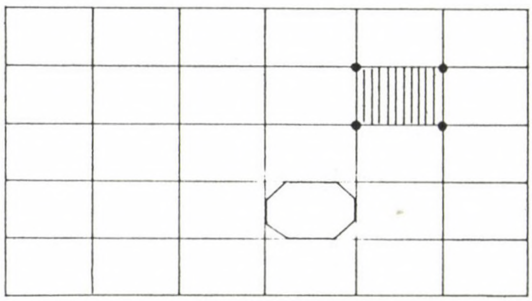


Fig. 14.

Rearranging Eq. (38) and using Eqs (39) we obtain

$$\sigma + p_1\dot{\sigma} - k = q_0\varepsilon + q_1\dot{\varepsilon}\,. \tag{40}$$

While deriving the creep compliance formula, by the definition of this concept, the input stress is selected as

$$\begin{aligned} &\sigma = P\Delta(t) \\ &p = \text{const} \qquad \Delta(t) = \begin{matrix} 0 & & t < 0 \\ & \text{if} & \\ 1 & & t \leq 0 \end{matrix} \end{aligned} \tag{41}$$

and the function $\varepsilon = \varepsilon(t)$ is looked for.

In order to determine $\varepsilon(t)$ Laplace-transformation is to be applied on both sides of Eq. (40) and inserted the Laplace-transform of Eq. (41) into the result. Thus,

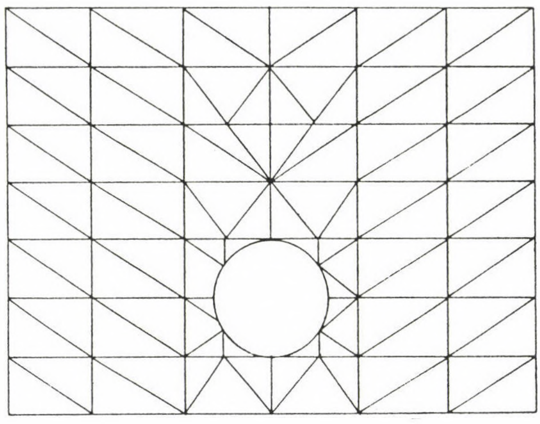


Fig. 15.

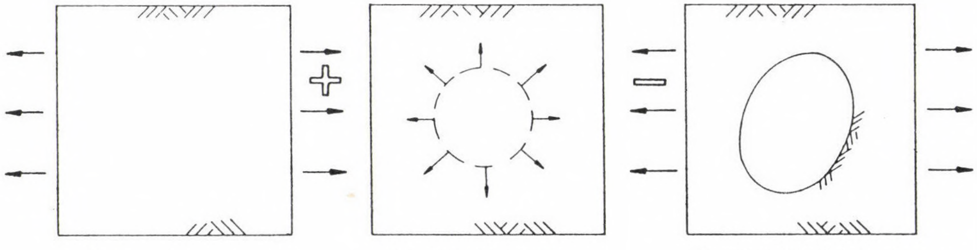


Fig. 16.

referring to the rule of the transform of the derivative (Korn and Korn 1961), one has

$$\sigma + p_1 s\sigma - \frac{k}{s} = q_0\varepsilon + q_1 s\varepsilon \tag{42a}$$

and since

$$\sigma = \frac{P}{s} \tag{42b}$$

one gets

$$(1 + p_1 s)\,\frac{P}{s} - \frac{k}{s} = (q_0 + q_1 s)\varepsilon. \tag{43}$$

That is

$$\varepsilon = \frac{P}{s}\,\frac{1 + p_1 s}{q_0 + q_1 s} - \frac{k}{s(q_0 + q_1 s)} \tag{44a}$$

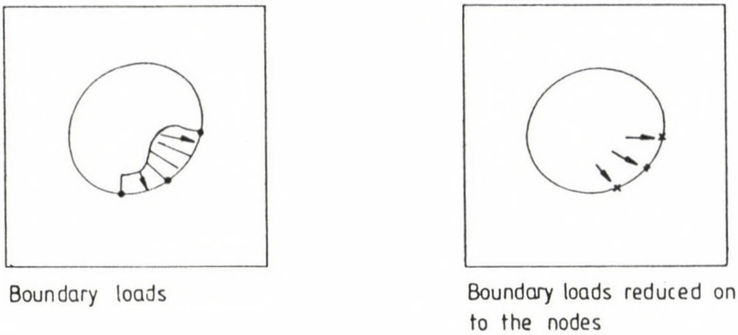


Fig. 17.

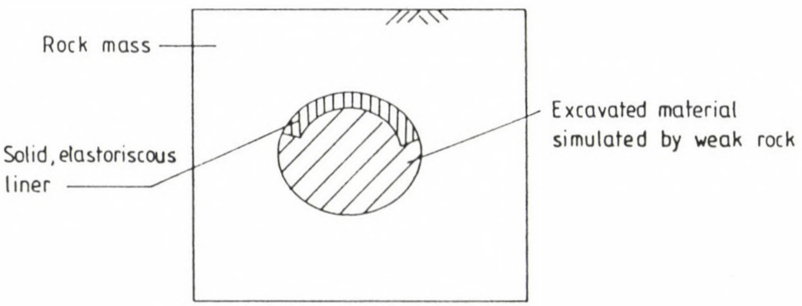


Fig. 18.

or in a rearranged form

$$\varepsilon = \frac{P-k}{s}\frac{1}{q_0+q_1 s} + Pp_1\frac{1}{q_0+q_1 s}. \tag{44b}$$

In order to determine $\varepsilon(t)$ as inverse Laplace-trasform of $\varepsilon(s)$ we have to mind that by the definition of the Laplace transforms and by basic rules of integral calculus (Flügge 1975, Korn and Korn 1961)

$$\mathcal{L}\left\{e^{-\frac{q_0}{q_1}t}\right\} = \frac{1}{s+\frac{q_0}{q_1}} \tag{45a}$$

also

$$\mathcal{L}\left\{\frac{q_1}{q_0}\left[1-e^{-\frac{q_0}{q_1}t}\right]\right\} = \frac{1}{s\left[s+\frac{q_0}{q_1}\right]}. \tag{45b}$$

(This latter formula makes use of the previous result, too.)

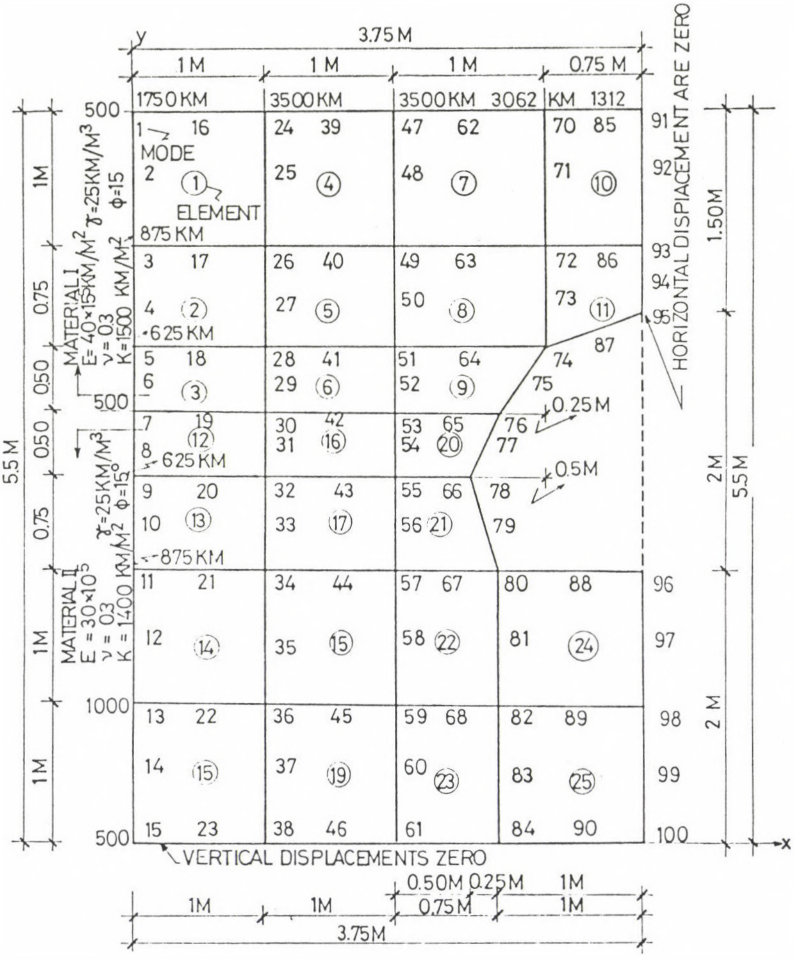


Fig. 19.

Now Eq. (44b) is rearranged once more again in order to use results of Eq. (45)

$$\varepsilon = \frac{P-k}{q_1}\,\frac{1}{s\left[s+\frac{q_0}{q_1}\right]} + \frac{Pp_1}{q_1}\,\frac{1}{s+\frac{q_0}{q_1}}. \tag{46}$$

Therefore, reading Eqs (45a) and (45b) in a reverse order, it can be written

$$\varepsilon(t) = \frac{Pp_1}{q_1}\,\frac{q_1}{q_0}\left[1-e^{-\frac{q_0}{q_1}t}\right] + \frac{P-k}{q_1}\,e^{-\frac{q_0}{q_1}t} \tag{47}$$

or

$$\varepsilon(t) = \frac{P-k}{q_0} + \left[\frac{P-k}{q_1} - \frac{Pp_1}{q_0}\right]e^{-\frac{q_0}{q_1}t}. \tag{48}$$

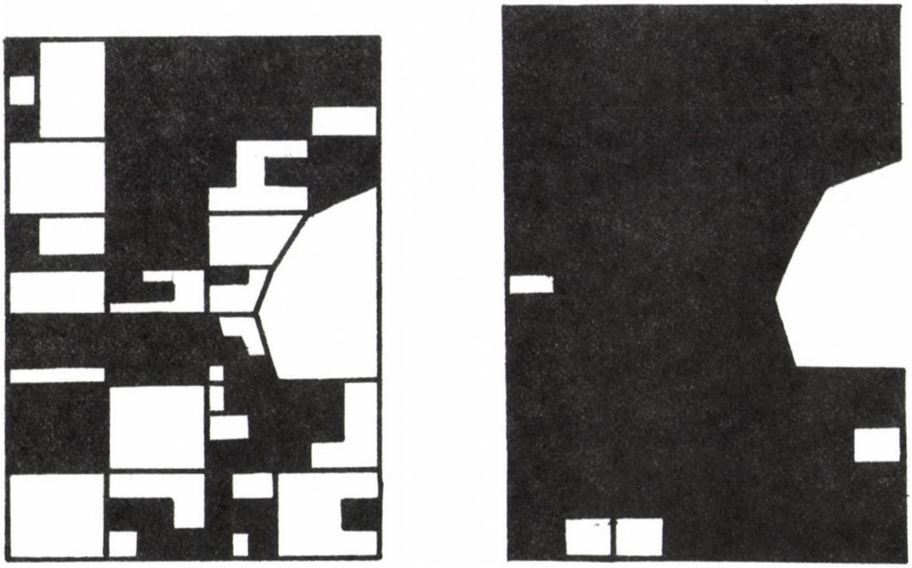


Fig. 20.

A short form, with the notations

$$\frac{P-k}{q_0}=a \quad \frac{Pp_1}{q_1}-\frac{P-k}{q_0}=b \quad \frac{q_0}{q_1}=\lambda \tag{49}$$

reads as

$$\varepsilon(t)=a-be^{-\lambda t}. \tag{50}$$

This is similar to the creep compliance function of a three parameter model.

Analyzing the result, the strain as a function of time is presented while the stress P is a parameter (Fig. 9), or the strain as a function of the stress P, while time is fixed (Fig. 10). Considering the particular sketch of Fig. 11 we define there the secondary modulus $E_{\mathbf{II}}$ as slope of the strain hardening line and also the (secondary) hardening parameter $H_{\mathbf{II}}$; for which

$$H_{\mathbf{II}}=\tan\beta_{\mathbf{II}}=\frac{P-\sigma_y}{\varepsilon_p(t_0)} \tag{51}$$

is true, that is $H_{\mathbf{II}}$ is considered as a reduced elasticity modulus. Recalling the notations (39), the formula of $E_{\mathbf{II}}$ reads after some calculation as

$$E_{\mathbf{II}}=\frac{E}{1+\dfrac{E}{H}\left[1-e^{-\frac{\mathbf{H}}{\mathbf{F}}t_0}\right]}. \tag{52}$$

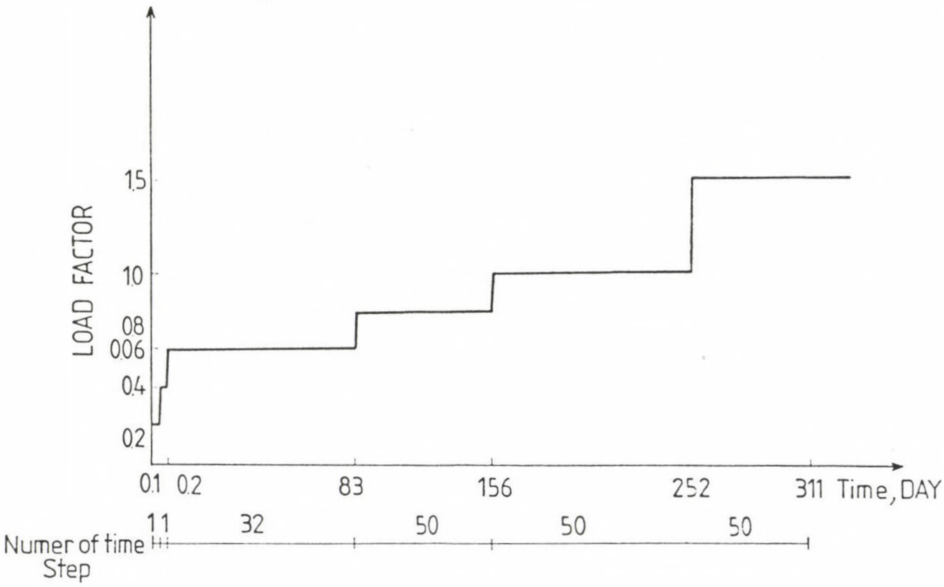


Fig. 21.

The diagram in Fig. 9 is quite similar to the creep compliance curve of the standard Poynting-Thomson viscoelastic material which is on the other hand fully linear at all stages and levels of the stress. Besides the diagram in Fig. 10 presents a strain hardening elastoplastic material with particular elasticity moduli in the completely elastic and also in the strain hardening stage. So the following two theorems can be stated:

Theorem 1. *The four parameter elasto-viscoplastic model built up by a spring, a dashpot and a slider behaves in the strain hardening stadium like a three parameter elasto-viscous model built up by two springs and a dashpot. This result can be denoted as the material analogy of the standard models.*

Theorem 2. *At a definite instant the four parameter elasto-viscoplastic model built up by a spring, a dashpot and a slider behaves like an elasto-plastic model the characteristic line of which is bilinear. The primary actual elasticity modulus is the usual Young's modul E while the secondary modulus is $E_{\mathbf{II}}$.*

$$E_{\mathbf{II}} = \frac{E}{1 + \frac{E}{H}\left[1 - e^{-\frac{\mathbf{H}}{\mathbf{F}}t_0}\right]} \tag{53}$$

This analogy can be denoted as the second material analogy, connecting the visco-elastoplastic and the elastoplastic models.

