Acta Geodaetica et Geophysica Hungarica

VOLUME 32, NUMBERS 1–2, 1997

Akadémiai Kiadó, Budapest

AGG 32 (1–2) 1–262 (1997) HU ISSN 1217–8977

ACTA GEODAETICA ET GEOPHYSICA HUNGARICA
A Quarterly Journal of the Hungarian Academy of Sciences

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

AGG is published in yearly volumes of four issues by

AKADÉMIAI KIADÓ
H-1117 Budapest, Prielle K. u. 19–35

Manuscripts and editorial correspondence should be addressed to

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

Subscription information

Orders should be addressed to

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

Subscription price for Volume 32 (1997) in 4 issues US$ 112.00, including normal postage, airmail delivery US$ 20.00.

CONTENTS

Book reviews

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 3–13 (1997)

MODELLING OF SOLIDIFICATION PROCESSES IN MAGMA CHAMBERS COOLED FROM ABOVE

PEEYUSH TEWARI[1], U S GUPTA[1], SRI NIWAS[2]

[Manuscript received October 1, 1996]

Moving Boundary Formulation is used to analyze one dimensional model of magma chamber which gradually cools and solidifies after emplacement in the region. Cooling of a layer of pure melt (magma chamber) from above leads to a temperature profile that increases with depth and is thermally unstable. The vigor of convection thus produced, is determined by Rayleigh-Nusselt number based on the finite depth of liquid layer remaining at a particular time and some other parameters like initial heat flow and Stefan number etc. We solve the nonlinear coupled set of differential equations by Runge-Kutta-Fehlberg method and results for different parameters are compared for phase boundary movement.

The effect of the phase boundary or solidification process is shown for various parameters in detail. It is observed that the evolution process is much smoother when the exponent in the Rayleigh-Nusselt relation is decreased i.e. < 0.33.

Keywords: magma chamber; moving boundary problem; solidification; Stefan number

1. Introduction

The diversity of igneous rocks observed at the surface of earth motivates one to understand the process of the solidification of magma coming from mantle and related convection processes in the magma chamber. Compositional differences in a region in magma chambers are supposed to drive convection and its cooling causes large scale zonation of minerals (Huppert and Worster 1992). For understanding and investigating the solidification process an in-depth study of various fluid mechanical processes is essential. Among the factors that effect fluid mechanical processes, convection pattern (i.e. whether time dependent convection or not), multi-component convection and conductivity changes etc. have been investigated by many workers from diverse fields such as applied mathematics, engineering, metallurgy and geology. In the present paper we are studying the effect of varying viscosity on the solidification processes by numerical simulation and formulating it as moving boundary problem (MBP) which are also known as Stefan problems (Crank 1984).

Background

Moving boundary problems were initially developed for melting and freezing of ice and subsequently its applications were extended to various metallurgical processes, oxygen diffusion problems and gas dynamics etc. Recently this type of formulation has successfully been applied to solve geophysical, geological and some

[1]Department of Mathematics, University of Roorkee, Roorkee–247676, India
[2]Department of Earth Sciences, University of Roorkee, Roorkee–247676, India

1217-8977/97/$ 5.00

tectono-thermal problems like lithosphere thinning (Spohn and Schubert 1982), metamorphic processes (England 1978), magmatic intrusion (Manglik and Singh 1996) and solidification of magma chambers (Huppert 1990). These physical systems are nonlinear. The position of the boundary in these systems are not known in advance and has to be determined as the part of solution (Crank 1984 and Hill 1987).

Magma chambers are reservoirs of molten rocks which solidify due to cooling. This cooling process have been studied by various workers using different analytical and numerical approaches (Spera et al. 1984, Huppert and Worster 1985, Canright and Davis 1989, Langer 1987, Ananth and Gill 1988, 1989, Glicksman et al. 1986, Loper 1987, Worster et al. 1993, Kerr et al. 1990). Some of these approaches are similarity solutions (Ivastnov 1947, Langer 1980), numerical simulations (Brown 1988) and by fluid mechanical effects (Huppert 1990, Huppert and Worster 1992, Worster 1983). In the third approach an important constituent of the analysis is convection. It is important to know the details of convection occurring in the domain i.e. whether the convection is layered or not or convection is occurring in whole or partial area and many other things. The vigor of convection however is measured by Rayleigh number which is inversely proportional to viscosity. Emphasis is mostly given to the efficient self-regulating mechanism linking to heat loss and internal temperature of earth. This has helped to explore "Parameterization" of convective heat transfer (Davies 1980, Schubert 1979, Turcoutte 1980). Whole mantle convection has been assumed in parameterized model of thermal history of earth. For mathematical treatment all the complexities of the flow pattern and the temperature field of convection cell are ignored and a simple relation between Rayleigh number Ra, and Nusselt number Nu (a measure of global efficiency of convective heat transport) is generally applied and taken as

$$Nu\,\alpha\,Ra^{\beta}\,.$$

In the classical models of constant viscosity convection, β in Rayleigh-Nusselt number relationship is chosen as 0.20–0.33. Boundary layer theory determines β as 1/3. It is where the variable viscosity models vary significantly. The range 0.0–0.33 of β have been extensively reviewed by Christensen (1984, 1985). He concluded that these values should lie between 0.05–0.10, however, it may go up to maximum of 0.15 for variable viscosity convection models of thermal evolution of earth. We have used above modified range to study the effect of various physical parameters on solidification process and phase boundary movement.

Closed form solution of MBP are not available in general, though efforts have been made to investigate and simulate these problems using various approximations of parameters. Very few analytical methods are available in literature like heat balance method, isotherm migration method and asymptotic solution through perturbation (Crank 1984 and Hill 1987). Among the numerical methods, finite difference method (Crank 1984), finite element (Hsiao and Chung 1984) and Fourier method (Glicko and Rovensky 1985) are mostly used. In this paper we have used Runge-Kutta-Fehlberg fifth order method to solve the problem for coupled set of differential equations derived for the physical system.

2. Mathematical model

One dimensional model (Fig. 1) has been chosen to investigate the effect of different physical parameters and movement of phase boundary with time. A semi-infinite layer of liquid (pure melt) which is initially at a uniform temperature T_o is suddenly subjected to a maintained temperature T_A ($z = 0$) which is below the solidification temperature T_s i.e. $T_A < T_s < T_o$, where T_s is solidification temperature. The upper boundary is taken at $z = 0$ and the base is at $z = H$.

The lower boundary in the model is assumed as insulated, the temperature of melt T_m is kept uniform by convective motion of melt and decreases with time owing to the convective heat transfer to the solidifying roof. Though the temperature profile in the solid phase should be evaluated from the solution of diffusion equation, it is assumed as linear for simplicity (Eq. 6). $z = s(t)$ denotes the moving boundary which moves with time. Two conditions must be satisfied at the phase boundary i.e. at $z = s(t)$. Firstly, the temperature at any time at the interface is given by solidification temperature which can be expressed mathematically as

$$T(z = s(t), t) = T_s \tag{1a}$$

with initial conditions

$$T_m = T_0\,, \quad s = 0\,, \quad t = 0\,. \tag{1b}$$

The second condition is due to conservation of heat energy which gives

$$k_s \left.\frac{\partial T}{\partial z}\right|_{z=s(t)} = F_T + [c_m(T_m - T_s) + L]\frac{ds}{dt}\,, \tag{2}$$

where F_T the thermal flux from vigorously convecting fluid to the phase boundary, is governed by the equation

$$F_T = \gamma k_m \left(\frac{\alpha g}{\kappa_m \eta}\right)^{\beta} (\Delta T)^{1+\beta}(H - s)^{3\beta-1}\,. \tag{3}$$

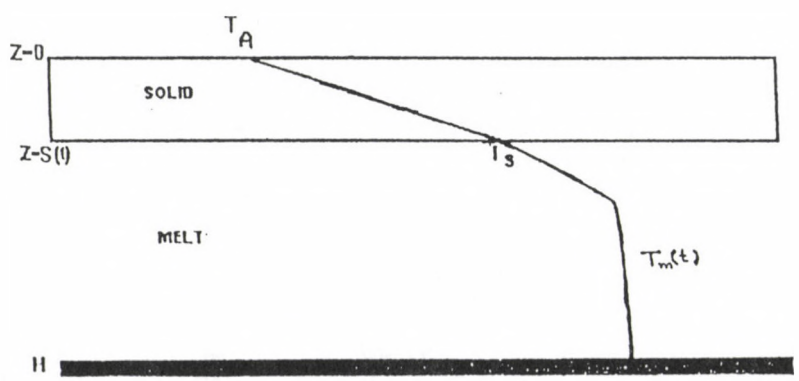


Fig. 1. Diagram of a layer of fluid of depth H which is being cooled and solidified from above. The upper boundary $z = 0$ is maintained at a constant temperature T_A which is lower than the solidification temperature of the melt T_s. The lower boundary is assumed as insulated