An explanation is shown in Fig. 11, where a separate elastic part and another separate, plastic strain hardening part of the actual strain are seen. As a conse-

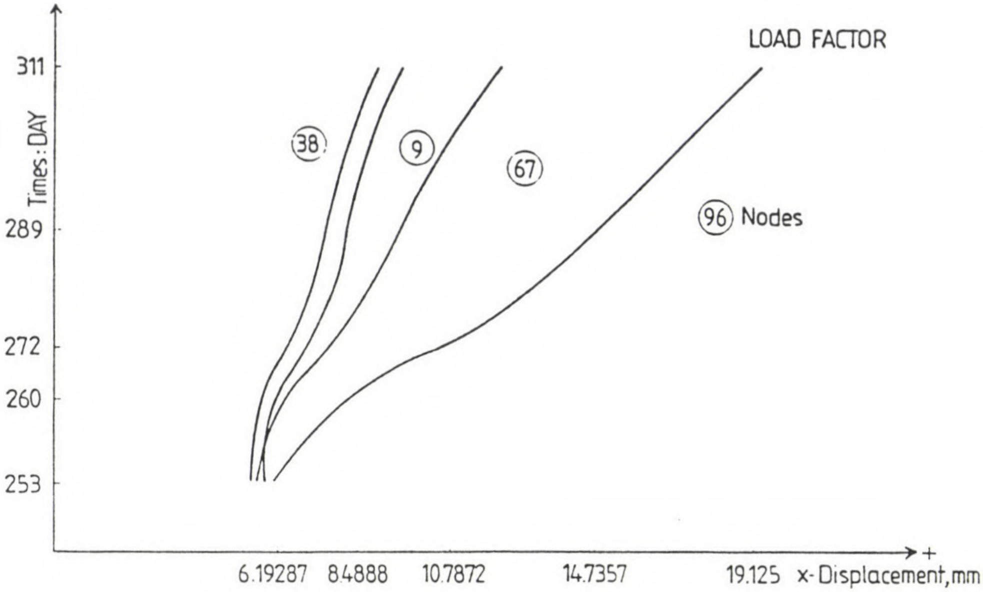


Fig. 22.

quence we state the next theorem:

Theorem 3. *At a definite instant the elasto-viscoplastic strain of the particle can be divided into an elastic and a plastic part, hence*

$$\varepsilon(t_0) = \varepsilon_e(t_0) + \varepsilon_p(t_0). \tag{54}$$

Now we have to mention a controversary of the linear and nonlinear characteristic curves, being explained by Fig. 12. This indicates that permanent stresses which are of proportionally increased intensity (Fig. 12a), induce to linearly behaving materials strains that can be generated from each other by shifting their compliance graphs parallel to the axis σ. On the other hand, if the material character is nonlinear, stresses altered in the same way are accompanied by firmly modified deformation curves. Meeting uniaxial, elastic state of stress, instead of the Hookean formula

$$\sigma = E\varepsilon \tag{55}$$

the Ramberg-Osgod relationship

$$\frac{\varepsilon}{\varepsilon_0} = \frac{\sigma}{\sigma_0} + \frac{3}{7}\left[\frac{\sigma}{\sigma_0}\right]^n \tag{56}$$

is proposed (Kaliszky 1990), complicating nevertheless the numerical calculations.

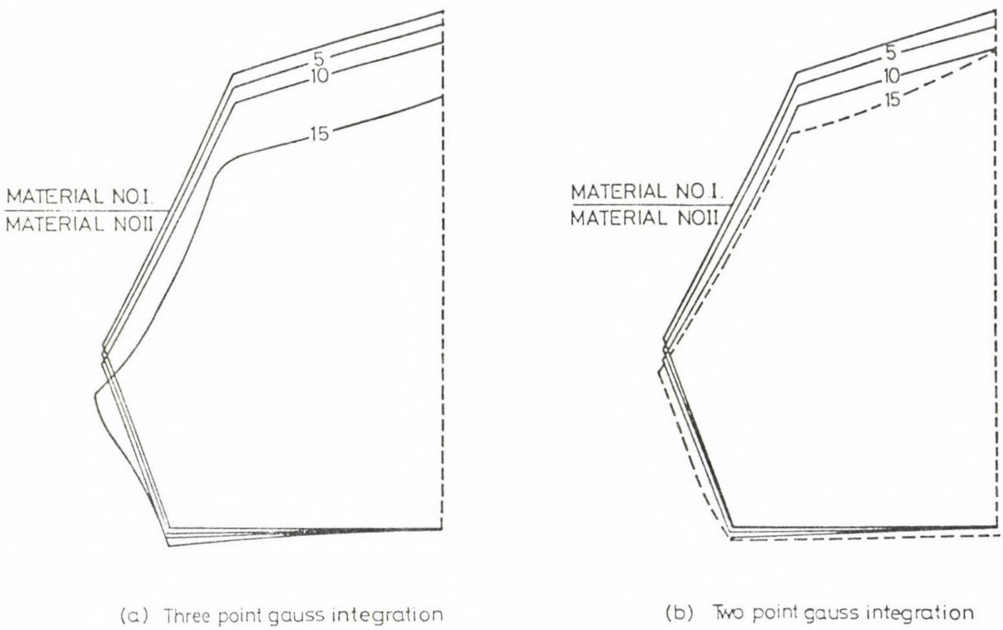


Fig. 23.

In case of viscoelastic material, the constitutive equation

$$\varepsilon(t) = \int_0^t \frac{d\sigma(\tau)}{d\tau} J_1(t,\tau)\, d\tau + \int_0^t\int_0^t \frac{d\sigma(\tau_1)}{d\tau_1}\frac{d\sigma(\tau_2)}{d\tau_2} J_2(t,\tau_1,\tau_2)\, d\tau_1, d\tau_2 \ldots \tag{57}$$

is written instead of Eq. (3), starting out in particular from the consideration that altered compliance functions are valid in each separate instant (Findley et al. 1976).

Application of the material law based on Eq. (57) meets in case of triaxial state of stress insolvable difficulties at the time being.

7. Modelling of the rock mass

The Department made a considerable success in the field of mechanical modelling of rock continua. Three kinds of models were dealt with in recent years:

i) Truss model

ii) Model of the mass consisting of completely rigid blocks coupled by elastic joints

iii) Continua replaced by a set of finite elements.

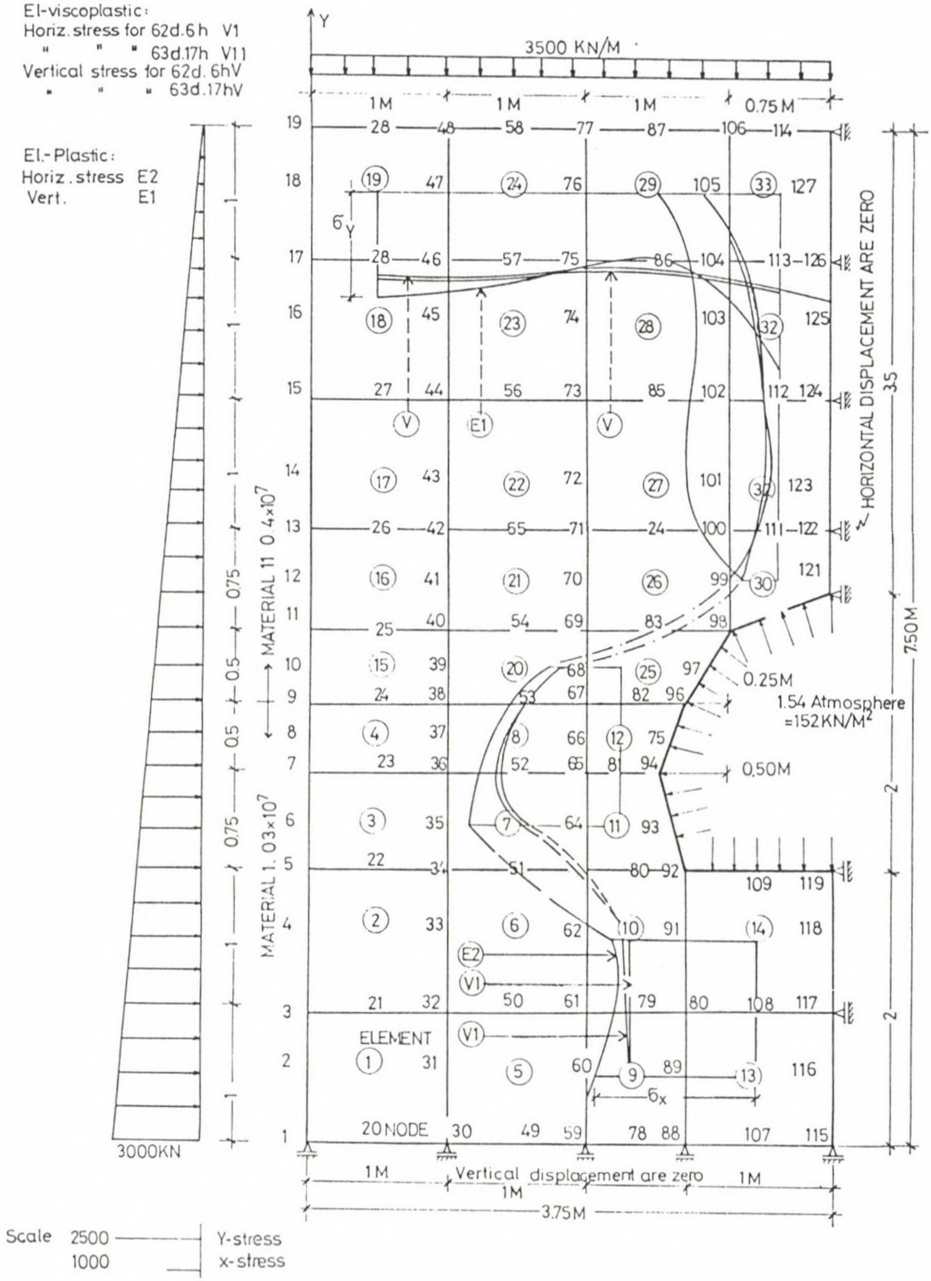

El-viscoplastic:
Horiz. stress for 62d.6h V1
" " " 63d.17h V11
Vertical stress for 62d. 6hV
" " " 63d.17hV
El.-Plastic:
Horiz. stress E2
Vert. E1
3500 KN/M
1M
1M
1M
0.75 M
HORIZONTAL DISPLACEMENT ARE ZERO
MATERIAL 11 0.4×10^7
MATERIAL 1. 03×10^7
0.25M
1.54 Atmosphere =152KN/M^2
0.50M
ELEMENT
20 NODE
3000KN
Vertical displacement are zero
3.75M
7.50M
Scale 2500 Y-stress
1000 x-stress

Fig. 24.

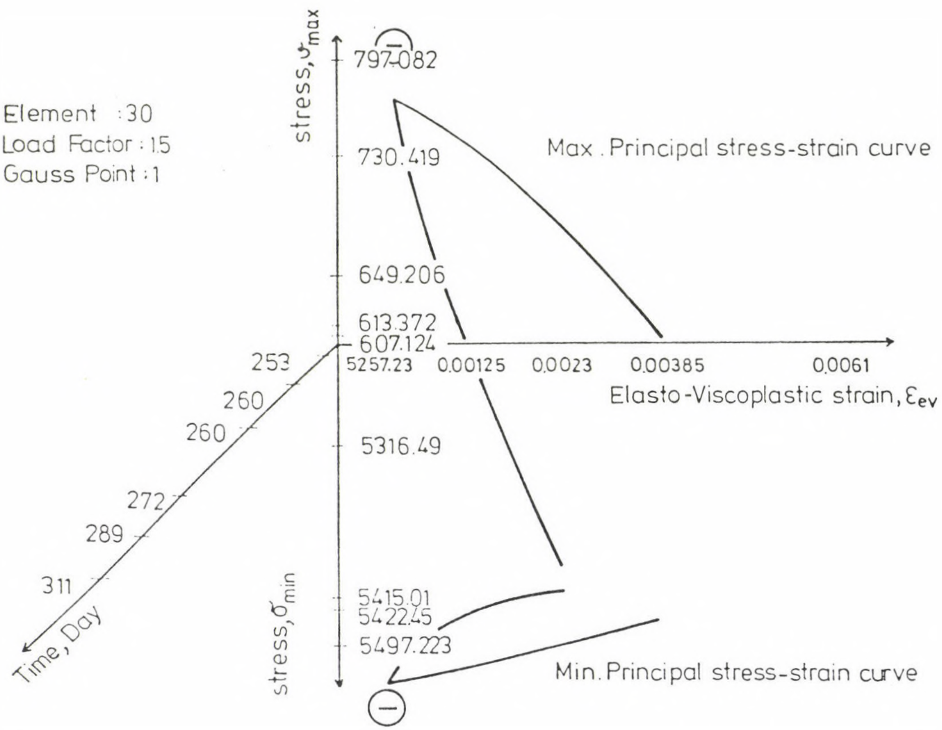


Fig. 25.

An optional arrangement of the truss model is presented in Fig. 13. It is not worth applying double bracing, since it does not correspond to the nature of the rock. The altering state is described by the matrix integral equation

$$\mathbf{A}(t,t)\,\mathbf{x}(t) + \int_0^t \mathbf{V}(t-\tau)\,\mathbf{x}(\tau)\,d\tau = \mathbf{b}(t), \tag{58}$$

where

$\mathbf{x}(t)$ — a vector containing the unknown nodal displacements and the bar forces,

$\mathbf{b}(t)$ — vector of prescribed nodal load components and thermal effects.

Integral equation (58) may be solved *e.g.* by iteration which is convergent meeeting the circumstances of engineering problems, but this trend needs lots of computer time. Theoretical and practical connections of the problem are treated by Roller (1980).

An advance of the model made of rigid blocks is that it can imitate in a quite realistic manner the gaps situated between layers, so the stratified continuum is simulated well, provided the preliminary geological exploration is ample enough.

The algorithm concerning the elastic case has been published by Roller (1970) a long time ago, with respect to shell structures. The analysis is recalled by Fig. 14.

The finite element state altering analysis of elastic, elasto-viscous, or elasto-viscoplastic rock continua consisting of triangles and isoparametric rectangles (optionally also of other finite elements), referred in Fig. 15 has been elaborated by the Department, too. A detail problem will be presented in the next paragraph. Here we mention just the constitutive law of the elastoplastic state. Instead of the bilinear law of the uniaxial case a reduced material stiffness matrix is to be applied in case of a triaxial state of stress (Owen and Hinton 1980) as follows

$$D_{ep} = D_e - \frac{\mathbf{D}_e^* \mathbf{g}\mathbf{g}^* \mathbf{D}_e}{\mathbf{g}^* \mathbf{D}_e \mathbf{g}} \tag{59}$$

$\mathbf{D}_e$ — matrix of the elastic material stiffness

$\mathbf{g}$ — gradient of the Prager-Drucker yield function, being derived by the help of the partial derivatives with respect to the stress components.

Having compiled the matrix D_{ep}, the analogon of Hooke's law can be formally established that relates the incremental strain to the incremental change of the stress (Owen and Hinton 1980) as

$$d\boldsymbol{\sigma} = \mathbf{D}_{ep}\, d\boldsymbol{\varepsilon}\,. \tag{60}$$

If the material is elasto-viscoplastic, then the viscoplastic strain increment reads as

$$d\boldsymbol{\varepsilon}_{vp} = \gamma < \phi > \mathbf{g}, \tag{61}$$

where

γ — reciprocal viscosity coefficient

$< \phi >$ — a potential function depending only on the stresses, defined exclusively in the viscoplastic state.

Another problem of establishing realistic models is the determination of the stress disturbance caused by the excavation of a hole in the mass, provided the primary state is prescribed and also the stress influence due to a pointwise internal load can be determined, that is the Green functions of the stresses are at disposal. These latter are obtained from the Theory of Elasticity. The fundamental idea is shown in Figs 16–17. In the first figure the superposition method of evaluation of the stresses is presented related to the opening of the hole, while the second shows the basic principle of the boundary element method, according to which the nodal loads are to be determined developing at the opening of the case by removing the rock material, via applying appropriate interpolation functions, similarly to the procedure valid at the computation of the load vector in the finite element method.

The flow chart of the algorithm applied at the finite element method calculation is presented after Owen and Hinton (1980).

8. Numerical example carried out by the finite element method

A program in concordance with the flow chart presented above has been worked out at the Department, partly as further development of the former investigation, partly in frame of Ph.D. thesis supervised by the author (Fard). Some input data and results of a numerical problem which has been solved by applying this code are presented in this paragraph.

The subject of the example is the investigation of the displacement and stress process around a hole opened in the rock mass in order to build a tunnel. The structure is situated in a rock environment containing two separate layers. The layout, the position of the strata, also the selection of the finite elements and the account numbers are shown in Fig. 19. (The load taken into account is presented in Fig. 24.) The problem has been worked out both by elasto-plastic and elasto-viscoplastic constitutive laws. Gauss-Legendre integral based on two respectively three points at each element were used. Thus the yielded elements and zones have been determined by selecting both number of Gauss points. It turned out that the selection of 3 Gauss points is more realistic, therefore its application is reasonable even if the computer consuming is greater than by use of 2 Gauss points.

Using elasto-viscoplastic material, Eq. (61) of the constitutive equation has been applied in case of yielding. Then the loading process is described by Fig. 21. Figure 22 indicates separately the displacements of some points of interest, indicating secondary creep. Figure 23 presents the planar convergence of the bottom and of the crown, respectively. This effect is quite significant, in accordance with the practical experience. Figure 24 shows the spatial and time distribution of different stress components, especially at the contour of the layer, respectively at the symmetry line. Finally Fig. 25 indicates the time distribution of the maximum and minimum principal stress on some selected elements.

References

Bojtár I, Bagi K 1989: *Acta Technica Acad. Sci. Hung.*, 102, 207–225.

Bojtár I, Kirchner I, Lovas A 1990: Finite Element Analysis of human femur. Third Polish-Hungarian Joint Seminar on Mechanics of Inelastic Solids and Structure, Cracow

Fard M A 1992: Altering state changing around underground opening. *Periodica Polytechnica* (in press)

Findley W N, Lai J S, Onaran K 1976: Creep and relaxation of nonlinear viscoelastic materials. North Holland, Amsterdam, New York, Oxford

Flügge W 1975: Viscoelasticity. 2nd Revised Edition. Springer, Berlin–Heidelberg–New York

Gajári Gy 1990: *Acta Geod. Geoph. Mont. Hung.*, 25, 153–173.

Kaliszky S 1989: Plasticity. Theory and Applications. Akadémiai Kiadó, Budapest

Kaliszky S 1990: Mechanik II. - Festigkeitslehre (in Hungarian). Tankönyvkiadó, Budapest

Kirchner I 1989: *Periodica Polytechnica, Civil Eng.*, 33, Nos 3–4, 201–210.

Korn G A, Korn T M 1961: Mathematical Handbook for Scientists and Engineers. McGraw Hill, New York–Sydney

Kurutz-Kovács M 1981: *Periodica Polytechnica, Civil. Eng.*, 25, 169–189.

Kurutz-Kovács M: On non smooth stability-analysis. *Periodica Polytechnica, Civil. Eng.*, (in press)

Owen M R, Hinton C W 1980: Finite Elements in Plasticity. Pineridge Press Ltd, Swansea, U.K.

Popper Gy, Csizmás F 1987: *Acta Techn. Acad. Sci. Hung.*, 100, 259–267.

Roller B 1970: *Számítástechnikai Közlemények*, No. 14, 25–48.

Roller B 1976: *X. Oszt. Közl.*, 9, 217–244.

Roller B 1978: *Acta Geod. Geoph. Mont. Hung.*, 13, 395–424.

Roller B 1979: *Periodica Polytechnica, Civil. Eng.*, 23, No. 2, 101–115.

Roller B 1980: Altering state analysis of viscoelastic bar structures (in Hungarian). Akadémiai doktori értekezés, Budapest

Roller B 1984: *Építés- Építészettudomány*, 16, 73–79.

Roller B 1989: *Acta Techn. Acad. Sci. Hung.*, 102, 339–354.

Roller B 1990: Systems theory, linear viscoelasticity and plasticity. International Colloquium Stability of Steel Structures, Budapest, Final Report, 321–320.

Roller B, Kesserű Zs 1976: Anwendung der Methoden der endlichen Elemente für die Untersuchung der Unterbauten Hangendschichten in Falle veränderlicher materieller Eigenschaften. III. Int. Symp. für Markscheidewesen, Leoben, Band 4, 189–199.

Roller B, Szentiványi B 1976: *Periodica Polytechnica, Civil. Eng.*, 20, No. 3–4, 111–126.

Simoni L, Schrefler B A 1989: *Meccanica*, 24, 98–106.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 371-379 (1992)

SOME PROBLEMS WITH CALCULATING VARIOGRAMS

R HARGITAI[1]

[Manuscript received May 20, 1991]

The summarizing analysis and the estimation of a variable is based on the computation of experimental variograms and on fitting theoretical models to them.

This paper lists some problems appearing in this stage of the geostatistical works; and shows the reason(s) of them in some cases.

Keywords: data population; distance class; geostatistics; sampling; variogram cloud; weighting factor

Introduction

The geostatistical literature gives a slightly misleading impression about the calculation of variograms, saying that to calculate them and to fit a variogram model on these experimental points, is easy. Everyone who has ever tried to do this will know that this is not true. The experimental variograms of the raw data are often highly fluctuating and generally it takes a long time and much effort to find out why. It is very important to determine what are the causes of this phenomenon.

This report gives a selection of some examples of the more common problems in structural analysis and suggests some ways to remedy them.

The theoretical background

It is an important fact in geostatistics, as pointed out by Matheron (1965) that the experimental variogram is highly variable for large values of h.

Alfaro (1979) and Brookeer (1982) have demonstrated that in cases when the underlying variogram is known, there can be marked differences between the experimental variogram and the theoretical model using simulated data. This fact can also be shown by constructing a random walk using random number tables, even though the theoretical variogram is $|h|$. From the experimental data it is simple enough to choose sequences of values that have a flat (*i.e.* pure nugget effect) variogram, or, alternatively have one which rises faster than h^2 (i.e. appears nonstationary). It is surprising to find such a fundamental problem after so much study into the theoretical structure and practical calculation of variograms over a long period of time.

[1]Ecole Nationale Superieure des Mines de Paris, Centre de Geostatistique, 35, Rue St Honoré, 77305 Fontainebleau, France 25
Permanent address: University of Miskolc, Faculty of Geodesy and Mine Surveying, H-3515 Miskolc, Egyetemváros, Hungary

The variogram

Let $Z_{(x)}$ be an intrinsic random function (e.g. the ash content of the analysed samples of coal deposit).

The variogram of $Z_{(x)}$ can be defined by the equation of:

$$\gamma_{(h)} = 1/2Var[Z_{(x+h)} - Z_{(x)}]$$

where x and $x+h$ refer to two points in an n-dimensional space where n = 1, 2, or 3.

Assuming that the mean of $[Z_{(x+h)} - Z_{(x)}]$ is zero, $\gamma_{(h)}$ is just the mean square value of the difference:

$$\gamma_{(h)} = 1/2E[Z_{(x+h)} - Z_{(x)}]^2 .$$

The graph of $\gamma_{(h)}$ plotted against h presents the following features:

— it starts at 0 (for $h = 0$, $Z_{(x+h)} = Z_{(x)}$)

— it generally increases with h

— it rises up to a certain level called the sill and then flatterns out.

Alternatively it could just continue rising.

The experimental variogram can be calculated using the following formula:

$$\gamma^*_{(h)} = 1/2N_{(h)} \sum_{i=1}^{N_{(h)}} [Z_{(x_i+h)} - Z_{(x_i)}]^2$$

where:
$Z_{(x_i)}$ – are the data values
x_i – are the locations of values
$N_{(h)}$ – is the number of pairs of points $(x, x+h)$ separated by a distance h (with directional considerations taken into account).

Some experimental cases

1. Poor choice of distance classes

Figure 1 represents a very erratic behaviour caused by a poor choice of distance classes. The variogram of the zinc content of a polymetallic deposit has been calculated with a class of 20 m.

Although it was not expected that the 60 m grid of drillholes could cause such a problem, this was the most probable explanation (Fig. 2).

To check this, a scatter diagram of the square differences, $[Z_{(x+h)} - Z_{(x)}]^2$ against distance h, (Fig. 3) was calculated.

This scatter diagram, which has been called the variogram cloud by Chauvet (1982), has proved very handy for diagnosing problems.

The variogram cloud indicated a problem which could be solved by recalculating the variogram using a new distance class, which is not a submultiple of 60 m. The resulting variogram will be a smoothed version of the first one.

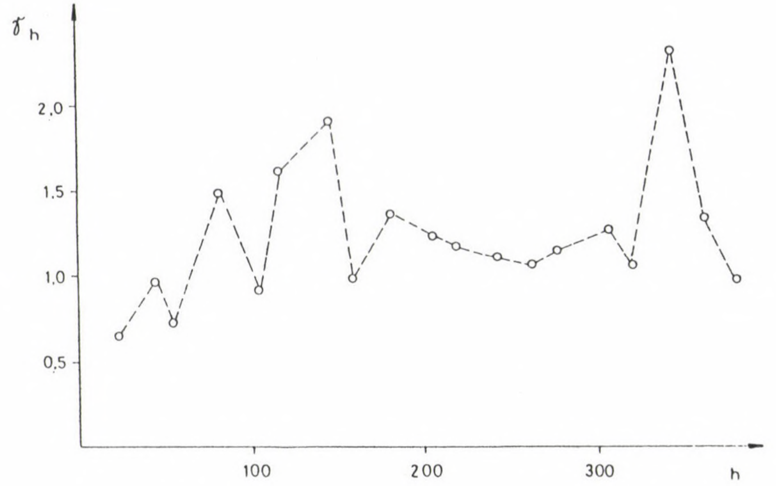


Fig. 1. Typical erratic behaviour (poor choice of distance)

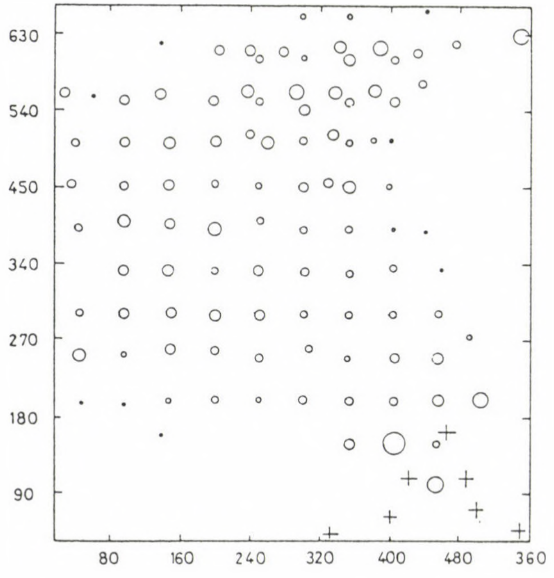


Fig. 2. The drillholes on the surface

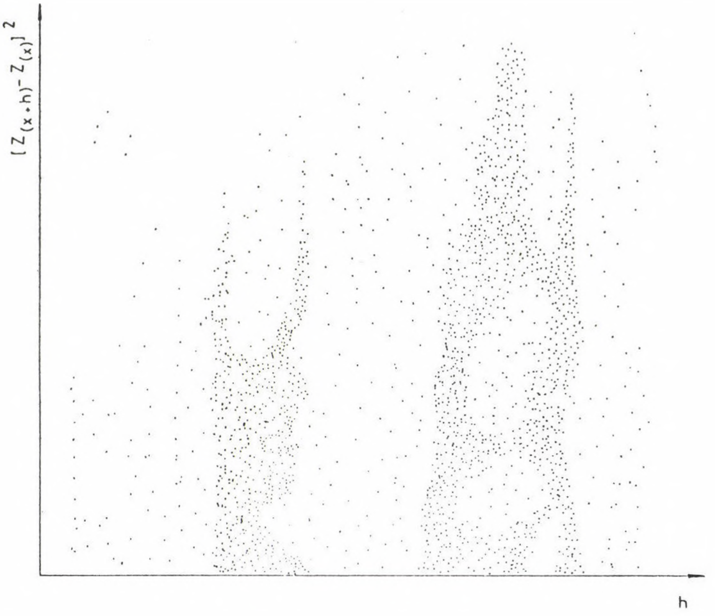


Fig. 3. The "variogram cloud"

2. Width of d.angle

To calculate variograms for the data in two or three dimensions we fix the direction (α) and the tolerance, called d.angle ($d\alpha$) (Fig. 4).

The d.angle has to be choosen cautiously. Using too small a ($d\alpha$) may result in an unusable variogram as in Fig. 6.

Mixed populations

A mixed population means that the data came from two (or more) statistically different populations.

1. Geographically distinct populations

The variogram in this case represents a mixture of the features of the different areas. The geographically distinct populations are observed when examining histogram, as bimodality (Fig. 7). If data are divided into as many parts as populations it may happen that data are not enough for the separate analyses.

2. Mixed errors

The same problem can occur when the data arise from different sampling or analysis methods, or the research work has been done in different campaigns, as

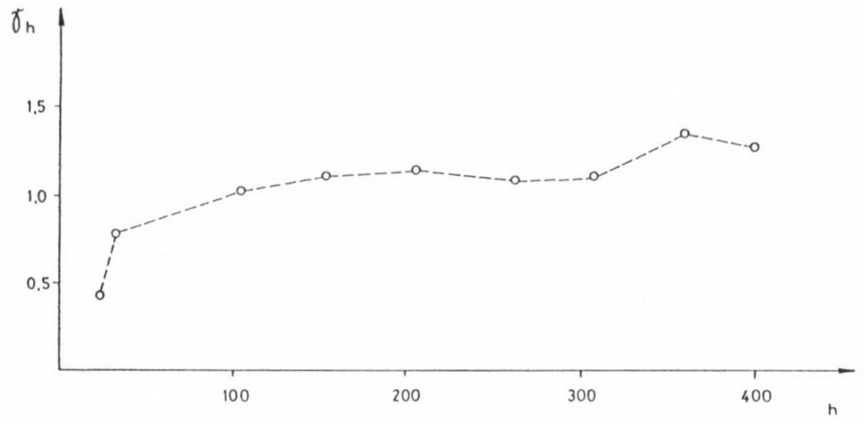


Fig. 4. The recalculated variogram

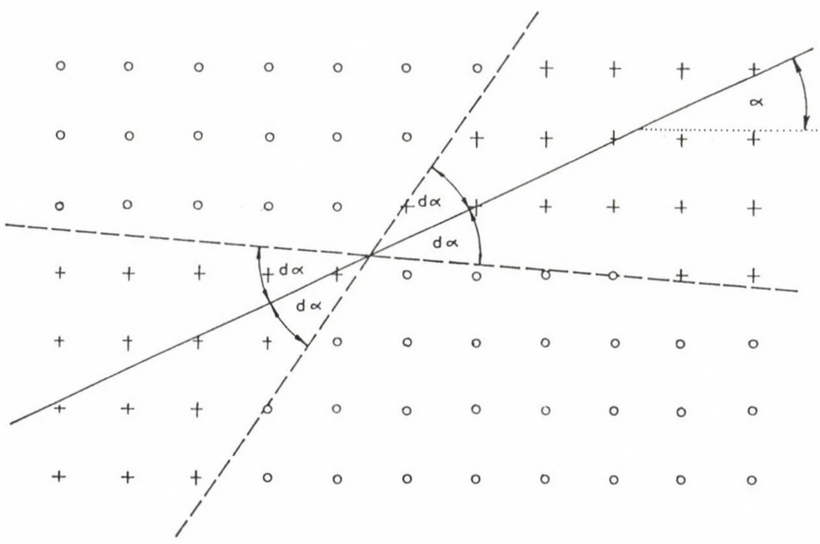


Fig. 5. The behaviour of the variogram depending on the width of dx

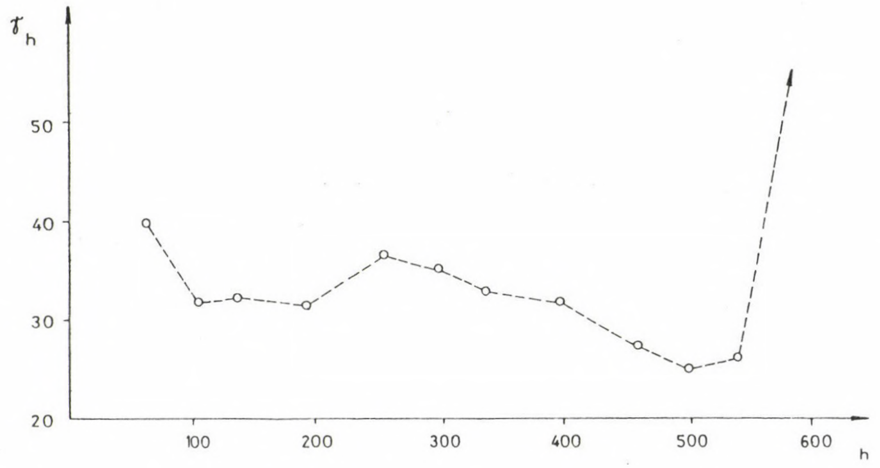


Fig. 6. Choosing too small $d\alpha$, the variogram gets in usuable

mentioned by Armstrong (1984).

3. Problems of extreme values

This is almost the same problem as that in the case of distinct populations, but in this case it is caused by some extremely high or low data.

Figure 8 shows the variogram of lead, calculated from 1214 samples, and Fig. 9 the density function of these data.

Masking out the extremely high values — ($Z_{(x_i)} > 2.95$) — and calculating the variogram in the same direction and using the same parameters as before, we should have a more utilisable form (Fig. 11).