In the boundary layer theory, β is taken as 1/3. Then F_T is given by

$$F_T = \gamma k_m \left(\frac{\alpha g}{\kappa_m \eta}\right)^{1/3} (\Delta T)^{4/3} . \tag{4}$$

We specify that Eq. (3) has a very important term $(H - s)^{3\beta-1}$ which is not present in Eq. (4). Equation (3) is much more general equation. Therefore Eq. (4) may be taken as special case of Eq. (3). This term has been incorporated by us using the Christensen theory (1984, 1985) to study the phase boundary movement. We use Eq. (3) in the following analysis and all the investigations are done for the acceptable range of beta. Here κ_s and κ_m are thermal conductivities of solid and melt phases respectively, η the viscosity, g the acceleration due to gravity, γ the empirical constant whose value is 0.14 and α_m is the coefficient of thermal expression. c_m the specific heat per unit volume of the melt, T_m the mean temperature of the melt and L is the latent heat as melt transforms into unit volume of solid.

Due to convection in the melt temperature decreases which can be given as

$$F_T = -c_m(H - s)\frac{dT_m}{dt} . \tag{5}$$

The linear temperature profile in the solid phase is taken as

$$\left.\frac{\partial T}{\partial z}\right|_{z=s(t)} = \frac{(T_s - T_A)}{s} . \tag{6}$$

After introducing dimensionless variables as

$$u = \frac{T_m - T_s}{T_0 - T_s} \qquad v = \frac{s}{H} \qquad \tau = \frac{\kappa_m t}{H^2} \tag{7}$$

Eqs (7), (2) and (3) together give

$$\frac{1}{v} = N'(1 - v)^{3\beta-1} Q u^{1+\beta} + (Qu + \overline{S})\frac{dv}{d\tau} . \tag{8}$$

Here u represents the remaining superheat of the melt and v represents the fraction of the solid layer. N' is modified Nusselt number and $\overline{S}$ is modified Stefan number. Time t is non-dimensionalised by H^2/κ_m (conductive flux)

$$N' = \gamma \left(\frac{\alpha g}{\kappa \eta}\right)^{\beta} (T_0 - T_s)^{\beta} H^{3\beta} \tag{9a}$$

$$Q = \frac{T_0 - T_s}{T_s - T_A} \tag{9b}$$

$$\overline{S} = \frac{\kappa_m}{\kappa_s} \frac{L}{c_m(T_s - T_A)} . \tag{9c}$$

Substituting values of F_T from Eq. (3) into Eq. (6) we get

$$\frac{du}{d\tau} = -u^{1+\beta}(1-v)^{3\beta-2}\,. \tag{10}$$

The initial conditions transform to

$$u = 1\,, \quad v = 0\,, \quad \tau = 0\,. \tag{11}$$

The system of Eqs (8), (10) and (11) form a nonlinear coupled set of first order differential equations which is solved by using Runge-Kutta-Fehlberg fifth order method to investigate the effect of various physical parameters on the phase boundary.

3. Computational details

In our analysis of solidification of magma chambers the exponent beta is the key parameter which affects the movement of phase boundary. The movement of phase boundary has been discussed for its various values with distinct physical parameters. Beta takes the values from 0.05–0.10 up to a maximum of 0.15. Out of this range of permissible values of beta, we have chosen three particular values 0.05, 0.10 and 0.15 so as to fix the data points suitably. The phase boundary movement has been shown in the analysis with the values of some other parameters i.e. Stefan number, initial heat flux and Nusselt number. Among the two important aspects of modelling i.e.

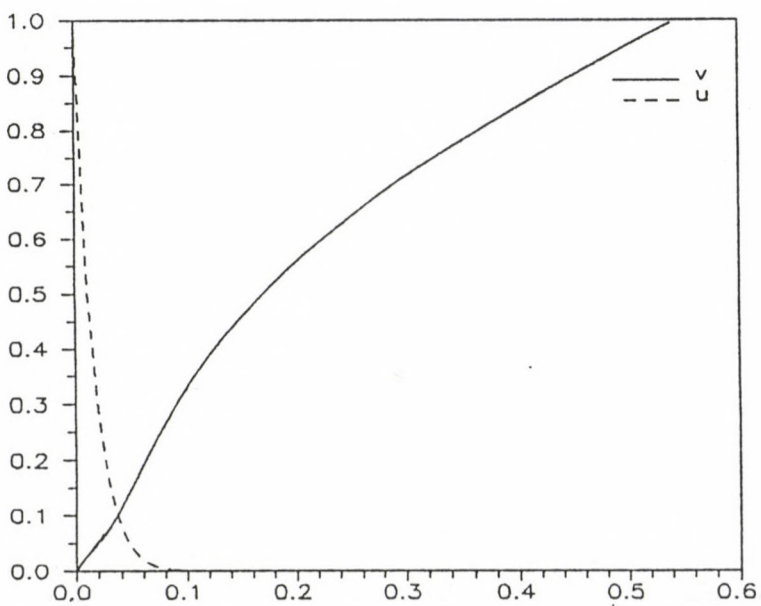


Fig. 2. Schematic diagram of the solidification of a layer of liquid using dimensionless parameters as $Q, \overline{S}$ and N. The dimensionless temperature of the melt u and the depth of solid at the roof v as a function of time t are shown here. β is taken as 0.05

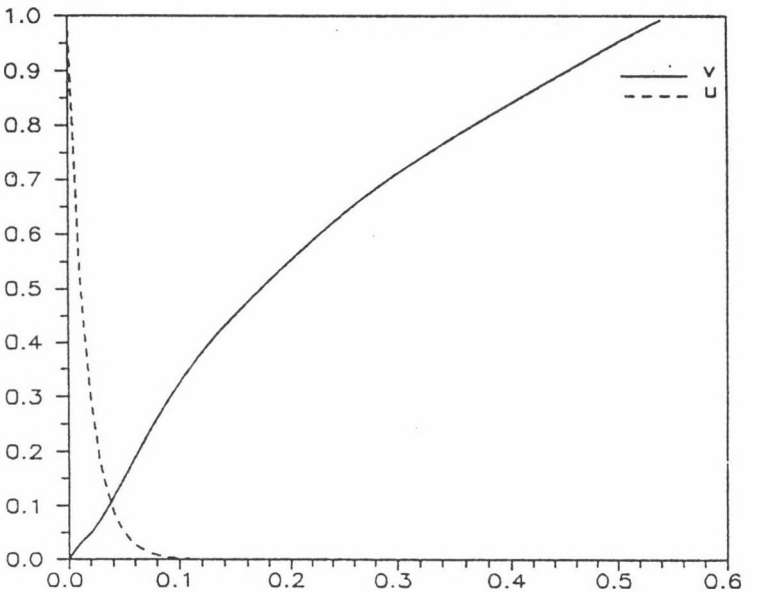


Fig. 3. Diagram of the solidification of a layer of liquid using dimensionless parameters as $Q, \overline{S}$ and N. Temperature of the melt u and the depth of solid at the roof v as a function of time t are shown β is taken as 0.10

viscosity ratio between top and bottom is constant or whether the top boundary is fixed, second on is chosen because it gives appropriate parameterization for cooling a planetary body and has been confirmed in the studies based on variable viscosity convection models (Christensen 1984 and 1985).

Three other parameters which have been analyzed are N' (the modified Nusselt number, a measure of global efficiency of convective heat transport), Q (initial superheat) and $\overline{S}$ (the modified Stefan number). Though the values of these parameters N', $\overline{S}$ and Q can be calculated from the formulae given by Eqs (9a), (9c) and (9b), we take these values as 60, 1.0 and 1.0 respectively for studying the movement of phase boundary. In other cases comparative studies of different values of $\overline{S}$, N' and Q are made and their effect on the solidification front is analyzed.