4. Masking out the extremely high values is a method too drastic to use it generally. Two alternatives can be offered:

4.1 Transformations

4.11 The experimental variogram is transformed to a lognormal one which is easier to fit a theoretical model to. After fitting, both the experimental and theoretical variograms can be retransformed, using the following equation:

$$\gamma_{(h)} = e^{-2\eta+2\sigma^2} - e^{\sigma^2[1-\rho_{(h)}]}$$

where m, σ^2, and $\rho_{(h)}$ are the mean, the variance and the covariance function of the transformed data.

4.12 The other way to solve the problem is kriging using the lognormal transformed data and variograms — as first in 1974 — made by Marechal who called it lognormal kriging. After this, the estimated data are transformed back.

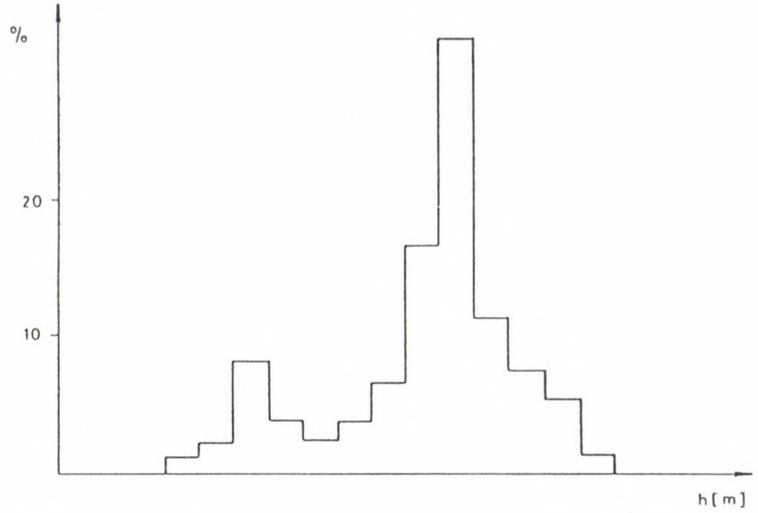


Fig. 7. Geographically distinct populations

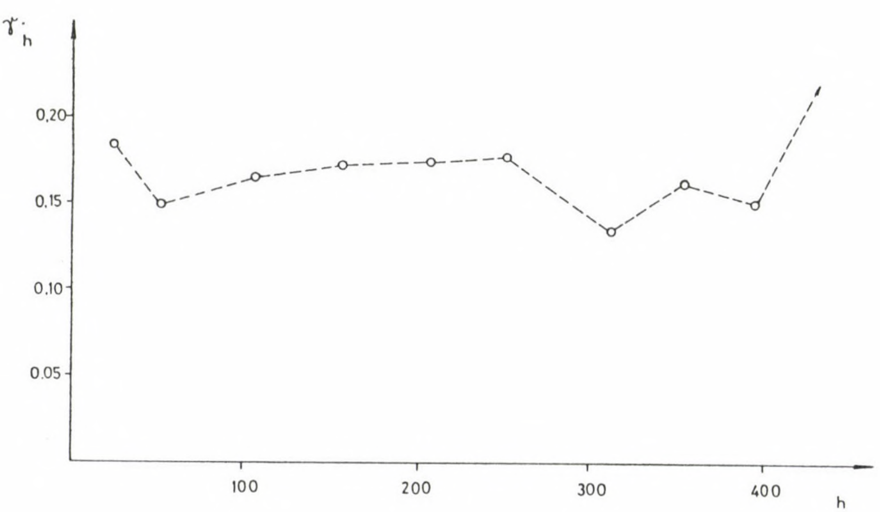


Fig. 8. Problems of extreme values

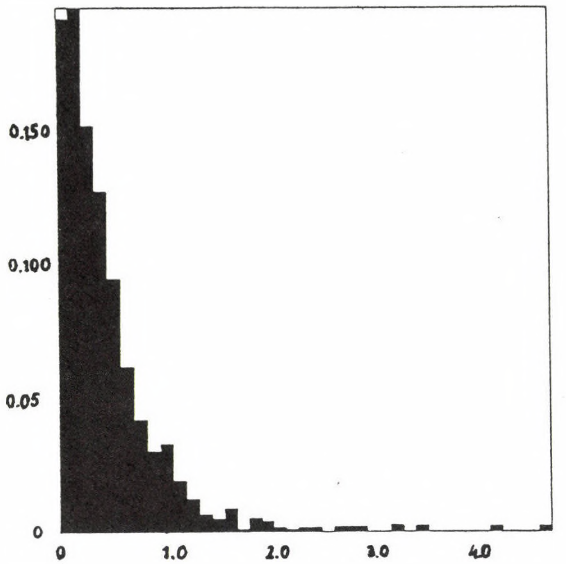


Fig. 9. The density function in case of extreme values

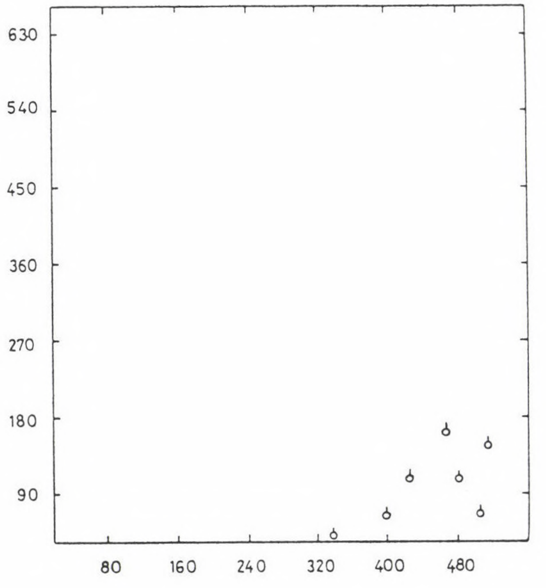


Fig. 10. The spacing of the extreme values

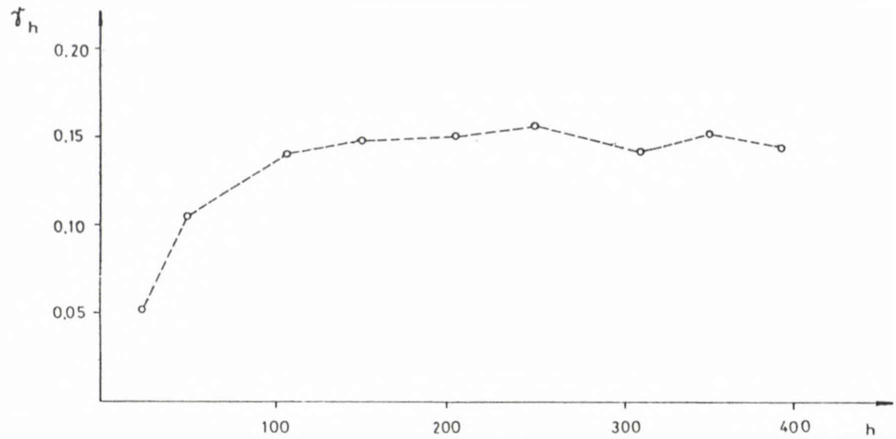


Fig. 11. The variogram, after masking out or weighting down the extreme values

4.2 The method of down-weighting

Let us multiply each $Z_{(x_i)}$ by a weighting factor $\lambda_{(x_i)}$:

$$\lambda_{(x_i)} \min \gg \lambda^e_{(x_i)} \min$$

where $\lambda^e_{(x_i)}$ are the weights of the extreme values. A special case of this method is masking out the extreme values using zero weights

$$\lambda^e_{(x_i)} = 0 .$$

This short list cannot be comprehensive but perhaps it can help to avoid the obstacles which faced by specialists working in geostatistics, and to spare time and energy for them.

References

Alfaro M 1979: Etude de la robustesse des simulations aléatoires. Thése de Docteur-Ingénieur, Ecole Nationale Superiieure des Mines de Pariis

Brookeer P 1982: AIME Annual Meeting, Dallas, EIME Preprint No. 82–11.

Chauvet P 1982: In: Proceedings of the 17th APCIM Symposium, Colorado School of Mines, Golden, Colorado, 757–764.

Marechal A 1974: Krigeage normal et lognormal. Ecole des Mines de Paris, CGMM, Fontainebleau, unpublished note No. N–376.

Matheron G 1965: Les variables regionalisées et leur estimation. Masson et cie, Paris

Additional information

Armstrong M 1982: Improving the estimation and modelling of variogram. Presented at the 6th Australian Statistical conference. Melbourne University

Armstrong M 1984: *J. Math. Geol.*, 16, 305–313.

Cressie N, Hawkins D 1980: *J. Math. Geol.*, 12, 115–126.

Rendu J M 1979: *J. Math. Geol.*, 11, 407–422.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2-4), pp. 381-388 (1992)

SOME DATA TO THE CRITICAL ANALYSIS OF MINERAL RESOURCES VALUATION

G FALLER[1]

[Manuscript received May 30, 1991]

In East-Middle-Europe a new epoch of economic history has been started. In this epoch a significant role will be played by mineral resources valuation. Critical analysis of its past may contribute to expedient forming of the relevant methods. This paper is an experiment to outline a possible mode of the critical analysis.

Keywords: COMECON; history of mining; mineral resistivities; valuation

When outlining the frames of this topic it seems to be expedient to start from the fact that the success of mining *i.e.* the value of the mineral raw material deposit is fundamentally influenced by natural potentialities. Natural potentialities having an impact on expenditures (depth, thickness, tectonics, natural hazards, *etc.*), are called "bonity of the deposit", natural potentialities relating to the value of the mining product (ore content, thermal value, ash, coking capacity, *etc.*) are called "quality of the deposit". The result obtained from the difference of product value and expenditures (and the rentability to be quantified as the quotient of the above two respectively) has two constituents as depending variables, further the natural potentialities as independent variables which are in a quantitative relationship between each other. These stochastic relationships are also reflecting the level of technical perfection. But this reflection is limited by the fact that no free "selection" is provided from the generally wide-ranged stock of means of the engineering because natural potentialities are restricting the sphere of applicable solutions in actual cases (in deposits).

1. Main directions of long-term development

It follows from the outlined relationships and circumstances that in the ore mining of the early Medieval-Medieval age (*e.g.* in the ore mining of the Carpathian Basin) the factors of bonity (rock-strength, depth, vainy facies, *etc.*) were relatively homogeneous. But the quality factors (taking into account that they are basically fond-resources) were very variable. Available information was enough to determine the economy of production and the value of the deposit (and also their changes) which relates also to the metal content of the ore (and also to its changes). According to these factors it was spoken about poor or rich ("beneficient") vains (about impoverishment of rich vains and *vice versa*). In the 17-18th centuries the role of

[1]Dísz tér 8, H-1014 Budapest, Hungary

bonity has appeared in the value judgement due to the increase of expenditures caused first of all by increasing mining depth as beginning of a historical development which ensured bonity in the oil mining an almost ruling position in the economic judgement; this happened in the first half of this century. In this case one has flow-resources where quality factors are less variable than those of bonity.

In interaction with these processes mining science developed from mining craftmanship. One of the first representants of this science was Christopher Delius Traugott (1728–1779) who studied mathematics and natural sciences in Wittenberg and later mining and metallurgy in Selmecbánya. He became the first professor at the third chair of the Mining Academy there. After his installation in 1770 he wrote a book "Anleitung zu der Bergbaukunst ..." in which the practical and theoretical mining knowledge in the second half of the 18th century was summarized. This work was for many decades an authoritative text-book for the European higher education. It included geology, mining, ore dressing and mining economy.

From the present point of view the most important part of the book is the appendix: "Principles of Sciences of the Chamber of Mines" where Delius Traugott wrote as follows: *"The benefits of the nation from the mining are obviously fourfold: notably*

- *first of all the profit gained directly or indirectly by the treasury;*
- *secondly the benefit improving general means of subsistence;*
- *thirdly, that in connection with the mining through minting the produced gold and silver the national capital power increases and general circulation of money expands;*
- *fourthly, that through export of non-precious metals and other mining products a lot of money is rolling into the own state."*

Regarding the question: *"Whether it is reasonable to operate the mines with deficit?"* — Delius's opinion: *"Should the question refer only to gold and silver mining, the answer must be in any case affirmative, due to increasing capital power and money circulation of the country. Regarding non-precious metals such a statement is not justified, because they are to be considered as other industrial products. Notwithstanding this can be opposed that in this way at least the leakage of money can be hindered but it can be stated with full justification that the attention of the nation should rather be directed to the production of more profitable industrial products than to those with which the competition of other nations cannot be avoided."* It is easy to recognize that our present mineral resources management has to be based fully on this principle.

Since the time of Delius mineral resources valuation has became a more and more significant part both of mining and mining economy. This activity has obtained a particular name (even at present used in German). This name is expressively characterized in the "Introduction" of the "Science of Mining Art" by G Réz: *"... finally we shall particularly deal with appraisement of mines, mining properties*

(Schätzung der Gruben) which is a chapter of mining knowledge absolutely necessary for all mining engineers." All over the world and in Hungary, too, the system and practice of categorizing and recording mineral resources from various respects has developed in this sense.

Mine appraisement has been further expanded by József Finkey (1889–1941). In more detailed form this topic was dealt with in the "Mine Appraisement" by Péter Esztó (1885–1965). The basic principle is that the "appraised value" of a mine (*per analogiam*: mineral resources of a free area, undetected occurrence) is the actual capital value *i.e.* the sum of present values calculated as the discounted result obtained from the differences of annually attainable return of sales and relevant expenditures during the period between the date of appraisal and completion of exploitation. This principle has been applied when mine appraisal was not only necessary in the case of trading with mines and deposits between national capital owners, but also *e.g.* in the case of MAORT which operated on concession basis where the share of property between the American and Hungarian parties was 96 to 4 percent. This principle was used in the regulations of the Government in conditions of the market economy with all exact legal (mining law) definitions and with the controlling activity of a single governmental centre (in the ministry). The concession contract concluded with the company MAORT was signed on June 8, 1933. It prescribed that MAORT was obliged *"to carry out and maintain the production in a professional way and provided that this activity will ensure reasonable earnings for the New Company, to increase the production up to a level when the national consumption of Hungary will be covered."* From the amount of crude oil produced over the Hungarian demand, 85 percent were due to the American party and 15 percent to the Hungarian State as inland revenue. (The definition "inland revenue" shows that no significant changes in bonity or quality have been taken into account.) It is easy to recognize that the method of "mine appraisal" used in the early times nearly corresponds to the present-value calculation of economy-considering methods generally applied in present market economies and due to an undisturbed development also in mining. Present-value calculation has been developed (as recently summarized by Rudawsky (1986)) by applying more and more abundantly geostatistics to estimate mining risks as cautiously as possible.

2. The "detour" of four decades

During the last four decades mineral resources had been recorded in the East-Middle-European region under control of the state in accordance with the very detailed, uniform regulations developed in the Soviet-Union regarding categorization based on the degree of exploration and development of the occurrences. These same regulations had been applied in all COMECON-countries with minor alterations in each country based on scientific considerations. This system was used in Hungary, too but with a much higher scientific content as average. The third aspect of categorization, namely economy of exploitation, value (expressed in money) of mineral resources, did not appear at the very beginning. In this form the system was in harmony with the plan-directive system of the national economies and complied with the voluntarist economic development programmes having extremely high material and energy demands. These programmes had striven at a national autarky despite few mineral resources. The more and more successful exploration of the mineral resources in the COMECON countries and a certain division of labour made it in some countries (including in Hungary) possible and the depletion of sources of extensive development necessary (especially in our case) to carry out an economy-based selection. In interaction with this possibility and demand different methods based more or less on economic aspects have been developed for the economic valuation of mineral resources. This took place first in Hungary. These methods attained different scientific practical levels pending on the kind and success of reforms. In this sense there was no uniform "socialist" method and practice: the Hungarian one has developed (not without serious difficulties) as a simulation of market economy conditions. This method — even if it could not be separated fully from the distorted scale of values of the economy — could (in some cases would) serve on one hand the structural improvement of the economy but on the other hand it would be able functioning in an unchanged form in realistic market conditions.

The economic valuation of mineral resources (similarly to those of most natural resources) cannot be separated from its optimal application programme. The programme must be done for an adequate interval and possible (domestic an import) sources should be selected which satisfy the needs of the economy for the aim with the smallest possible expenditures. Application-expenditures of sources are to be taken into account and forecasted resp., by special calculations due to the fact that prices of inputs (in case of imports, the exchange rate of the currency) have been consciously deflected from their real value. The calculations were to be done as far as possible without distortion. Therefore it became necessary to deviate from the stately regulated calculation system, as (in the interest of our objective) only *expenditures in the future* with their compound interest have been taken into account, but not the costs of *investments carried out* previously.

The real value of the mine product is much better reflected in the cost limit computed by the expenditure-demand of the worst and most expensive source-element selected for the optimal programme than in distorted prices. This was the starting point of the considerations. Compared with this marginal source-element all selected-elements have the benefit of a differential allowance during execution of the programme and this is that differential allowance which is suitable in a given situation to define the *in situ* monetary value of the actual source-element (deposit, mine). The quotient of the two terms of the difference characterizes the degree of workability. Here the application-expenditures of the resource-elements are emphasized as the system and mathematical stock of the theory of location of mines developed as attested by two books of János Zambó: in 1960 the first edition of the "Analytics of Mining Locations" has been published and in 1966 the first edition of the "Theory of Location in Mining". Their fundamental importance is shown by the fact that this theory founded the turn of the attitude to economic history, as

— the revival of the notion rentability, of the efforts to achieve cost-minimum in mining has been started on the field of mine location,

— this method had fundamental influence on the first structure-optimizing programmes of the sixties first of all thereby that instead of comparison of unique alternatives a general solution, the analysis of cost-functions have been used; this influence also appeared in the first economic qualification of mineral resources of Hungary.

Later cost-forecasts used as fundamental criterium that costs should be settled in relation to the main optimal parameters of the mines and deposites. When ranking the single resource-elements and summing up the sequence of expedient mine-unit capacities resp., it had to be started with the capacity belonging to the cost-minimum of the actual mine, which has the smallest cost-minimum among all mines. This has been followed

— either by one of the production increments of that first mine

— or by the own optimal capacity of the mine having the second smallest cost-minimum,

(— or by one of the import-possibilities,)

according to which one of these represented smaller increment-prime cost, and so on. This means that from a more and more expanding circle the third, fourth *etc.* source-element is to be selected.

This sequence of ideas was to demonstrate that a correct mineral resource-valuation must not avoid economic mathematics of location-optimizing investigations especially those cost-functions serving capacity-optimizing. In a simulation of market conditions (by constraints) the programme ensures the minimum of the sum of increment-costs adequately reckonned (as measure of expenditures) of the mineral raw-material sources. The marginal (usually import) source-increment costs derived from this programme are considered (as cost limits) as the measure of value of the product. The *in situ* value defined by this method is in close relationship with the net present value and with the "estimated value" of the mine and the workability index is essentially a rentability index suitable for the economic ranking of deposits and mines, *i.e.* like the internal rate of return. This method of valuation could never be fully accepted in COMECON: it made real "storms" that instead of the retrospective price principle world market price forecasts have been used which was condemned as "antimarxist" by those who vulgarized the labour-value theory. From the whole process of development three conditions and subjects are considered to be conspicuous. These are:

a) The domestic method developed for economic valuation of mineral resources is uniform and does not depend on kind, state, morphogenetics *etc.* of the mineral resources established the methodological basis of economic valuation of different natural resources.

b) Practical experiences in the first decade of controlled and unmanipulated application of the method have provided a sound basis for the introduction of a property application fee introduction of which kept delayed up to now.

c) This system of economic valuation is uniformly regulated and constructed "from below upwards" like a hierarchy: integration of information relating to mining blocks value-judgements-balance summaries at deposit-, basin-, finally of national-level are established.

An enormous amount of data required handling (reckonning, flowing, deriving, archiving, *etc.*) personel and equipment have been continuously developed and are available.

Coming back to that basic principle of the method that the numerical value is the adequately reckonned amount of result during the life-time of the mine, it is obvious that for adequate reckonning is the synchronization of the expenditures and results necessary. Should the synchronization be made for the end of life-time, formally the *in situ* value of the mineral resources is obtained, with the peculiarity that annual results of current costs are be derived from some kind of price but from a cost-limit with the content given above. This is done according to a study by János Zambó: *"Analysis of mining investments economy may be different in each country. The situation is quite different in countries with up-to-date industry and a convertible currency and which have what to convert, too, compared with countries where the same possibilities are not and will not be available for longer time. Unfortunately our country is also like this. In this exigence human phantasy inevitably starts to operate and the concept of cost-limit is brought to world. This idea has reality only in countries with convertible currency. Where this is missing it is as long imaginary as convertibility is achieved."*

This is a highest level confirmation that the Hungarian location-theoretical school developed a method of valuation of mineral resources which fully complies with the conditions of that time (having a well-operating phantasy) and which with a return to market economy and with the convertibility to be established will get rid of its "simulation" character. (But will not get rid of the difficulties of forecasting which include that rate of exchange of currencies change in inverse proportion to the availability of convertible exchange commodities. This time their decreasing shortage will moderate the false allowance of domestic raw-material sources.) In this manner this valuation method will be again a real basis of mine estimation. Its result will provide an estimated value depending on the date of estimation within the life-time. Zambó's study shows this for the beginning of investment and for the start of operation but it is obvious that based on the Net Cash Flow presented in the study and when synchronizing future costs and results for adequate times the assessed value and the Net Present Value of the deposit and those of an operating mine can be defined. The Internal Rate of Return shown in the study as marginal interest factor can be calculated making above values equal to zero and are used in the function of the workability index applied in the solution of simulation for economic ranking of the deposits and mines. All these prove the "relationship" between the simulation method and the real market economy valuation as mentioned above.

This synthesis in the study makes it clear that by returning to the valuation method which had a historically organic development and which can be asserted now without simulation in the present new era of the economic history, the scientific content of the Hungarian school of location-theory will have high significance.

3. Initial efforts tending to a new synthesis

According to our critical survey we have to return after the detour of four decades to the long term and global main tendency of valuation. This is to be realized on the most expedient and efficient way by using recent results from market economies (first of all in countries of mineral resources potentialities being similar to our ones) and from the Hungarian simulation method. This is the way how mineral resources valuation can provide a sound basis for mineral resources management of the Government to fulfil functional requirements freezed in for four decades which are to be resuscitated now. An adequate application would support a more efficient utilization of the mineral resources with participation of foreign capital and hinder a bargain on a "below the value" price. This process was started by the quantification of reliability and risk resp. further to problems of capital apport and concession fees in case of joint ventures. This study work provided a sound basis to one- or two-stage modification of the valid Mining Law which is unavoidable due to the change of ownership. These modifications have to comply with the national regulating practice (which developed historically and can be qualified as the theoretically correct one). This practice proves that mineral resources management and its legal regulation shall be fundamentally uniform, *i.e.* not raw-material specific.

In connection with the codification of the new conditions we have to quote Lajos Litschauer who formulated about one hundred years ago. *"Countries where the whole mining is in state power do not exist any more; but we do not know if a country would exist where mining business would be charged solely on private ventures; the most frequent and also in Hungary existing case is that both the state and private ventures are mining."*

For such conditions he formulated unambiguously: *"It is task of conscience of the State ... and also its obligation to clear away all obstacles which are hindering reasonable and productive mining; it must be kept especially in mind that to obtain a licence of mining should require as low costs and efforts as possible. The most important requirement to ensure credit for the mining is legal relations for the easy alienation of mine-property ... and the safety of mine ownership."*

These ideas became actual in Hungary again in spite of the fact that in the meantime some mines are exclusively property of the state or of the private sector.

References

Delius Traugott C 1773: Anleitung zu der Bergbaukunst ... J T Trattner, Wien
Esztó P 1942: Mine Appraisement. Postgraduate Institute of Engineering, Budapest
Litschauer L 1890–1891: Bányamíveléstan. Joerges Ágost özv. és fia, Selmeczbánya

Réz G 1910: Science of Mining Art. Joerges Ágost özv. és fia, Selmeczbánya
Rudawsky 1986: Mineral Economics. Elsevier, Amsterdam, Oxford, New York, Tokyo
Zambó J 1960: Analytics of Mining Locations. Műszaki Könyvkiadó, Budapest
Zambó J 1966: Theory of Location in Mining. Műszaki Könyvkiadó, Budapest
Zambó J 1986: *J. of Mining and Metallurgy, Mining*, 122, 733–744.

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 389–397 (1992)

GEOSTATISTICAL MODEL OF ENVIRONMENTAL POLLUTION

A FÜST[1]

[Manuscript received November 18, 1991]

The study informs about a four-dimensional geostatistical model which is applicable to describe processes of environmental pollution and to prepare forecasts. Application of the method is illustrated by an example.

Keywords: environmental pollution; forecast; geostatistics

All kinds of environmental pollution are processes enacted in space and time, their intensity and degree is influenced by the media emitting and receiving pollution resp. (air, water, rock) further by numerous effects of mostly random character. For the investigation of such processes observation systems are established with measuring stations having the task of continuous or discrete measurements. From the results of the measurements maps showing the degree of pollution are elaborated which serve as basis for forecasts of expectable conditions.

For modelling of random processes in time mathematical methods and among them transformed geostatistical procedures are applied. In the followings the latter will be surveyed.

1. The system of observation and collection of samples

Within the area investigated n measuring stations with known coordinates exist. In these stations measurements have been carried out for any of the parameters in a discrete point of time. The result of the measurement should be denoted by $Z_{ijk}(x_{ij}, y_{ij}, z_{ij}, t_{ik}), (i = 1, 2, \dots n; j = 1, 2, \dots m_i; k = 1, 2, \dots T_i)$, where m_i represents the number of measurements carried out at the measuring station i and T_i is the time coordinate of the last measurement. Thus the number of all observations is

$$n' = \sum_{i=1}^{n} m_i .$$

As sampling at the measuring station i has not been carried out at the same time interval, further, the epoch of sampling may also differ in the number n of measuring stations, for the processing data are to be recasted. In the interest of easier management the following notations are introduced:

$$Z_{ijk}(x_{ij}, y_{ij}, z_{ij}, t_{ik}) = Z_{ijk},$$

[1]Central Institute for Mining Development, POB 115, H-1300, Budapest, Hungary

a_{ij} — is the weight factor of the estimation in the x, y plane which refers to the given coordinate z and epoch;

b_{ij} — is the weight factor of the estimation referring to epoch at a certain measuring station;

c_{ij} — is the weight factor of the estimation of the z direction at a definite measuring station.

A parameter estimation referring to any station characterized by the space coordinates x, y, z and by the time coordinate t is realized in three steps. First at all measuring stations an estimation is made for all possible epochs for the required coordinate z, estimations are based on samples relating to the same epoch. Let us estimate the value of Z^*_{izt}:

$$Z^*_{izt} = \sum_{j=1}^{m'_{iT}} c_{ij} Z_{ijt}, \tag{1a}$$

where m'_{iT} is the number of samples applied in the estimation:

$$m'_{iT} \leq m_i .$$

After this at each measuring station, an estimation relating to the required t epoch is done:

$$Z^*_{it} = \sum_{k=1}^{m''_i} b_{ij} Z^*_{izt}, \tag{1b}$$

where m''_i is the number of previously estimated values applied in the estimation:

$$m''_i \leq m_i .$$

Thereafter an estimation is made for the required station having the coordinates x, y:

$$Z^* = \sum_{i=1}^{N'} a_{ij} Z^*_{it}, \tag{1c}$$

where N' is the number of values estimated for the epoch t and for coordinate z applied in the estimation:

$$N' \leq n .$$

Summarizing the relations (1a), (1b) and (1c):

$$Z^* = \sum_{i=1}^{N'} a_{ij} \sum_{k=1}^{m''_i} b_{ij} \sum_{j=1}^{m'_{iT}} c_{ij} Z_{ijt} . \tag{1d}$$

At the same time, the following conditions are realized:

$$\sum_{j=1}^{m'_{iT}} c_{ij} = 1; \quad \sum_{k=1}^{m''_i} b_{ij} = 1; \quad \sum_{i=1}^{N'} a_{ij} = 1;$$

$$\sum_{i=1}^{N'}\sum_{k=1}^{m_i''}\sum_{j=1}^{m_{iT}'} a_{ij}b_{ij}c_{ij} = 1$$

Should the estimation be carried out using any distance-dependent weighing procedure, the calculation of coefficients is rather simple. If kriging is applied, a semivariogram of each sampling station or properly summarized (site-independent) z-directional one ($\gamma_z(h)$) and a similarly formed time-semivariogram ($\gamma_t(h)$) is required, further an other semivariogram bound to given z, t coordinates. These will be identified later.

Hence the environmental pollution process acting in time and space is described using a geostatistical model in which stationarity in time and space is assumed, further the z-directional and temporal change is considered site-independent and that in the x, y plane as isotropic. If universal kriging is applied the postulate of stationarity can be dropped.

2. Site-independent z-directional and temporal estimation

The vertical and temporal estimation may be done by interpolation or by extrapolation. The latter will be dealt with together with forecasting methods. In case of interpolation, estimation may be done by kriging, by any weighing procedure, spline function and by fully fitting trends.

In kriging, semivariograms describing the vertical change of the parameter ($\gamma_z(h)$) and that one expressing the temporal changes ($\gamma_t(h)$) are required:

$$\gamma_z(h) = \frac{1}{2\sum\limits_{i=1}^{n} M_{it}(z)}\sum_{i=1}^{n}\sum_{j=1}^{M_{it}(z)}\left[Z_{ijt} - Z_{i(j+h)t}\right]^2,$$

$$\gamma_t(h) = \frac{1}{2\sum\limits_{i=1}^{n} M_i(t)}\sum_{i=1}^{n}\sum_{j=1}^{M_i(t)}\left[Z^*_{izt} - Z^*_{iz(t+h)}\right]^2,$$

where

$M_{it}(z)$ — is the number of pairs of samples relating to the given z in an epoch t at a measuring station i,

$M_i(t)$ — is the number pairs of values estimated to a given t at the same measuring station i.

Adapting a theoretical function to the semivariograms $\gamma_z(h)$ and $\gamma_t(h)$ the theoretical semivariograms $\gamma_z^*(h)$ and $\gamma_t^*(h)$ are obtained. In case of $\gamma_t^*(h)$ this is interpreted in the following time limits:

$$\begin{aligned}
\gamma_t^*(h) &= 0 && \text{when} \quad h = 0\\
\gamma_t^*(h) &= f(h) && \text{when} \quad 0 < h \le a_t\\
\gamma_t^*(h) &= (C + C_0)_t && \text{when} \quad h \ge a_t
\end{aligned}$$

where

$f(h)$	—	is the theoretical function approaching the ascending branch of the empirical semivariogram,
a_t	—	is the sphere of time (and idea analogous to the idea of range),
$(C+C_o)_t$	—	is the threshold level of the temporal theoretical semivariogram.

When insted of kriging if no high accuracy is required any of the weighing procedures can be applied, the calculation of the semivariograms $\gamma_z(h)$ and $\gamma_t(h)$ is unrequired. For example when estimation is made with the inverted value of any power of the time difference, then at any measuring station the estimation for a discrete epoch is as follows denoting the time difference by Δt_k and the exponent by c:

$$Z_{it}^* = \sum_{k=1}^{m_i'} \frac{1/\Delta t_k^c}{\sum\limits_{k=1}^{m_i''} 1/\Delta t_k^c} Z_{izt}^* .$$

In spline interpolation, the starting point is that at a measuring station i at the epoch $t_{ij}(j = 1, 2, \ldots, m_i)$ the measured parameter values Z_{ijt} are given. These be points on the plane $(t_j; Z_{jt})$. Neighbouring points are to be connected with a third degree polynomial from which the value of the parameter is calculated for any epoch.

A fully fitting trend function can only be applied for interpolation when within the series of data the number of samples is relatively small. In the case of too many spots the postulation of full fitting encounters limits due to computer capacity.

Should the samples applied for interpolation be not of point-size but relating to any depth- (z-directional) or time-interval, then digitalization seems to be expedient.

3. Preparation of forecasts

If no possibility of interpolation exists a z-directional or temporal estimation, extrapolation is to be applied in which case different methods of forecasting can be used.

Forecasts of environmental protection are made in a four-dimensional space. But this task can be split into temporal, z-directional and x, y-plane forecasts. Should the temporal and z-directional forecasts been carried out, then the x, y-plane forecast reduces to an interpolation.