4. Results and discussion

Coupled set of nonlinear differential Eqs (8), (10), (11) are solved by Runge-Kutta-Fehlberg fifth order scheme. The scheme uses fourth and fifth order method to move from one value to another value. Two consecutive values of functions are calculated, one by fourth order formula and another by fifth order formula and the values are compared with error in the step. The error is estimated by the difference in two estimates i.e. estimate by fourth order and estimate by fifth order method. After assuming an initial desired accuracy level say Δ_0, the step sizes are modified by a formula. We have taken Δ_0 as 0.00001. Step sizes are modified at each step and then values for next step are calculated. We have described the effects of various

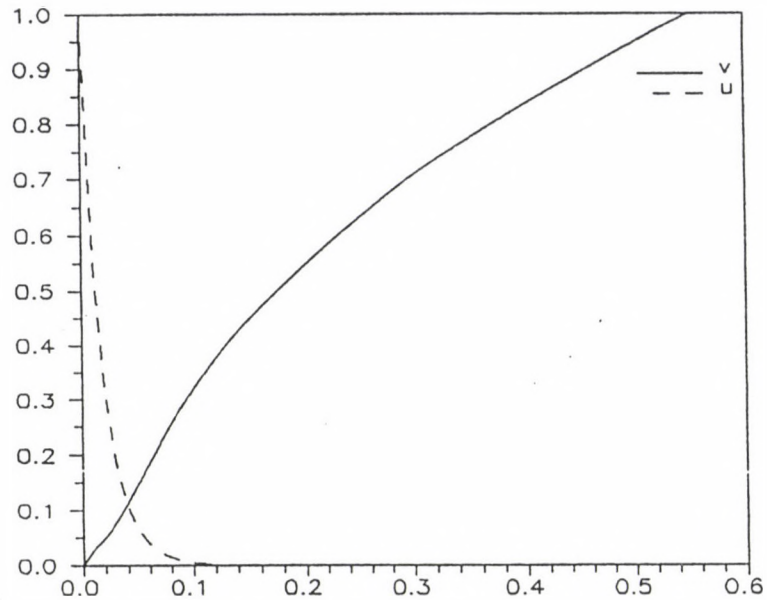


Fig. 4. Schematic diagram of the solidification of a layer of liquid using dimensionless parameters as $Q, \overline{S}$ and N. The temperature of the melt u and the depth of solid at the roof v as a function of time t are shown here. β is taken as 0.15

factors in four cases and movement of phase boundary are estimated and compared with standard solutions.

Case 1

Here different values of β are chosen and their effect on the moving front is shown (Figs 2, 3, 4). In this case initial heat flow is chosen as 1.0 and value of Stefan number is taken as 1.0. The value of the modified Nusselt number is taken to be 60. The stages of thermal evolution shows that superheat u decays very fast to a small value which facilitates the solidification process and allows the system to solidify quickly. The superheat decays in time 0.12 (nondimensional) and the total time taken for solidification is around 0.5399 units of time. When we increase the value of beta from 0.05 to 0.10 (Fig. 3), we find that the time taken for vanishing the superheat increases and it completely vanishes at 0.14 unit of time (values are correct to four decimal places). At this time the value of v is 0.4292 (units). Phase boundary gains the stabilized position at 0.5430 units of time. When beta is further increased to 0.15 (Fig. 4) the superheat vanishes at 0.17 units of time and value of v at this instant is 0.4987. The boundary takes its final position i.e. complete solidification state, at 0.5499 units of time. It can therefore be inferred that solidification time increases as the exponent beta increases and this gives more smooth pattern to the solidification process. Since the superheat u does not increase much, the solidification is fast and convection still dominates in the whole process.

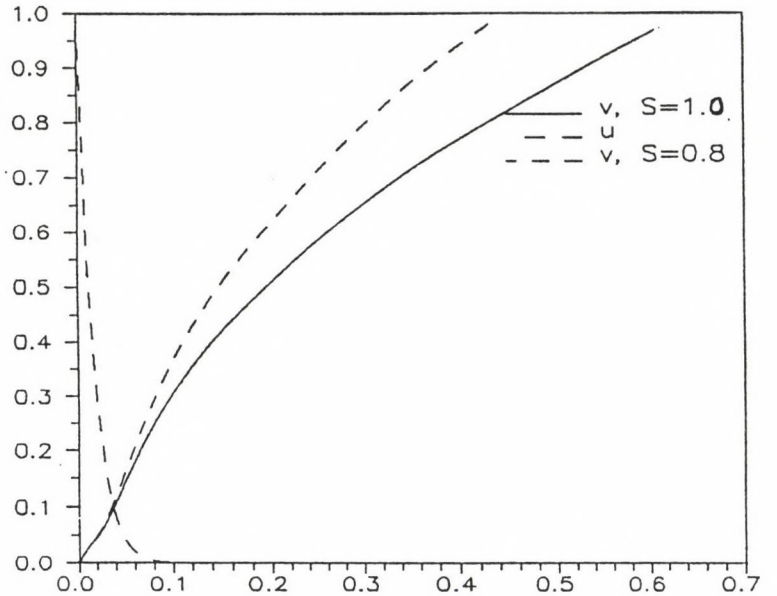


Fig. 5. Comparative values for the two values of Stefan number are shown here. The values in this case are taken as $\beta = 0.05$, $N = 60$ and $S = 1.0$ (—). In the second case the values are $\beta = 0.05$, $N = 60$ and $S = 0.8$ (– –)

Case 2

We describe the effect of parameter Nusselt number keeping other parameters of the model fixed. Exponent beta and $\overline{S}$ the Stefan number are taken as 0.1 and 1.0 respectively in all three cases under consideration. As N' is increased from 60 to 90 it is found that superheat u decays more quickly as compared to the case when N' is take as 60. The solidification process becomes a little bit faster but total solidification time does not increase much. Similar effect on the process of solidification is found if the value of N' is increased from 90 to 120. We find that superheat again decreases quickly and is even less as compared to $N' = 90$, but total time for complete solidification does not change. Therefore it is concluded that as the convection intensity increases, the solidification time decreases relative to the time scale for thermal diffusion. This range of Nusselt number should not be violated much because other values may not be favorable to majority of geological situations. In this case initial superheat Q was taken as 0.1.

Case 3

In this case effect of varying the initial superheat Q for its three values chosen as 1.0, 0.10 and 0.05 is discussed. The fixed parameters beta, Nusselt number N' and Stefan number $\overline{S}$ are taken as 0.1, 60 and 1.0 respectively. Here we find that in all the three cases superheat u again decays fast. When Q is taken as 1.0 it decays in 0.14 units of time. When $Q = 0.1$, superheat vanishes in 0.12 units of time and

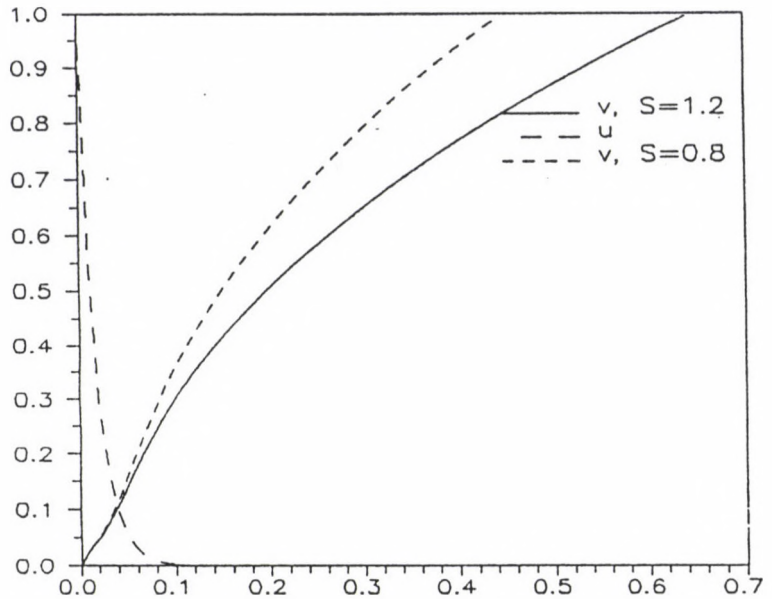


Fig. 6. Comparative values for the two values of Stefan number are shown here. The values in this case are taken as $\beta = 0.05$, $N = 60$ and $S = 1.2$ (—). In the second case the values are $\beta = 0.05$, $N = 60$ and $S = 0.8$ (– –)

if $Q = 0.05$ it takes 0.11 unit of time to decay. Total time taken for the complete solidification is 0.53 when Q is taken is 1.0. However, total time for the solidification reduces to 0.50 and 0.49 when Q is reduced from 1.0 to 0.05. We again see that the superheat u decays rapidly in these two cases also. Now although the Rayleigh number based upon the value of superheat is still very large, the convective heat flux are always small in comparison with the heat flux conducted through the roof and solidification pattern remain same for all the cases approximately. So convection still plays a role in reducing the time for complete solidification.

Case 4

In this case the effect of variations in Stefan number is analyzed. Three values of this are taken as 1.0, 0.8 and 1.2 to measure the effect on the phase boundary. In this case we have fixed beta as 0.05 and Nusselt number and heat fluxes are taken as 60 and 1.0 respectively. When Stefan number is taken as 0.8, solidification of whole region completes in 0.43 units of time. When it is chosen as 1.0 the total time for solidification is increased (0.54 units). Here also the superheat vanishes quickly (Fig. 5). If we further increase the Stefan number i.e. up to 1.2, it is found that the time taken for phase boundary to gain its established position increases (Fig. 6). This means that increase in the Stefan number increases the solidification time but the superheat is still small to exemplify that convection is dominant in the whole process.

5. Conclusion

The analysis has been presented for many aspects of the solidification processes in a domain. This can be applied as an approximate model for explaining the evolution of lava lakes, which are formed when lava from an eruption flows into a depression on earth surface. The lava could be a few meters deep and cooled by contact with atmosphere at the upper surface. This model takes into account three important aspects. The first is that any superheat is removed quickly and another is that the vigorous convective motions will take place in the region. All these conclusions and considerations are in agreement with the field data obtained from lava lakes and remains valid even if more factors are added like conductivity and compositional differences. However, the present analysis is valid only when whole region convection is taken into account. If convection is layered then solidification time is to be verified. What governs the heat flow through surrounding region? Is convection layered or not? These are the questions to be studied in future. This analysis can further be useful for studying the models with different boundary conditions. Consideration of mushy zone can also be incorporated in the study in future.

Acknowledgements

The authors would like to thank Dr R N Singh (CMMACS, BANGLORE) for his guidance and support during this study. One of the author would also like to thank U.G.C. for their financial support.

References

Ananth R, Gill W N 1988: *Chem. Engg. Comm.*, 68, 1–44.
Ananth R, Gill W N 1989: *Jour. Fluid. Mech.*, 208, 575–593.
Brown R A 1988: *AICHEJ*, 34, 881–911.
Canright D, Davis S H 1989: *Metal. Tran.*, A-20, 225–235.
Christensen U R 1984: *Phys. Earth Planet. Int.*, 35, 264–282.
Christensen U R 1985: *J. Geoph. Res.*, 90, B–4, 2995–3007.
Crank J 1984: Free and moving boundary problems. Clarendon Press, Oxford
Davies G F 1980: *J. Geoph. Res.*, 85, 2517–2530.
England P C 1978: *Tectonophysics*, 46, 21–40.
Glicksman M E, Coriell S R, McFadden G B 1986: *Ann. Rev. Fluid. Mech.*, 18, 307–335.
Glicko A O, Rovensky O N 1985: *Izv. Akad. Nauk USSR*, 21, 416–419.
Hill J M 1987: One dimensional Stefan problems. Longmans
Hsiao J S, Chung B T F 1984: ASME–84, HT–2.
Huppert H E 1990: *Jour. Fluid Mech.*, 212, 209–240.
Huppert H E, Worster M G 1985: *Nature*, 314, 703–707.
Huppert H E, Worster M G 1992: In: Chaotic processes in geological sciences. D A Yuen ed., Springer Verlag, New York, 141–173.
Ivastnov G P 1947: *Dokl. Akad. Nauk USSR*, 58, 567–569.
Kerr R C, Woods A W, Worster M G, Huppert H E 1990: *Jour. Fluid. Mech.*, 216, 323–342.

Langer J S 1980: *Rer. Mod. Phys.*, 52, 1–28.
Langer J S 1987: Chance and matter. ASI, North Holland, 629–712.
Loper D E 1987: Structure and dynamics of partially solidified systems. Matinus, Nijhoff
Manglik A, Singh R N 1996: *J. Geoph. Res.*, 100, 18031–18043.
Schubert G 1979: ICARUS, 38, 192–211.
Spohn T, Schubert G 1982: *J. Geoph. Res.*, 87, 4682–4696.
Spera F J, Yuen D A, Kemp D V 1984: *Nature*, 310, 764–767.
Turcoutte D L 1980: *Earth Planet. Sci. Lett.*, 48, 53–58.
Worster M G 1983: Convective flow problems in the geological fluid dynamics. PhD Thesis, University of Cambridge, Cambridge
Worster M G, Huppert H E, Sparks R S J 1993: *J. Geoph. Res.*, 98, 15891–15901.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 15–25 (1997)

A STABLE ITERATIVE SCHEME FOR INVERSION OF DIRECT CURRENT RESISTIVITY SOUNDING DATA

SRI NIWAS[1] and AMIT MEHROTRA[1]

[Manuscript received October 2, 1996]

This paper deals with a scheme for inversion of d.c. resistivity data wherein ill-conditioning of the system matrix is reduced by making initial guess in terms of reflection coefficients instead of resistivities and ridge regression inversion is performed by composing the system matrix wherein differential weights (w_r and w_t) are provided to the set of columns pertaining to reflection coefficient and layer thickness respectively. Four synthetic data sets are chosen to test the inversion scheme and the solutions are compared with those obtained by classical scheme. It is found that the former solution is better than the latter in terms of convergence. One field data set is inverted to obtain roughly same solution while using three different initial guesses.

Keywords: inversion; resistivity sounding

1. Introduction

Direct current resistivity method is extensively used for groundwater exploration, environmental studies and engineering geophysics. The success of these programs largely depend on reliable inversion of measured data. Inversion is relatively difficult and inverse solutions so obtained suffer from non-uniqueness due to various mathematical, observational and computational limitations (Backus-Gilbert 1967, 1968, 1970, Berdichevsky and Zhdanov 1984). Inverse theory for linear problems is well developed. However, to use it for non-linear geoelectrical problems quasi-linearization in the vicinity of an initial guess becomes essential. The choice of initial guess of the model parameters imports subjectivity affecting quality of inverse solution. Ill-conditioning of the linear system when some model parameters are linearly dependent results in the existence of extremely small eigenvalues. These vanishing eigenvalues are the cause of the unstable inverse solution when data contains error. The quantitative measure of ill-conditioning of the system is the condition number (η_j) given by

$$\eta_j = \frac{\mu_1}{\mu_j},$$

where (μ_1) is the first eigenvalue and (μ_j) is the jth eigenvalue of the system matrix. Twomey (1977) has demonstrated by taking hypothetical case that direct inversion and least square inversion are intrinsically unstable in the sense that small change in data produce large variation in solution because of ill-conditioned system matrix.

[1]Department of Earth Sciences, University of Roorkee, Roorkee–247667, India

1217-8977/97/$ 5.00

The only procedures which can be expected to improve the situation are those which in some way increase the magnitude of the eigenvalues. This is achieved by imposing additional conditions or criterion (not derived from data) which are essentially arbitrary. Ridge-regression and damped least square are such procedures to improve stability of the solution (Marquardt 1963, 1970, Tikhonov and Arsenin 1977, Akí and Richard 1980), however, at the cost of increased error and more number of iterations (Inman 1975). In ridge regression inversion all the eigenvalues are suitably increased (by Marquardt parameter p) to enhance the visibility of small eigenvalues so that stability is achieved. However, the choice of p and its rate of decrement may be chosen judiciously, otherwise ridge-regression solution may also diverge (Aki and Richard 1980).

We are investigating here a sound basis for choosing initial guess to reduce the subjectivity and the ill-conditioning of the system matrix so that a stable solution is obtained without increasing error and number of iterations. The subjectivity in choosing initial guess is minimized by seeking solution in terms of reflection coefficients (instead of resistivities). In comparison to layer resistivities which has large range of variation, reflection coefficient ranges between -1.0 to $+1.0$. The condition number is reduced by composing system matrix column-wise with partial derivative of apparent resistivity function with respect to reflection coefficients (r_j) and layer thicknesses (t_j) and by providing differential weight (w_r and w_t) to these two sets of column. The inverse solution is obtained using ridge regression inversion.