Among one-dimensional forecasts, numerous simple methods are available. For example different exponential forecasting methods, using the optimalization of the rate of change, the harmonized weighing, the moving average method, *etc.*

But among the possiblities of the one-dimensional forecasts, trend calculation is the simplest method. The calculated trend must not be full fitting. The procedure is as follows:

— check whether the investigated series of data has "white noise" character;

— if not, separate a not fully fitting polynomial trend and

— investigate the rest, whether they can be considered as "white noise";

— if not, then separate a not fully fitting harmonic trend again and

— repeat the investigation regarding "white noise";

— should the rest be no "white noise", then it is handled as a stochastic process and

— the forecast is to be calculated as the sum of the polynomial trend, the harmonic trend and the stochastic process;

— otherwise the basis of the forecast is the sum of the polynomial trend and of the harmonic trend.

Let us speak about the error of the forecast. The forecast is based on m_i' equidistant samples ($m_i' \ll m_i$). By producing test-forecasts for the sites of the known samples, the error of the forecast can be determined. The difference between the estimated and measured value be for a t_e forecast distance $\Delta_k(t_e)$. Calculating the average

$$\bar{\Delta}(t_e) = \frac{1}{s}\sum_{k=1}^{s} \Delta_k(t_e) \quad \text{and the scatter}$$

$$S_\Delta(t_e) = \sqrt{\frac{\sum_{k=1}^{s} \left[\Delta_k(t_e) - \bar{\Delta}(t_e)\right]^2}{n}}$$

where

s — is the number of forecasts

t_e — is the forecast period

at the given probability level, the error of the forecast can be calculated:

$$Z^*(t) \pm \left\{\bar{\Delta}(t_e) + t\left[S_\Delta(t_e) + \bar{S}_t\right]\right\}$$

where $\bar{S}_t$ is the average of the standard error (S_{t_i}) of the trends (of the functions applied for forecasting). Should be trends fully fitting ones, then $S_{t_i} = 0$ and $\bar{S}_t = 0$ and so the latter relation reduces to

$$Z^*(t) \pm \left[\bar{\Delta}(t_e) + tS_\Delta(t_e)\right].$$

A further reduction of the formula is possible when the change of the parameter has no definite trend. Namely in this case $\bar{\Delta}(t_e) \approx 0$ and if measurements are not loaded by regular errors the former relation is reduced:

$$Z^*(t) \pm tS_\Delta(t).$$

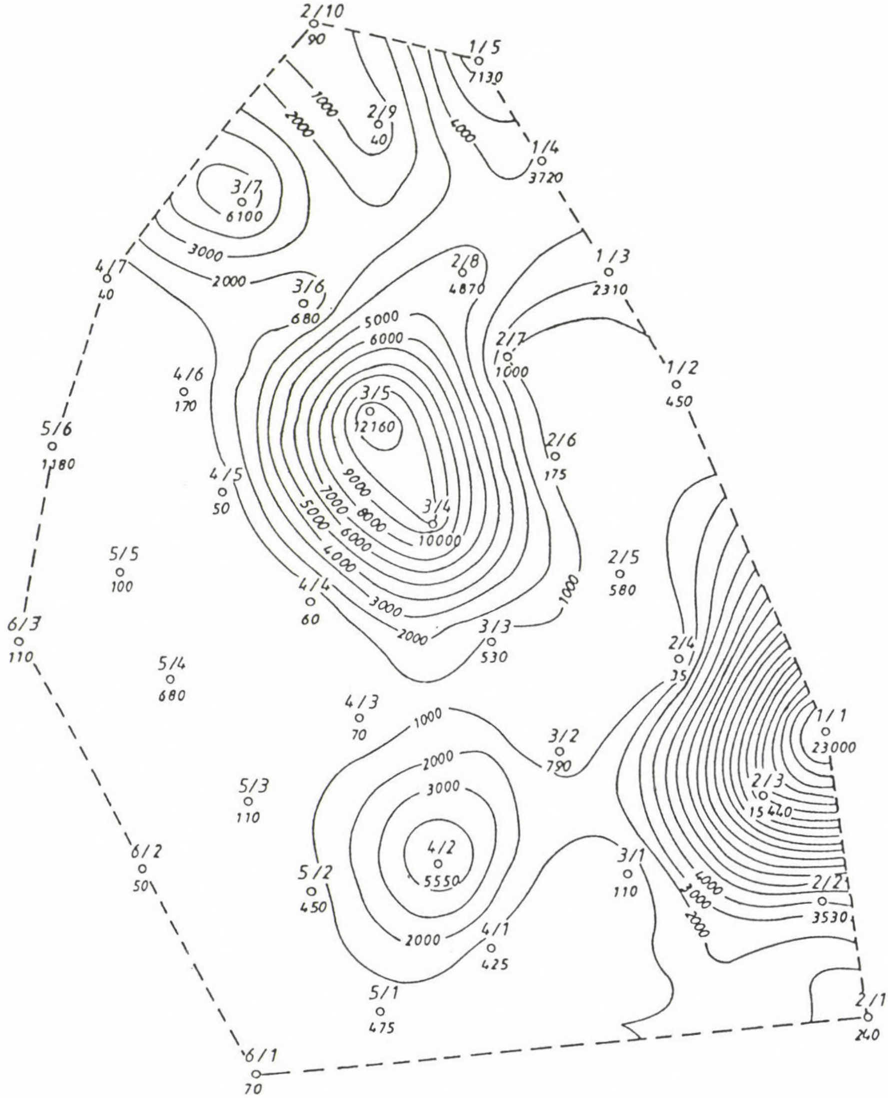


Fig. 1. Hydrocarbon pipeline. Polluting material content at a depth 1.25 m. Epoch: July 1, dimension: mg/kg, $M = 1 : 1000$

In the case of samples taken in irregular time intervals, the basis and error zone of the forecast may be determined as outlined above. The basis width (temporal length) be denoted by B, its starting point by $x_{\min}$ and the end by $x_{\max}$. The steps of the basis are denoted by L and the length of the forecast by E. The not fully fitting function applied on the samples within the basis B should be denoted by $f(x)$. The measured value of the parameter should be y, its value computed from the function $f(x)y'$. For all samples over the basis B in all possible conditions of the basis the series of differences $y' - y = d$ can be calculated. The equation of the straight line applied on these should be denoted by $f_1(\mathbf{x})$. Split the forecasting distance E in arbitrarily selected (e) intervals. Let us take first the interval with the starting point $x_{\max}$ and with the end $x_{\max} + e$ in which the lower limit is not included, but the upper one is so. Formulate in this the standard error

$$S_t = \sqrt{\frac{\sum_{i=1}^{m} (y_i' - y_i)^2}{m}}$$

where m represents the effective number of measurements within the interval. The coordinate of the standard error computed as above: $x_{\max}+e/2$. In the next interval having the limits $(x_{\max} + e)$ and $(x_{\max} + 2e)$ a standard error may be calculated again which has the coordinate $x_{\max} + 3/2e$ *etc.* Accordingly one got a concrete $f(x)$ function in the basis B by means of which the forecasting is to be done, further a straight line $f_1(\mathbf{x})$ and a standard error relating to the interval centres. Let us apply on the latter a function

$$f_s^t(\mathbf{x}) = ab^x$$

on a probability level t. Thus the error zone of the forecast is described with the following function $f_h(x)$:

$$f_h(x) = f(x) \pm \left[f_\ell(\mathbf{x}) + f_s^t(\mathbf{x}) \right] .$$

Now the calculation is repeated by increasing B. The increase of the basis length reduces the standard error, but the standard error is in a certain extent independent from the distance of the forecasting.

The direction of the tangential straight line of the function $f_s^t(\mathbf{x}) = ab^{\mathbf{x}}$ is 0 if $\mathbf{x} = -\infty$ and ∞ if $\mathbf{x} = \infty$.

Let us choose the allowed values of the directional tangent to w. Thus

$$ab^{\mathbf{x}} \ln b = w, \quad \text{and}$$

$$\mathbf{x} = \frac{\ln w - \ln a - \ln(\ln b)}{\ln b}.$$

In the case of a given basis length it is obtained up to which distance the forecasting can be made.

4. Case study

At a Hungarian hydrocarbon site a pipeline got broken. From the underground pipeline hydrocarbon oozed into the soil. To investigate the degree of pollution 38 boreholes were sunk in six rows. In these the polluting material has been measured in different depths and epochs. We have to make the kriged map showing the polluting material content on July 1, in a depth of 1.25 m.

As described, the temporal and vertical estimations have been done in all boreholes for July 1, in a depth of 1.25 m. The calculation is shown with the estimations in boreholes 3/5 and 1/1.

In the borehole 3/5 in the depth of 1.25 m the following data are known (Table I):

Table I.

Date of measurement	Time-coordinate (T) (day)	Pollution (mg/kg)
June 3	3	10642
June 18	18	11148
July 2	32	12247
July 9	39	12193
Aug. 23	84	12168

The time-coordinate is counted from June 1.

The equation of the site-independent semivariogram describing the temporal change is as follows:

$$\begin{aligned} \gamma_t(h) &= 0 & h &= 0 \\ \gamma_t(h) &= 500\left[1.5\tfrac{h}{23} - 0.5\left(\tfrac{h}{23}\right)^3\right] + 12 & 0 &< h \leq 23 \\ \gamma_t(h) &= 512 & h &\geq 23 \end{aligned}$$

Based on the 23 days time of influence, the points having the time-coordinates 18, 32 and 39 can be taken into account in the estimation. The estimated value of pollution is for $T = 31$ *i.e.* on July 1 12160 mg/kg.

In the borehole 1/1 measurements have been carried out at different epochs, thus also on July 1, in different depths. The measured values are as follows (Table II):

Table II.

Depth from the surface x (m)	Pollution (mg/kg)
0.1	2108
0.2	2650
0.4	2000
0.7	1100
1.0	3507

No measurements have been made deeper than 1 m, thus for the depth of 1.25 m a forecast has to be carried out. For the forecast a trend-function composed of a polynomial and a harmonic part will be applied with the following equation:

$$y = 1777.014 + 10600x - 42640x^2 + 38380x^3 + 88.52 \cos \frac{\pi x}{0.5} - $$
$$- 238.3 \sin \frac{\pi x}{0.5} - 185.7 \cos \frac{2\pi x}{0.5} + 234.0 \sin \frac{2\pi x}{0.5}.$$

The forecasted value is 23000 mg/kg.

After knowing the measured or calculated values in all boreholes for July 1, in the depth of 1.25 m the area semivariogram can be calculated. The equation:

$$\begin{array}{lcll} \gamma(h) & = & 0 & h = 0 \\ \gamma(h) & = & 13275650\{1 + \sin[\pi(h - 22.5)/45]\} + 816241.7 & 0 < h \leq 45 \\ \gamma(h) & = & 27367540 & h \geq 45 \end{array}$$

The kriged map of the polluting material content is shown in Fig. 1.

Naturally the map can be plotted from the interpolated values for any epoch. But in case of extrapolation (whether in depth or in time) one must be very careful since when the forecasting distance is by 1/3 longer than the basis of the forecast then the applicability of the calculated result is questionable.

References

Box G E P, Jenkins G M 1970: Time Series Analysis, forecasting and control. Holden Day. San Francisco, Cambridge, London, Amsterdam

Füst A 1990: Geostatistics (in Hungarian). Publication of the ELTE Chair of Applied Geology. University lecture notes. Budapest

Makridakis S, Wheelwright S C 1978: Interactive Forecasting. Holden Day. Inc. San Francisco, Düsseldorf, Johannesburg, London, Panama, Singapore, Sidney

Mecseki A, Molnár S, Szabó B 1987: Izvestiya Tsentralnogo Instituta po Razvitiyu Gornoj Promishlennosti, Budapest, No. 30, 117–120.

Varga S ed. 1980: Planning of Industrial Enterprises (in Hungarian). KG-INFORMAIK, Budapest

Acta Geod. Geoph. Mont. Hung., Vol. 27(2–4), pp. 399–404 (1992)

APPARENT RESISTIVITY COMPUTATIONS FOR THE MODELS WITH TRANSITIONAL LAYERS

M Israil[1]

[Manuscript received June 11, 1992]

Matrix method recently developed by the author has been successfully used for the computation of apparent resistivity over a horizontally stratified earth (1D case), where each layer has constant resistivity. Efficiency, accuracy, consistency and reliability of the method are well established (Israil 1989). Certain kinds of clay deposits contain varying amount of sand, the model assumes them as a layer in which resistivity varies as function of depth. The scope of the matrix method has been extended here for the models having layers in which the resistivity varies continuously with depth. The procedure for the computation of apparent resistivity for a generalised electrode configuration has been presented. A numerical example is demonstrated for a simple model with linear variation of resistivity *vs.* depth in a transitional layer.

Keywords: apparent resistivity; matrix method; resistivity kernel; transitional layer

Introduction

In 1D resistivity inversion, horizontally stratified earth is assumed with each layer having constant resistivity. Electrical well logging data indicate that the situation is more complicated in nature and a continuous variation of resistivity may be more realistic. An approximate mathematical function of depth is acceptable for the resistivity variation. Theoretical as well as field investigations have been carried out (e.g. Mallick and Roy 1968, Jain 1972, Niwas and Upadhyay 1974, Patella 1977, Koefoed 1979, Mundry and Zschau 1983 *etc.*). Raghuwanshi and Bijendra Singh (1986) presented generalised formulas considering three types of resistivity variations with depth in transitional layers as; (i) exponential variation, (ii) power law variation and (iii) linear variation. For numerical computations they have suggested a filter method (Ghosh 1970). Recently Israil (1989) presented a matrix method for the resistivity computations for horizontally stratified earth with each layer having constant resistivity. The matrix method is an efficient alternative to the filter method. It may be directly used for an arbitrary electrode configuration and for any abscissa distribution. Accuracy, stability and efficiency of the matrix method are well established for very high resistivity contrasts, for perfectly conducting and for perfectly insulating basement.

In the present paper the scope of the matrix method has been investigated for models having layers with a continuous variation of resistivity *vs.* depth. A

[1]Department of Geophysics, Kurukshetra University, Kurukshetra – 132119, Haryana, India

simple and straightforward procedure is presented for the computation of apparent resistivities for an arbitrary electrode configuration. Once the kernel values for the desired model are computed, they can be used to derive apparent resistivity values through a simple matrix operator relation. A numerical example is also demonstrated.

The matrix method

The matrix method is based upon the kernel approximation theory. Here a brief outline of the method is given, for details see Sri Niwas and Israil (1986, 1987). The resistivity kernel is represented by an exponential series with a finite number of terms. Accuracy can be controlled by using optimum values of exponents and of number of terms in the series representation of the kernel function which is written as

$$K(\lambda) = \sum_{i=1}^{p} f_i e^{-\xi_i \lambda}, \tag{1}$$

where ξ_i-s establish the position of the approximating function along the abscissa. Equation (1) may be substituted in the potential expression developed for horizontally stratified earth due to a point current source of strength I which is (Stefanescu 1930):

$$U(r) + \frac{I\varrho_1}{2\pi}\left[\frac{1}{r} + 2\int_0^\infty K(\lambda) J_0(\lambda r) d\lambda\right] \tag{2}$$

where
$J_0(\lambda r)$ is the Bessel function of the first kind and zero order;
$K(\lambda)$ is the kernel function
λ is a real integration variable.

Substituting Eq. (1) into Eq. (2) and by using the Lipschitz integral (Watson 1966)

$$\int_0^\infty e^{-\xi_i \lambda} J_0(\lambda r) d\lambda = \frac{1}{\sqrt{(\xi_i^2 + r^2)}} \tag{3}$$

the potential expression reduces to the following form

$$U(r) = \frac{I\varrho_1}{2nr}\left[1 + \sum_{i=1}^{p} f_i^p G_i(r)\right] \tag{4}$$

with $G_i(r) = \dfrac{2r}{\sqrt{(\xi_i^2 + r^2)}}$.

Using Eq. (4), a matrix operator equation is developed for computations of apparent resistivities for an arbitrary electrode configuration as;

$$\vec{R} = \mathbf{S}\vec{K}, \tag{5}$$

where $S = G(E^tE)^{-1}E^t$.

$$\vec{R} = \left[\frac{\varrho_a(r_1) - \varrho_1}{\varrho_1},\ \frac{\varrho_a(r_2) - \varrho_1}{\varrho_1},\ \ldots,\ \frac{\varrho_a(r_j) - \varrho_1}{\varrho_1}\right]^t \tag{6}$$

$$\vec{K} = [K(\lambda_1),\ K(\lambda_2),\ \ldots,\ K(\lambda_j)]^t \tag{7}$$

where t denotes transpose operation.