2. The scheme

Apparent resistivity expressions for various electrode configurations are well known. For example Schlumberger expression for apparent resistivity can be written as (Flathe 1955, Koefoed 1979)

$$\varrho_a(s) = \varrho_1 s^2 \int_0^\infty \Theta_n(\lambda) J_1(\lambda s) \lambda \, d\lambda \,, \tag{1}$$

where $J_1(\lambda s)$ is the Bessel function of first order of the first kind, s is the half-current electrode separation, λ is the integration variable and $\Theta_n(u)$ is the Slichter Kernel function of layer thicknesses t_j $(j = 1, 2, \ldots n)$ and reflection coefficients r_j $(j = 1, 2, \ldots n+1)$, given by

$$\Theta_n(u) = \frac{N_n(u)}{D_n(u)} \,. \tag{2}$$

n is the number of layers resting over half-space,

$$N_{j+1}(u) = N_j(u) + r_j u_j^2 N_j(u^{-1})$$

and

$$D_{j+1}(u) = D_j(u) - r_j u_j^2 D_j(u^{-1})$$

$$N_{j+1}(u^{-1}) = N_j(u^{-1}) + r_j u_j^{-2} N_j(u)$$

$$D_{j+1}(u^{-1}) = D_j(u^{-1}) - r_j u_j^{-2} D_j(u)$$

$$N_1 = D_1 = 1, \qquad u_j = e^{-\lambda h_j}, \qquad r_j = \frac{\varrho_{j+1} - \varrho_j}{\varrho_{j+1} + \varrho_j}.$$

For using linear inversion scheme Eq. (1) is quasi-linearised in the vicinity of initial guess

$$m^0 = \{r_1^0, r_2^0, \ldots r_n^0; t_1^0, t_2^0, \ldots t_n^0\}^T$$

of true model parameters

$$m = \{r_1, r_2, \ldots r_n; t_1, t_2, \ldots t_n\}^T.$$

T being transpose operation.

The quasi-linearized version of Eq. (1) for ith half-current electrode separation s_i is,

$$\varrho_{a0}(s_i) = \varrho_{ac}(s_i)|_{m=m^0} + \sum_{j=1}^{M}(m_j - m_j^0)\frac{\delta \varrho_{ac}(s_i)}{\delta m_j}\bigg|_{m=m^0} \tag{3}$$

$$\begin{aligned} i &= 1, 2 \ldots N \\ j &= 1, 2, \ldots M,\ M = 2n \end{aligned}$$

where $\varrho_{a0}(s_i)$ and $\varrho_{ac}(s_i)$ are the observed and computed apparent resistivity values. Equation (3) may be posed as matrix equation

$$\underset{\text{NX1}}{\delta d} = \underset{\text{NXM}}{G}\ \underset{\text{MX1}}{\delta m} \tag{4}$$

with

$$\begin{aligned} \delta d_i &= \varrho_{a0}(s_i) - \varrho_{ac}(s_i), \quad i = 1, 2, \ldots, M \\ \delta m_j &= m_j - m_j^0 \qquad\qquad\quad j = 1, 2, \ldots, M \end{aligned}$$

$$G_{ij} = \frac{\partial \varrho_{ac}(s_i)}{\partial m_j}\bigg|_{m=m^0}.$$

The coefficient matrix is composed with partial derivatives

$$\frac{\partial \varrho_a(s_i)}{\partial m_j} = \varrho_1 s^2 \int_0^\infty \frac{\partial \Theta_n(\lambda)}{\partial m_j} J_1(\lambda s)\lambda\, d\lambda \tag{5}$$

which may be obtained through recurrence relation

$$\frac{\partial \Theta_n(u)}{\partial m_j} = \frac{D_n(u)\frac{\partial N_n(u)}{\partial m_j} - N_n(u)\frac{\partial D_n(u)}{\partial m_j}}{D_n^2(u)}.$$

The two sets of column of G, due to reflection coefficients and layer thicknesses may be provided with different weights (w_r and w_t)

$$\begin{aligned} G &= (G_1, G_2) \\ G_1 &= \frac{\partial \varrho_a(s_i)}{\partial r_j} w_r \\ G+2 &= \frac{\partial \varrho_a(s_i)}{\partial t_j} w_t \end{aligned}$$

to keep η within reasonable limit.

By composing Eq. (4) in this manner and using ridge regression inversion the estimated δm may be obtained as

$$\delta \hat{m} = (G^T G + p \cdot I)^{-1} G^T \delta d \,. \tag{6}$$

Initially the Marquardt parameter, p is kept large (say 1) and then decreasing it suitably in each advancing iteration till

$$\frac{1}{N \varrho_1} \delta d^T \delta d \leq 6 \,,$$

preassigned convergence level is achieved. The conventional ridge regression solution may be obtained by substituting w_r and w_t equal to 1 in Eq. (6) for comparison. The final model is obtained as

$$m = m^0 + \delta \hat{m}$$

and hence r_j and t_j.

The solution in layer resistivities are obtained once ϱ_1 is obtained through exterior means. Having knowledge of ϱ_1 is not very difficult for experienced electrical geophysicists.

3. Numerical examples

Variety of synthetic data computed for differing models have been inverted to demonstrate the effectiveness of the suggested improvement in the stable iterative scheme for better solution of the resistivity inverse problem. Included here are the following four models as representative ones.

Model 1: $\varrho_1 = 100$ ohm m, $t_1 = 10$ m, $\varrho_2 = 10$ ohm m, $t_2 = 20$ m, $\varrho_3 = 100$ ohm m
Model 2: $\varrho_1 = 100$ ohm m, $t_1 = 10$ m, $\varrho_2 = 200$ ohm m, $t_2 = 20$ m, $\varrho_3 = 100$ ohm m
Model 3: $\varrho_1 = 10$ ohm m, $t_1 = 10$ m, $\varrho_2 = 50$ ohm m, $t_2 = 50$ m, $\varrho_3 = 100$ ohm m
Model 4: $\varrho_1 = 500$ ohm m, $t_1 = 5$ m, $\varrho_2 = 500$ ohm m, $t_2 = 12.5$ m, $\varrho_3 = 1250$ ohm m, $t_3 = 50$ m, $\varrho_4 = 50$ ohm m.

Synthetic Schlumberger data are presented in Fig. 1 for all the four models. The system matrix G is composed by providing for solution (A) equal weights ($w_r = w_t = 1$) and for solution (B) different weights w_r and w_t. These weights are chosen on the basis of trial and error taking clue from the ratios of minimum to