The matrix $\mathbf{E}$ has the elements $e^{-\xi_i\lambda_j}$, $(i = 1, 2, \ldots, p;\ j = 1, 2, \ldots, q)$.

The elements of matrix $\mathbf{G}$ are derived from the appropriate G function for a particular electrode configuration as,

$$G_i(r_j, mr_j) = \frac{m}{m-1}G_i(r_j) - \frac{1}{m-1}G_i(mr_j) \tag{8}$$

for a symmetrical array, and

$$G_i(r_j) = \frac{2(r_j/\xi_i)^3 + 2(1-b)(r_j/\xi_i)}{[1 + (r_j/\xi_i)^2]^{3/2}} \tag{9}$$

for a dipole array.

Equations (8) and (9) may be used for any particular electrode configuration with numerical values of m and b, respectively, defined in Table I.

Table I.

symmetrical array		dipole array	
value of m	particular array	value of b	particular array
2.0	Wenner	0.0	azimuthal dipole
1.1	Schlumberger	0.5	radial dipole
∞	two electrodes	0.333	perpendicular dipole
1.1	half Schlumberger	$\frac{\cos^2\Theta}{3\cos^2\Theta - 1}$	parallel dipole

Matrix formulation of the problem

A generalised potential expression is for 1D models having layers with continuous variation of resistivity *vs.* depth (Raghuwanshi and Singh 1986)

$$U(r) = \frac{I\varrho_1}{2\pi}\left[\frac{1}{r} + 2\int_0^\infty K_{2n}(\lambda)J(\lambda r)d\lambda\right] \tag{10}$$

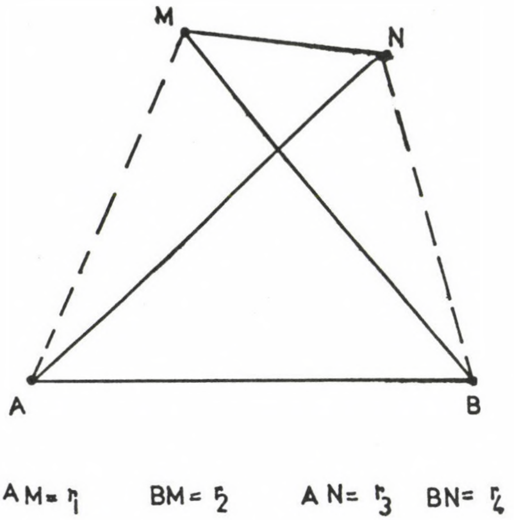


Fig. 1. General four electrode array

where

$$K_{2n}(\lambda) = \frac{\left(\frac{V_2/C_1-1}{V_2/C_1+1}\right)^{e^{-2\lambda h_1}}}{\left[1-\left(\frac{V_2/C_1-1}{V_2/C_1-1}\right)^{e^{-2\lambda h_1}}\right]} \tag{11}$$

V_2 and C_1 can be obtained using appropriate recurrence relations for arbitrary variation of the resistivity with depth in transitional layers. Applying the principle of matrix method to solve Eq. (10) analytically, the kernel function can be written in the form of an exponential series up to a finite number of terms p as given in Eq. (1). This allows the analytical solution of Eq. (1). Thus the potential expression reduces to the form of Eq. (4). For an arbitrary four-electrode array comprising two current electrodes (A, B) and two potential electrodes (M, N) (Fig. 1) the apparent resistivity can be written as

$$\varrho_a(r_1,r_2,r_3,r_4) = 2\pi r_1r_2r_3r_4(r_2r_3r_4 - r_1r_3r_4 - r_1r_2r_4 + r_1r_2r_3)^{-1}\frac{U_{\mathrm{M}} - U_{\mathrm{N}}}{I} \tag{12}$$

(assuming electrode separations AM $= r_1$, BM $= r_2$, AN $= r_3$ and BN $= r_4$).

Substituting the values of U_{M} and U_{N} in Eq. (12), the apparent resistivity is

$$\varrho_a = \varrho_1[1 + f_iG_i(r_1,r_2,r_3,r_4)] \tag{13}$$

where

$$\begin{aligned} G_i(r_1,r_2,r_3,r_4) &= (r_2r_3r_4 - r_1r_3r_4 - r_1r_2r_4 + r_1r_2r_3)^{-1}\cdot \\ &\cdot[r_2r_3r_4G_i(r_1) - r_1r_3r_4G_i(r_2) - r_1r_2r_4G_i(r_3)r_1r_2r_3G_i(r_4)] \end{aligned} \tag{14}$$

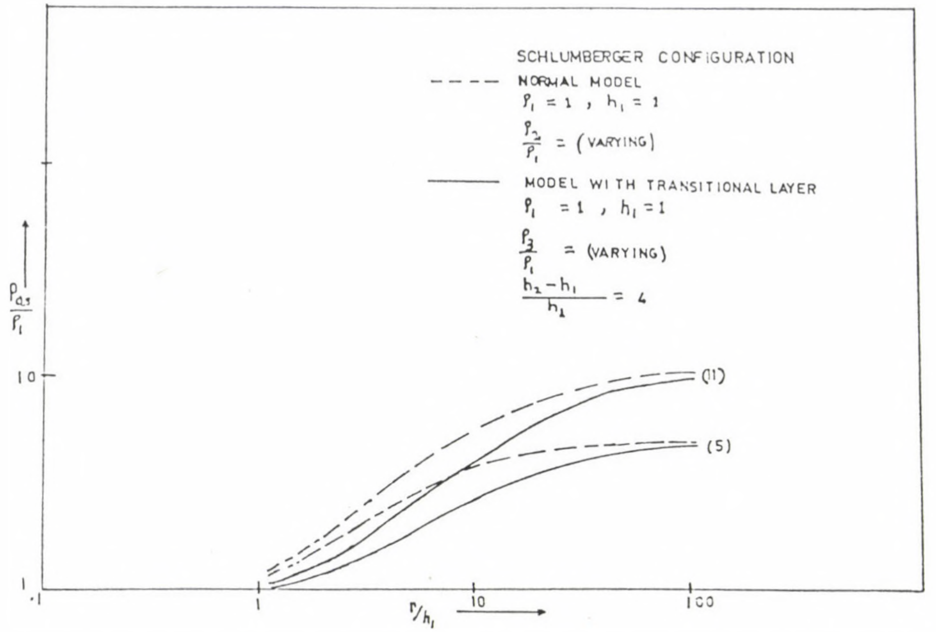


Fig. 2. Apparent resistivities for a normalised model with transitional layer of thickness $(h_2 - h_1)/h_1 = 4$ and a normal model with a transitional layer thickness of zero. Other layer parameters are: $\varrho_1 = 1$, $h_1 = 1$, $\frac{\varrho_3}{\varrho_1} = 5$ and 11

and $G_i(r)$ is given by Eq. (4).

Equation (13) reduces to an apparent resistivity expression for a particular configuration and Eq. (14) reduces to Eqs (8) and (9) for symmetrical and dipole configurations, respectively.

Results and discussion

The apparent resistivity as a function of electrode separation is analogous to the kernel as a function of $1/\lambda$. By writing the kernel function at discrete values of $1/\lambda_j$ $(j = 1, 2, \ldots, q > p)$ Eq. (1) is in matrix form,

$$\vec{K} = \mathbf{E}\vec{F} \tag{15}$$

where vector $\vec{K}$ contains the tabulated values of the kernel computed using Eq. (11) for the desired model with transitional layers at discrete values of $1/\lambda_j$ $(j = 1, 2, \ldots, q)$. The values of ξ_i-s are chosen on the basis of numerical experiments and fixed for all types of model calculations, the starting value of ξ_i is 0.2 which is increased in by a factor of 2 in each step. Twelve values of ξ_i-s are sufficient for the maximum abscissa value of 500.0 m as it produces an error of the order of 0.001 percent in the kernel approximation. Equation (15) can be made overdetermined by setting $(q > p)$ and the values of f_i-s obtained from the least square solution of

Eq. (15). The set of f_i-s is used to compute apparent resistivity values through Eq. (13). Following the procedure outlined above, the apparent resistivities are computed for normalised models with the parameters:

$$\begin{aligned} \varrho_1 &= 1, \quad h_1 = 1 \\ \varrho_2 &= \varrho_1 + \alpha(z - h_1) \text{ for } h_1 < z > h_2 \\ \frac{\varrho_3}{\varrho_1} &= 5 \text{ and } 11 \end{aligned}$$

and $(h_2 - h_1)/h_1 = 4$.

The computed apparent resistivity values for the models are presented in Fig. 2 along with normal models (transitional layer thickness is zero i.e. $(h_2 - h_1)/h_1 = 0.0$). Following the same method, the apparent resistivity can be computed for any model with an arbitrary variation of the resistivity with depth in the transitional layers using appropriate kernel values in Eq. (15).

Conclusion

The utility of the matrix method has been investigated for the models having layers with arbitrary variation of resistivity with depth. The method presents an efficient alternative to other existing methods. The present method simplify the computational procedure and extends the scope of the matrix method. This will not help in identifying the suitability of a particular model over another. It rests entirely upon the interpreter to select a particular model which is most suitable to represent the realistic situation.

Acknowledgements

The author is thankful to Professor K N Khattri and to Professor Sri Niwas for creating interest to solve research problems.

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 1989: Matrix method for the interpretation of direct current resistivity data. Ph.D. Thesis, Roorkee University

Jain S C 1972: *Geoph. Prosp.*, 20, 283–292.

Koefoed O 1979: *Geoph. Prosp.*, 27, 862–868.

Mallick K, Roy A 1968: *Geoph. Prosp.*, 16, 436–446.

Mundry E, Zschau H J 1983: *Geoph. Prosp.*, 31, 810–828.

Niwas S, Upadhyay S K 1974: *Geoph. Prosp.*, 22, 279–296.

Patella D 1986: *Geoph. Prosp.*, 34, 109–123.

Raghuwanshi S S, Singh B 1986: *Geoph. Prosp.*, 34, 409–423.

Sri Niwas, Israil M 1986: *Geophysics*, 51, 1594–1602.

Sri Niwas, Israil M 1987: *Geophysics*, 52, 1412–1417.

Stefanescu S S (in collaboration with Conrad and Marcel Schlumberger) 1930: *Le Journal de Physique et le Radium*, 7, Series 1.

Watson G N 1966: A treatise on the theory of Bessel functions. Cambridge Univ. Press

DFG DEUTSCHE FORSCHUNGSGEMEINSCHAFT (German Research Community) F SEIFERT ed.: *Kinetik gesteins- und mineralbildender Prozesse* (Kinetics of rock- and mineral-forming processes). Publication Nr. 19 of the Senate-Committee for Geoscientific Research Community. VCH Verlagsgesellschaft mBH, Weinheim, 1991. p. 106

The Senate-Committee of the German Geoscientific Research Community, acting as part of the Deutsche Forschungsgemeinschaft (German Research Community), regards as his goal and set as target to support research which covers several disciplines within geosciences.

The research program called "Kinetics of rock- and mineralforming processes" was started in 1980, and was carried out from 1981 till 1988.

The Geoscientific Committee dealt with the planning of this substantial program at several meetings.

In an earlier substantial program of DFG the task was to determine the mineral equilibrium in the Earth's crust and mantle at great pressure and high temperature as well as the physical properties and thermodynamic behaviour of different minerals and mineral parageneses. It was decided to study the processes of the genesis of rocks and minerals as well as their metamorphosis within the framework of a new research program.

The research teams choose as target the Ballachulish Pluton in Scotland and its contact zone of an especially advantageous geological structure which was formed by the metamorphosis of rocks having accompanied the Caledonian orogenesis. The magmatic rock body has a relatively simple form and the metamorphous rocks in the contact zone are various and a well separated regional metamorphous effect is recognized being earlier than the later contact metamorphosis.

The research substantial was coordinated by F Seifert, professor at the University Kiel. Research dealing with igneous rocks of the Ballachulish Pluton and with metamorphosis rock formations in its contact zone was directed by G Voll (University Köln). The research was carried out by a specialized staff of 183 persons with a cost of 7.9 million DM.

The publication of 106 pages reviewed here gives a short review on the results; some results were published earlier, while others wait for publication in different journals.

The community character of the research is especially evident for this rock-kinetic program. Many excellent scientists from different disciplines of geosciences joined this program. The fact is significant that scientists of other scientific branches showed interest in this research, even who else do not deal with geoscientific problems. Thus very effective discussions took place. These connections were continued after the end of the program, too.

A closing colloquium was held in Bad Honeff on January 18, 1989. The participants declared that by means of the rock-kinetic program a successful cooperation has been established which grew during its running time to an internationally accepted research area. The results of this comprehensive research are summed up in this booklet. The papers will be published soon in a more detailed form by Springer.

The booklet includes a detailed list of papers with results of the research. The distribution of papers is as follows: 20 papers on general problems, 112 on basic research, 38 diploma and final works and dissertations as well as 6 research reports. 167 lectures were held; 16 colloquia and workshops took place.

The rock-kinetic program aimed to learn better the processes of the genesis of rocks and minerals as well as their metamorphosis, and to solve the dynamic process behind actual geological problems. These time-dependent processes could have been cleared only by the coordinated field work and corresponding laboratory investigations.

The application of the experimental results needed, however, a critical extrapolation from the point of view of geology for the interpretation of the processes taking place in the nature, since experiments could not be carried out on scales of million years and of rock volumes of several km^3. Consequently the exact description and theoretical definition of the kinematic processes led with limited accuracy to conclusions about the processes in the nature. This was supported by the geophysical extreme value investigations, field observations, further by experiments on identical mineral associations. For these nuclear and macroscopic properties of mineral rock-components were necessary.

Nuclear properties were *e.g.* faults in the lattice structure of crystalline minerals as certained by spectroscopic methods (*e.g.* by Mössbauer and Rahman or infrared spectroscopy) or by nuclear magnetic resonance (NMR) and electron-spin resonance (SER) measurements. Sites of greater lattice faults are observable by electron microscope, while the investigation of lattice structures by X-ray delivered particular data on temperature- and pressure-dependent crystal lattice structures.

Conventional mineral investigation methods played an important role, too, such as by mineralogical-petrological microscope. These methods that the field of rock kinetics is broad, thus researchers working in crystallography, in mineralogy and in material-fabric closely cooperated with the field groups making petrographical, geochemical and geophysical observations.

Heads of groups give a short report on the results of their work:

1. Kinetics of homogeneous reactions:

   Kinetics of exchange of Al and Si in alkaline feldspars (experimental method). Kinetics of the change-over of the same elements in the sanidines of Eifel Mts. Sites of paramagnetic lattice faults and their effect on the kinetics of Al- and Si-distribution.

   Description of K-feldspar distribution observed in the rocks of the Kagenfels intrusion (Germany).

   Trace element contamination in quartz and its effect on the kinetics of the growth of crystals. Their effect on the development of rock fabric and crystal properties.

Incorporation of Fe^{3+} cation into the amethyst crystal and into the artificially produced quartz with iron content, further the kinetics of Fe^{3+}-containing lattice faults in the chrysolite. Diffusion of argon in mica. Kinetics of rock-glasses, volume relaxation at stiffening SiO_2-rock-glasses.

2. Nucleation and crystal growth:

   Transforming and separating processes in artificial and natural pyroxenes.

   Development and cessation velocity of crystal lattice faults in case of Mn- and Mg-amphiboles.

   Separation of garnet of pyroxene origin in ultramaphic rocks, its mechanism and kinetics.

   Effect of the structure of grain- and phase-borders to the metamorphosis kinetics of minerals.

   Kinetics of reversible chemical reactions of hydrothermal systems with feldspar content.

3. Heterogeneous equilibrium:

   Kinetics of element distributions between ternary feldspars and liquid phases present in this system, with the mechanism of isotope exchange taking place in metamorphic rocks, with the kinetics of the genesis of magmatic dykes and bedded veins and with the Na and K separation of feldspars.

4. Genesis of the Ballachulish Pluton

   Cooling process of the magmatic rock formation could be exactly described and temperature changes in the adjacent host rocks determined.

Naturally, all investigation methods could not be used for each research topic without restrictions, in spite of the fact that considerable results would have been reached by them in rock kinetics. This is valid for working groups which apply their investigation methods in nuclear or submicroscopic area. Namely here the importance and effect of the differing phenomena can only be cleared by further investigations.

The interdisciplinary character of up-to-date rock kinetics required a close co-operation between the research groups of adjacent scientific areas. Some sciences, as *e.g.* solid state physics and chemistry as well as ome fields of material sciences were not included into this program with proper weight. Hence it is desirable that the different related sciences should show appropriate interest in the future toward problems concerning geosciences.

The research groups consider it very important that the work should be continued and they hope proper support and understanding.

They consider a most important task to make available scientific information and research experience. They consider therefore the organization of a training course and the publication of its material in the form of a book, as very important.

M Bán

W Schröder, M Colacino, G Gregori eds: *Exploring the Earth: Progress in geophysics since the 17th century.* Interdivisional commission on history, IAGA, Bremen-Roennebeck, Germany, 1992

These volume contains the papers presented at the conference on history of IAGA (and geophysics in general) in August 1992, Vienna. The volume continues a series of similar publications, and several authors report on progress in previously started projects. The contributions can be divided into three groups: on people, on events and on history s.s.

The first group includes an account on the role of women in astronomy and geophysics in France and in the earliest time, in the 17th century in Central Europe (Mrs Hevelius and Maria Cumitz), followed by a long series of French scientists. John Allan Broun was a pioneer of geomagnetism in India, his life is treated by N J Skinner. An excellent person in the compilation of geomagnetic indices, Father Mayaud, lived a few decades ago. His activity is presented by Menvielle and Mazaudier. Schröder and Wiederkehr continue the history of G von Neumayer's life, here mainly his work in meteorology and geomagnetism. A most interesting aspect is his connection with polar research, especially with the first international polar year. G Fanselau is with Mayaud the second man with whom I had been lucky enough to have personal contact. His life is commemorated by W Mundt. C T Kwei and his activity in connection with the spread F conclude this group.

Events are best examplified by a most vivid account by C O Hines on the discovery of the gravity waves. L Eötvös, the excellent Hungarian geophysicist, who constructed the torsion balance, was also active in Paleomagnetism or more correctly, archeomagnetism (P Márton). Reports on three institutions from different ages may be added to this group: A cooperation called Global Interdisciplinary study on Environmental History (GISEH) (Colacino et al.), Isostatic Institute of the IAG (J Boavida) and the Cimento Academy (M Colacino, M R Valensise).

The group "history" includes Murty's Pioneers of geophysical research in India, Meloni et al's Historic measurements of the geomagnetic field (with valuable ancient data, Valensise's and Colacino's report on the climate idea in the 18th century, especially on Frencesco Alfarotti, Bernhardt's study on atmospheric physics at the Berlin university in the last century. M P Pavese collected a lot Latin inscription, connected with meteorological/geophysical events, e.g. volcanic eruptions, inundations *etc.* which led to interesting conclusions.

I left to the end two contributions which caught my special interest: L M Barreto on geomagnetism in Brazil, where he enumerated several Jesuits in the 18th century who carried out geomagnetic research there. They include Ignacio Azentmartony, more likely Szentmartony, being a clearly Hungarian name. He was previously unknown for me. Then A Udias reported on early geophysical observatories, among those in the 18th century that in Tyrnau, today Trnava, in Hungarian Nagyszombat. The many names illustrate the history in Central Europe, as cities had German, Hungarian and Slovak names. (This city is now in Slovakia.) The instruments for this and several other (astronomical) observatories were made by M Hell, an excellent astronomer, who observed the Venus transition in Norway with a Danish royal expedition. The same paper mentiones Father Fényi, director of the Kalocsa geophysical-astronomical-meteorological observatory, an excellent solar physicist, who was born in the city Sopron where these words are written.

As a whole, this book contains very remarkable contributions, and they are worth reading for everybody interested in the history of geophysics.

J Verő

F STEINER ed.: *The Most Frequent Value.* Akadémiai Kiadó, Budapest, 1991, p. 315

Since computation technics have increased rapidly in the second half of our century, new perspectives have been opened up before the development of mathematical methods of processing great data systems. Nowadays the iteration methods of mathematical statistics based on the new computation possibilities become more and more frequent.

Most of the methods proposed, however, have not the solid theoretical bases which characterize the classical methods of mathematical statistics.

The methods defined in a totally *ad hoc* way and analyzed in the book "Robust Estimates of Location" by Andrews et al. give a good example for that, and the inadequacy of the classical theories is proved unambiguously in the book cited. In such a sense the series of papers written by Ferenc Steiner and his coworkers means the beginning of a new era in applied statistics, as they deal with the data processing methods in accordance to the uptodate concept of mathematical statistics and give critical analyses of the characteristics of classical mathematical statistics.

The book titled "The Most Frequent Value" can be of good use both for those who apply the methods in practice, and for those having deeper interest in theoretical problems.

The book includes papers giving the theoretical basis of the methods as well as papers showing examples for practical application. Since Ferenc Steiner and his co-authors work on the field of geosciences, the practical examples are taken from this field (*e.g.* adjustment of telluric measurements), nevertheless the theoretical statements as well as the experiences in practice can also be used by those working on other fields.

The concepts expectance and scatter are critically analysed by the authors as statistical characteristics. They conclude that these characteristics are not sufficient from the professional point of view. The Gaussian distribution is by far not as frequent in the practice as having been thought so far, consequently it is not always suitable to use the corresponding characteristics (as expected value, scatter), or to say in a more general way: not even to use the method of least squares. In the papers dealing with robust statistics the methods of classical statistics are criticised in a similar way, but readers of this book must not have rudiments in this field, since the new concepts (most frequent value, dihesion) are introduced heuristically, and it is shown only thenafter, how these concepts can be organically deduced from information theory, practically on the basis of the very obvious requirement of minimizing the loss of information. Statistical methods based on the concept of the most frequent value get in this way a very strong theoretical fundament (which is missed by the *ad hoc* methods mentioned earlier), further the application of a well defined method is proposed in the book in spite of the very general results of information theory, and mathematical statistics, respectively, got in recent decades, which do not define concrete algorithms for the practical specialist, as knowledge about the types of error distributions would be necessary for that, too. From this point of view it must be emphasized that in the book the studies are based on such a type family which gives good models for the manyfold error distributions occurring in geosciences, and at the same time this leads to the computation technically most simple method out of the earlier known iteration weighting methods.

It is worth emphasizing interesting results:

a) the error decrease corresponding to the law of big numbers, proportional to $1/\sqrt{n}$, and to be taken in an asymptotic sense (where n is the number of sample elements) is fulfilled for the most frequent values already at small n-s, even if at the given type of error distribution the law is no fulfilled for the averages in the asymptotic sense, either;

b) the misleading effect of the concept scatter is made obvious in the book by the general investigation of the indefinite (Heisenberg) relation of the theory of Fourier transforms;

c) the method of computing the most frequent values can be generalized without any difficulty to adjustments with more variables;

d) at the computation of the most frequent values, or at the corresponding adjustment, the accuracy of the estimations can be further increased by taking the *a priori* weights of the data into account;

e) the last figure in the book shows a statistical efficiency of nearly 100 percent of the estimations after the most frequent values within a broad range of error distribution types (the same figure is on the cover), by comparing it with the efficiency of the classical statistics which rapidly decreases when moving away from the Gaussian type (it is well known that a greater statistical efficiency can have a considerable cost-decreasing effect on the measurements);

f) the table at the end of the book gives an overview about the mutual connections on the basis of which the recommended new method can be easily compared *e.g.* with the classical statistical method of minimizing the L_2-norm of the deviations.

Great importance is laid in the book to give detailed mathematical deductions in some cases. This cannot be done of course in all respects, due to the limits of the book, but by giving a complete bibliography with respect to the new method, it enables for the reader the adequate orientation.

In the era of the expansion of different robust methods it is especially recommendable to study this book of simple and well understandable composition, since it clears certain theoretical questions and in addition it shows the concrete way to an advantageous practical realization.

At the end of the book there is an index of subjects referring to papers and appendices, respectively, which are the most interesting for the reader.

L Zilahi-Sebess

E BUSCHMANN: *Gedanken über die Geodäsie. Einige naturwissenschaftliche, technische, philosophische und wirtschaftliche Aspekte.* Verlag Konrad Wittwer, Stuttgart, 1992. 152 pp, 2 figs, 5 tables

This is a most interesting book which describes some ideas on geodesy on the basis of natural scientific, technical, philosophical and economical aspect. Beside geodesist it is also worth reading for other geoscientists and even for people with wider field of interest.

Chapter 1 deals with the definition and task of geodesy. One can find beside the author's wording the definitions of different famous geoscientist.

Chapter 2 gives view about the geodetical recognition process. It deals with experiment, hypotheses, theory, model, observation, technical devices, geodetical information *etc.*

Chapter 3 treats the metrological basis of geodesy. One can read about metrological principle, reference systems, the movement of reference systems and survey objects, measurable elements *etc.*

Chapter 4 deals with the use of recognitions.

J Somogyi

Surveying. Volume 1. Baltimore, 1991. 300 pp, 43 figs, 43 tables

This book is the first volume of the technical papers of the 1991 ACM-ASPRS Annual Convention held in Baltimore.

The book collects the papers presented on surveying. The 46 papers deal with different problems on surveying, among others with leveling, geodetical computations, digitizing, measuring (electronic, GPS), automatization, cadastral *etc.*

The papers refers to up-to-date problems.

J Somogyi

Cartography and GIS/LIS. Volume 2. Baltimore, 1991. 353 pp, 87 figs, 16 tables

This book is the volume 2 of the technical papers of the 1991 ACSM-ASPRS Annual Convention held in Baltimore.

The proceedings collects 38 papers and 10 abstracts presented on cartography and GIS/LIS dealing, among others with digital topographic data base, cartographic line simplification, map generalization, digital line map, spatial data quality, GIS, integrated data analysis, automated mapping.

The collected papers deal with the actual problems of cartography and GIS/LIS and could be very useful who are interested on this topics.

J Somogyi

Remote Sensing. Volume 3. Baltimore, 1991. 522 pp, 122 figs, 84 tables

This book is the volume 3 of the technical papers of the 1991 ACSM-ASPRS Annual Convention held in Baltimore.

The book collects 59 papers and 5 abstracts presented on remote sensing dealing among others with image classification, use of satellite data, Landsat, interpretation, window problems, automatization, information management, remote sensing, GIS integration, digital data processing, computer programs, mapping.

In the last few years the remote sensing has played a very important role in geosciences. The proceedings summarize the present day situation and gives a very useful help to people who are interested in this field.

J Somogyi

GIS. Volume 4. Baltimore, 1991. 263 pp, 55 figs, 24 tables

This book is the fourth volume of the technical papers of the 1991 ACSM-ASPRS Annual Convention held in Baltimore.

The book collects papers presented on GIS. The 31 papers deal with different problems on GIS. For instance: determination using and displaying three dimensional data, 3-D GIS, GIS based classification, integrated GIS, raster vector integration, geocoding of remote sensing data.

The proceedings give a very informative summary about the GIS problems in USA.

J Somogyi

Photogrammetry and Primary Data Acquisition. Volume 5. Baltimore, 1991. 448 pp, 163 figs, 51 tables

This book is the fifth volume of the technical papers of the 1991 ACSM-ASPRS Annual Convention held in Baltimore.

This volume contains 49 papers presented on photogrammetry. The topics of the papers are the following: model orientation, digital block adjustment, multispectral imaging, close range photogrammetry, analytical calibration, panoramic photography, image matching, SPOT satellite imagery, GPS photogrammetry, orthophotos.

The collected papers refer to very actual problems.

J Somogyi

Auto-Carto 10. Volume 6. Baltimore, 1991. 444 pp, 155 figs, 38 tables

This book is the sixth volume of the technical papers of the 1991 ACSM-ASPRS Annual Convention held in Baltimore.

The book collects 30 papers presented on Auto-Carto 10 dealing among others with data base for multi scale GIS, map generalization operations, spline functions to represent geometric forms, orderings, automatic digitization, thematic mapping from imagery, answers to spatial questions, spatial reasoning, spatial analysis in GIS, spatial data management, error in digital elevation model, cartographic symbolization concerned with GIS, display operations in GIS, GIS independent user interface environment, spatial overlay, 3D information systems, integration of spatial objects in a GIS.

J Somogyi

CONSTANCE BABINGTON SMITH: *Air Spy. The story of photo intelligence in World War II.* American Society for Photogrammetry Foundation. Classic reprint. 266 pp, 53 photos

The author was a member of the Women's Auxiliary Air Force for six years, attaining the rank of Flight Officer. She started the Aircraft Section of the Central Interpretation Unit at Medmenham and was in charge of it until 1945.

In her very interesting book she gives a brief historical review about the development of photographic reconnaissance. She describes the important work of photographic pilots and interpreters in the Second World War. The reader can be persuaded of their contribution giving useful informations for the management of war.

J Somogyi

Proceedings: Resource Technology 90, Georgetown University Conference Center Washington, D.C., November 12–15, 1990. Published by American Society for Photogrammetry and Remote Sensing. 830 pp, 198 figs, 68 tables

This book contains papers and poster abstracts from the Second International Symposium on Advanced Technology in Natural Resources Management held in Washington, D.C.

The topics of the papers are the following: Model Driven Visual Simulation, Fire Management Technology, International Projects, Intensive Forest Management–Biomass for Energy and Strategic Data Management. Using GIS Technology for Managing, Planning and Monitoring Projected Areas. Soil Survey. Wildlife. The Social and Human Dimensions of Advanced Technology. Forest Management. Applications to Ecological Modeling: New Solutions to old Problems. Forest Pest Management. Minerals. GIS Applications.

The proceedings contain 85 papers and 29 poster session abstracts from the listed topics.

J Somogyi

R A MCDONALD ed.: *Proceedings: Space Imagery and News Gathering for the 1990s: So what?* Published by American Society for Photogrammetry and Remote Sensing. 121 pp, 14 figs, 15 photos, 3 tables

The proceedings contain papers from the Symposium on Foreign Policy and Remote Sensing held in the Patterson School of Diplomacy and International Commerce University of Kentucky, Februar 24–25, 1989.

The topics of the papers are the following:

1. Introduction and Overview
2. Broadcasting and Journalism Perspective
3. Technological Availability of Space Imagery
4. Human Factor Constraints on Using Space Imagery
5. Legal and National Security Implications of Using Space Imagery
6. Congress and its Role
7. Appendices

The proceeding contain 11 papers from the listed topics.

J Somogyi

SCHRÖDER W, TREDER H J eds: *Theoretical concepts and observational implications in meteorology and geophysics (Selected papers from the IAGA Symposium to commemorate the 50th anniversary of Ertel's potential vorticity).* Interdivisional commission on history of the International Association of Geomagnetism and Aeronomy (IAGA), Bremen–Roennebeck, 1993, 206 pp

Ertel discovered his important theorem in 1942, and the potential vorticity that bears his name helps to understand dynamics of flows. It is one of the basic concepts in theoretical meteorology and oceanography. Following the publication of Ertel's selected papers by the same way, the present volume reproduces Ertel's basic paper and its application as Bjerknes' vorticity theorem, the present book contains a collection of earlier papers dealing with Ertel's theorem and its consequences, further several papers from the 1991 Vienna symposium which commemorated the 50th anniversary of Ertel's publication.

Part 1 contain in addition to the mentioned papers by Ertel 10 earlier (1942–1989) papers on different general aspects of the theorem, including its derivation, proof, generalization, magnetic analogue, relativistic form *etc.* 5 papers from the 1991 symposium, presenting modern achievements in this field. Part 2 is devoted to meteorological applications, cyclonogenesis, front development and ozone depletion. Ertel's short biography concludes the volume with a list of publication.

This very useful volume will interest everybody dealing with theoretical meteorology and oceanography and also those who had opportunity to meet Hans Ertel personally.

J Verő

Announcement

Cultural Heritage Collected in Libraries of Geoscience, Mining and Metallurgy – Past, Present and Strategy for the Next Millenium

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 is 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.

We thank you in advance for your answer.

Yours sincerely – Glückauf!

Dr. Peter SCHMIDT
Bibliothek
Bergakademie Freiberg
Schliessfach 47
D-0-9200 FREIBERG (Sachsen)
Tel: 51 32 35

Dr. Lieselotte JONTES
Universitätsbibliothek
Montanuniversität Leoben
Franz-Josef-Strasse 18
A-8700 LEOBEN
Tel: (03842) 42 555/275

Akadémiai Kiadó, Budapest

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

VOLUME 27

AKADÉMIAI KIADÓ, BUDAPEST
1992

CONTENTS

Number 1

Book review

Numbers 2–4

Book reviews

AUTHOR INDEX

KEYWORD INDEX

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 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.* Fifty *offprints* are supplied free of charge.