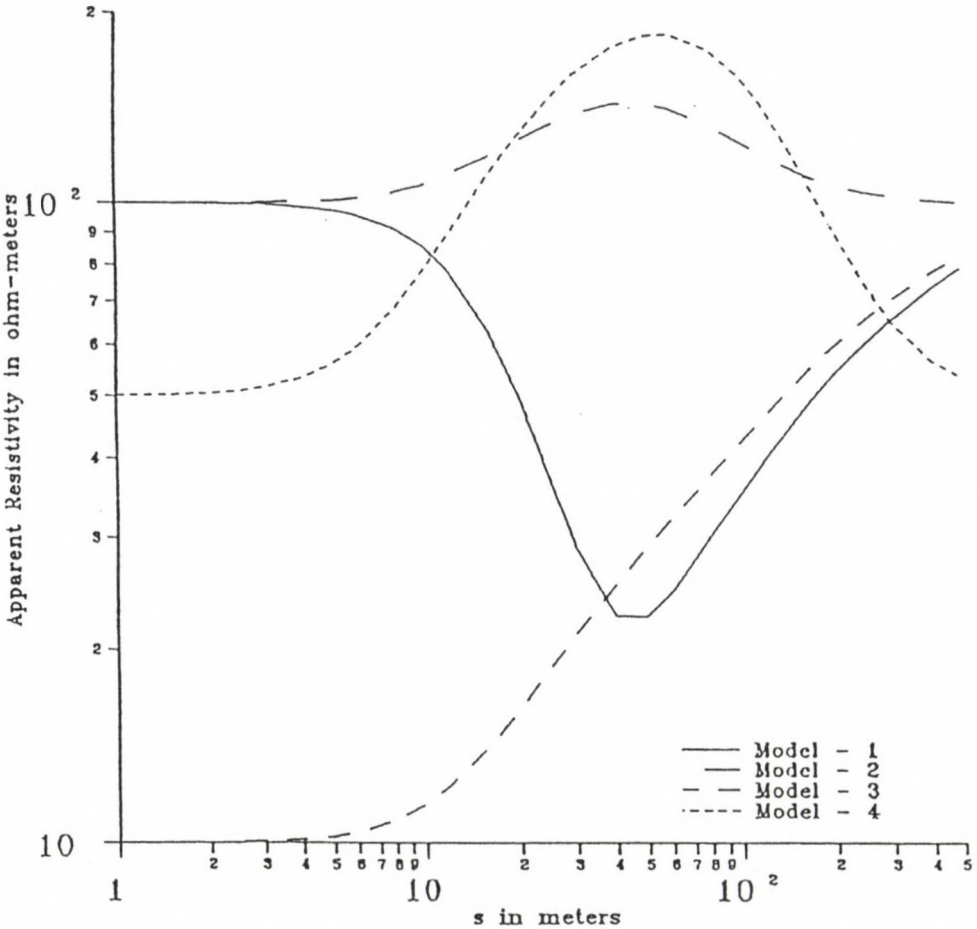


Fig. 1. Schlumberger apparent resistivity curve over Models 1, 2, 3 and 4

maximum eigenvalues which have eigenvectors which are chiefly pointing towards r (as w_r) and t (as w_t) range of variation of eigenvalue. Since the data generated are noisefree a very small average least square error ($\varepsilon < 0.005$) are chosen for achieving convergence.

For obtaining inverse solution of Eq. (4) one needs characterization of operational space of matrix operator G which includes parts of or full data space and parameter space. We can obtain an illuminating view of the linear system using Singular Value Decomposition (Lanczos 1961). The inverse problem are presented as overdetermined linear system. Number of expected eigenvalues is equal to the number of sought for parameters. These eigenvalues have associated parameter eigenvectors which essentially represent a combination of the original parameters that may be uniquely determined. Using Singular Value Decomposition of matrix G it has been

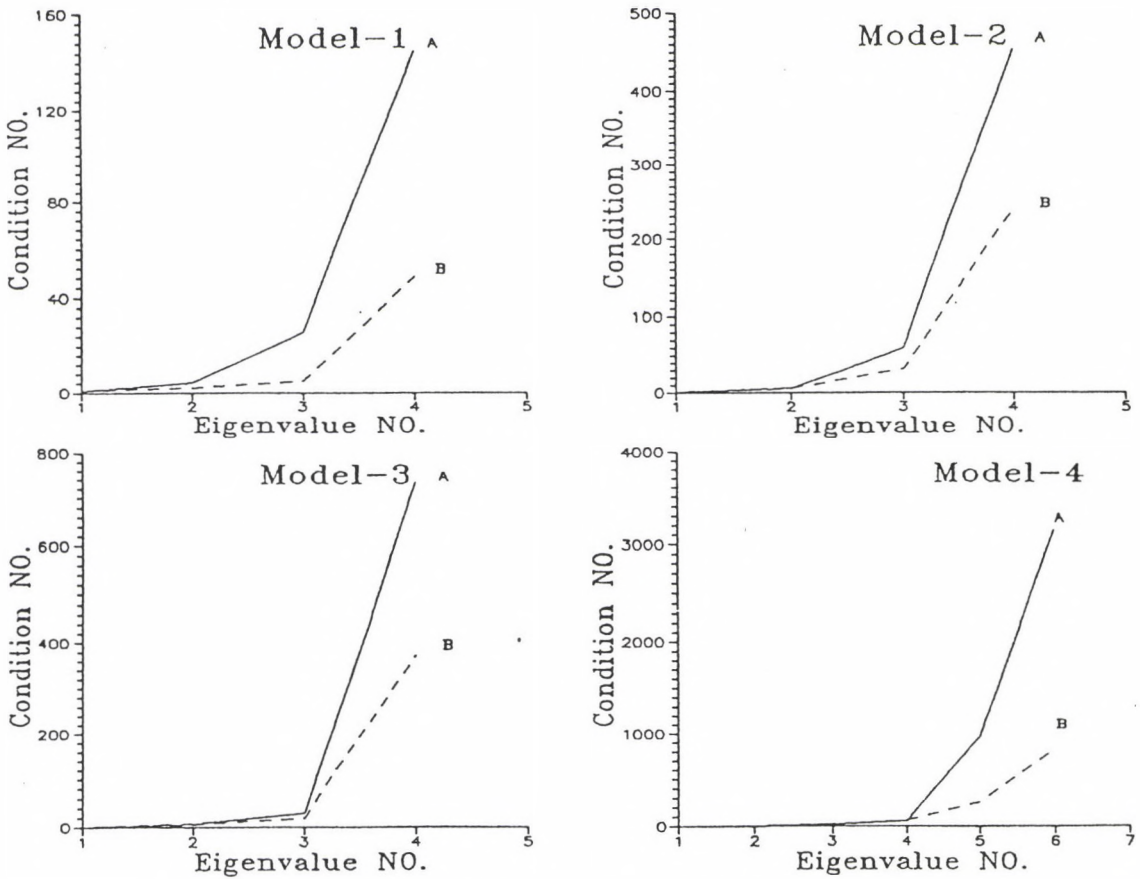
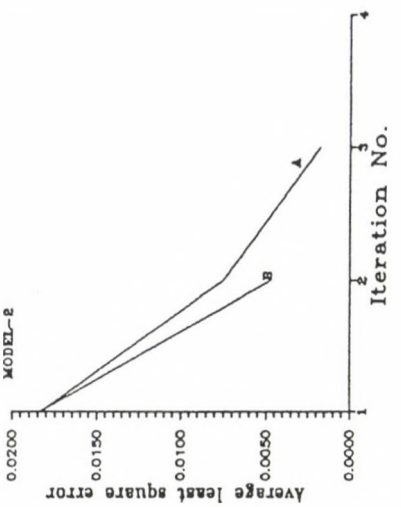
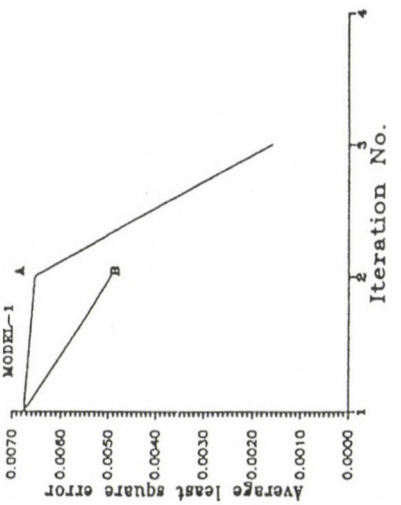
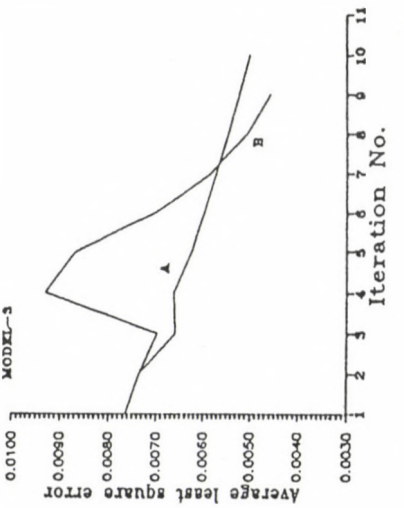


Fig. 2. Condition number plot for case A and case B for Models 1, 2, 3 and 4







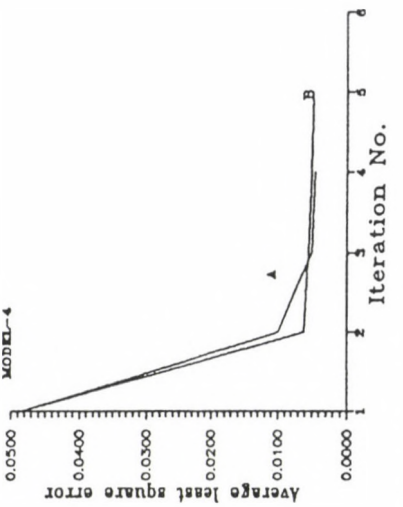


Fig. 3. Convergence of Solutions A and B for Models 1, 2, 3 and 4

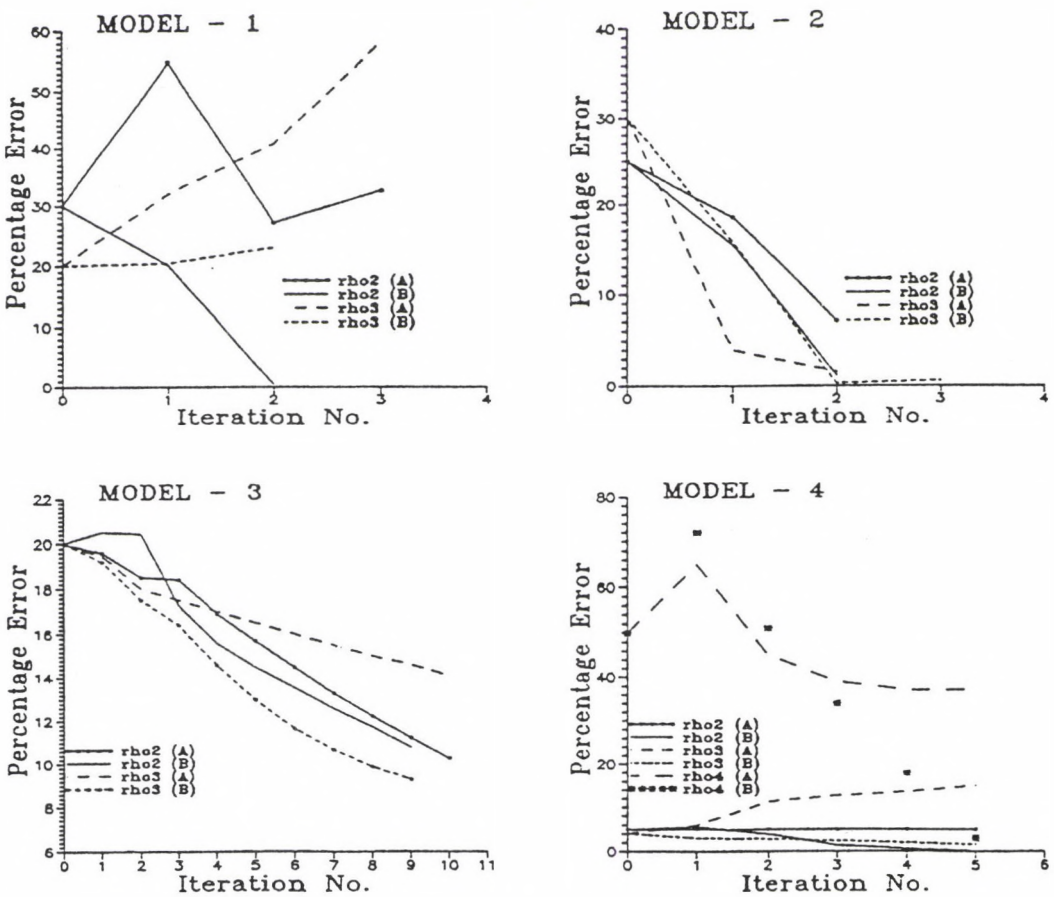


Fig. 4. Convergence of parameter ϱ_2, ϱ_3 for Model 1, 2, 3 and of ϱ_2, ϱ_3, ϱ_4 for Model 4 for both the Solutions A and B

observed that for case (A) the large eigenvalues ($\mu_1 = 12.0955$, $\mu_2 = 2.7766$ for Model 1; $\mu_1 = 8.5523$, $\mu_2 = 1.4348$ for Model 2; $\mu_1 = 37.53$, $\mu_2 = 5.0068$ for Model 3 and $\mu_1 = 34.9005$, $\mu_3 = 1.4666$ for Model 4) have associated eigenvector which points chiefly in the direction of reflection coefficients (r_1 and r_2 for Model 1, 2, 3 and r_1, r_2, r_3 for Model 4) whereas smaller eigenvalues ($\mu_3 = 0.4664, \mu_4 = 0.0834$ for Model 1, $\mu_3 = 0.1454, \mu_4 = 0.0189$ for Model 2, $\mu_3 = 1.2559, \mu_4 = 0.0511$ for Model 3 and $\mu_5 = 0.0362, \mu_6 = 0.0111$ for Model 4) have associated eigenvectors pointing strongly in the direction of layer thicknesses. The only exception is $\mu = 4.1656$ for Model 4 which is strongly correlated with t_1. This shows that the reflection coefficients may be resolved more accurately and layer thicknesses may be resolved less accurately. However, the range of variation of eigenvalues are very wide which results in very large condition number, which affects the resolvability of even reflection coefficients if proper care is not taken. The eigenvalues and associated eigenvectors for (case B) reduced considerably the variation of eigenvalues for case B, however, the strong linear dependence of eigenvectors associated with each eigenvalue is rationalized. Thus the application of weight reduces the gradient of condition number plot (Fig. 2, Curve B) in comparison to the condition number plot (Fig. 2, Curve A) for the case when no weight is introduced. This shows the possibility of better convergence of solution.

4. Discussion and conclusion

Once the matrix G is composed ridge regression inversion is performed according to Eq. (6) starting with Marquardt parameter $p = 1$ and decreasing it by one tenth in each iteration (Sri Niwas et al. 1982). The solution obtained using classical scheme ($w_r = w_t = 1.0$) is labeled as A in Figs 3 and 4 whereas the solution obtained using scheme worked out in this paper is labeled as B. For both solutions A and B same initial guess is used. Since the power of ridge regression solution in obtaining stable solution is well documented, efforts would be mainly directed to show that solution B is better than solution A. Figure 3 present the convergence of the solutions A and B for all the four Models, 1, 2, 3 and 4. It is evident that the convergence of solution B has improved in comparison to solution A by simply providing appropriate weight to the columns of matrix G pertaining to reflection coefficients and layer thicknesses. Figure 4 shows convergence of layer resistivities (ϱ_2 and ϱ_3) for Models 1, 2, 3 and ($\varrho_2, \varrho_3, \varrho_4$) for Model 4 for solutions A and B. It is clear from this presentation that improvement in solution B is quite impressive for Model 1 and Model 4 whereas it is noticeable for Model 2 and Model 3. Model parameters obtained for both the solutions A and B after desired convergence level is achieved are listed in Table I along with true model parameters as well as the values of weight w_r and w_t used to obtain these solutions. A comparative look at these values shows that solution B is superior to solution A. A general comment can also be made that Bell type model is better resolved in comparison to ascending type model.

A field data set from Saurashtra region, Western India is inverted using scheme B. Three different initial guesses are used and convergence to roughly same solution is obtained. Relevant data and solution are given in Fig. 5.

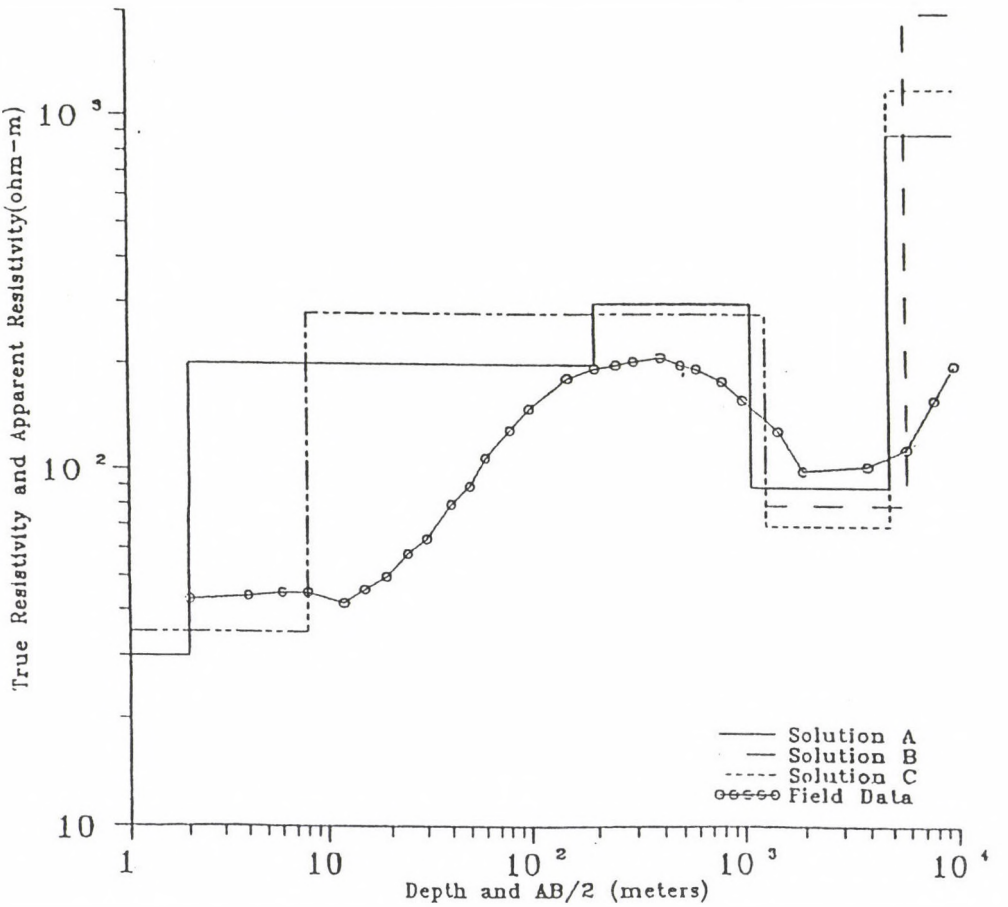


Fig. 5. Inversion result of a field data set from Saurashtra, Western India

The knowledge of correct value of ϱ_1 is essential to convert reflection coefficients obtained from inverse solution to layer resistivities. This may seem to be serious disadvantage of the method, however, experience of the interpreter may be expected to minimize the problem. Moreover, it is quite common in resistivity sounding to obtain curves with well defined asymptotes from which the resistivity of the first layer may be readily determined. On the basis of better result obtained in solution B it is recommended that the ridge regression inversion of resistivity sounding data may be performed in the manner suggested in this paper. This improvement in solution becomes possible by rationalising the distribution of eigenvalues in more equitable manner. Indirectly it reduces the effect of equivalence, however, it is not possible to completely remove the effect using d.c. resistivity data alone. This can only be achieved by jointly inverting d.c. and M.T. data. Sri Niwas and Pramod Kumar (1991) have presented the eigen analysis of same model for d.c. and M.T.

Table I. Obtained parameters from solutions A and B along with true value for model 1, 2, 3 and 4. w_r és w_t are weights

Model	Solution	ϱ_2 ohm m	ϱ_3 ohm m	ϱ_4 ohm m	t_1 m	t_2 m	t_3 m	w_r	w_t
1	True	10.0	100.0	–	10.0	20.0	–	–	–
	A	6.7	158.0	–	8.57	18.8	–	1.00	1.0
	B	9.96	123.0	–	10.2	19.7	–	0.01	0.25
2	True	200.0	100.0	–	10.0	20.0	–	–	–
	A	186.0	102.0	–	9.8	20.5	–	1.00	1.00
	B	208.0	99.4	–	10.2	19.9	–	0.05	0.10
3	True	50.0	100.0	–	10.0	50.0	–	–	–
	A	44.86	114.6	–	9.09	59.8	–	1.00	1.00
	B	45.0	109.3	–	8.1	59.6	–	0.05	0.10
4	True	500.0	125.0	50.0	5.0	12.5	50.0	–	–
	A	473.0	108.0	31.0	4.6	9.85	49.75	1.0	1.00
	B	512.0	123.0	48.5	5.0	11.55	50.0	0.02	0.08

cases and have shown that the distribution of eigenvalues are entirely different. This improved inversion scheme can also be extended to invert M.T. data in the same way as it has been applied to invert d.c. data.

5. Acknowledgements

The authors are thankful to Prof. Vinod K Gaur and Dr. P K Gupta for useful discussion and to Yesh Pal Singh and Anupama for their help in computation.

References

Aki K, Richard P G 1980: Quantitative seismology. Vol 2. Freeman and Company

Backus G, Gilbert F 1967: *Geophys. J. Roy. Astr. Soc.*, 13, 247–276.

Backus G, Gilbert F 1968: *Geophys. J. Roy. Astr. Soc.*, 16, 169–205.

Backus G, Gilbert F 1970: *Phil. Trans. Roy. Soc.*, 226, 123–192.

Berdichevsky M N, Zhdanov M S 1984: Advanced theory of deep geomagnetic sounding. Elsevier

Inman J R 1975: *Geophysics*, 40, 798–817.

Flathe H 1955: *Geophys. Prosp.*, 3, 268–294.

Koefoed O 1979: Geosounding principles 1, resistivity sounding measurements. Elsevier

Lanczos C 1961: Linear differential operators. D. Van. Nostrand

Marquardt D W 1963: *J. Soc. Ind. Appl. Math.*, 11, 431–441.

Marquardt D W 1970: *Technometric*, 12, 591–612.

Sri Niwas, Pawan Kumar, Wason H R 1982: *Proc. Ind. Acad. Sci., Earth and Planetary Sciences*, 91, 29–41.

Sri Niwas, Pramod Kumar 1991: *Acta Geod. Geoph. Mont. Hung.*, 26, 435–452.

Tikhonov A N, Arsenin V Y 1977: Solution of ill-posed problems. Winston and Sons

Twomey S 1977: Introduction to the mathematics of inversion in remote sensing and indirect measurement. Elsevier

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 27–42 (1997)

PC IMPLEMENTATION OF THE SODES2 PROGRAM SYSTEM AND OPTIMIZATION OF THE PAKS NETWORK

SAAD YAZJI[1]

[Manuscript received October 14, 1996]

This paper informs about the implementation of the software SODES2 program system, which was written in the University of Karlsruhe in Germany, on microcomputers to solve second order design (S.O.D.) problem of two-dimensional networks. This program system relies on the least-squares approximation of the inverse criterion matrix. It was written in Fortran-77 working only on a mainframe computer of type Siemens 7881. It consists of the components KRIT, SODES2 and NETZPLOT. This paper shows that this system can be used at research institutions and at private surveying offices as well.

The example chosen to explain this program was the Paks power station network. Some proposals are given depending on the results obtained for Paks network which consists of 37 points.

Keywords: criterion matrix; group weights; individual weights; least-squares approximation; observations; optimal network design; S.O.D.; SODES2-program system

1. Introduction

The history of the optimal design of geodetic networks tells us that this activity stopped ten years ago.

Kuang (1991) restored the important role of the optimization of the geodetic networks. Although Kuang (1992) introduced a new approach to solve second order design (S.O.D.) of geodetic networks, his study was also based on the least-squares principle and the criterion matrix as objective function.

The above-mentioned least-squares principle which has been used since more than 200 years in all sciences, is still of big importance in the field of optimization of geodetic networks. In addition to the least-squares principle, the criterion matrix concept as an ideal variance-covariance matrix of the coordinates is still valid.

Gáspár and Schmitt (1989) implemented a program to solve the S.O.D. problem. This program relies on the principle of discrete programming and on different criteria of the precision but not on the criterion matrix.

I found it interesting and beneficial to have a program to solve the S.O.D. by the least-squares principle and criterion matrices. The software package for the S.O.D. of two-dimensional networks (Müller and Schmitt 1985) I implemented including its three components on a microcomputer. This implementation was done for the two modules: KRIT and SODES2. The third component NETZPLOT was written by me in Turbo Pascal.

[1]Technical University Budapest, Department of Surveying, H–1111 Budapest, Műegyetem rkp. 3, Hungary

1217-8977/97/$ 5.00

2. The principles of the program system

2.1 The least-squares principle

The first application of least-squares was in astronomy by Gauss (referred by Mikhail 1976). Its practical importance has been enhanced by the introduction of electronic computers, formulation of its techniques in matrix notation, connecting its concept to statistics, and by the linear normal equations.

The mathematical principle of least-squares is briefly as follows:

At first the functional model is determined and the observations are assumed to be redundant.

The redundant observations are usually not compatible with the model because of the inherent stochastic properties of observations. In this situation it is necessary, with certain optimum properties, to derive a unique set of estimates for all the model variables.

If $\mathbf{L}$ is the observation vector, and $\hat{\mathbf{L}}$ the estimated observation vector, $\mathbf{V} = \hat{\mathbf{L}} - \mathbf{L}$ is the residual vector which plays an important role after the adjustment process.

Because of redundancy, an infinite number of estimates for $\mathbf{V}$ or $\mathbf{L}$, satify the model, so these estimates have to satisfy another criterion, which is the least-squares principle. This principle states that:

$$\Phi = \mathbf{V}^t\mathbf{W}\mathbf{V} \rightarrow \text{minimum}\,,$$

where $\mathbf{W}$ is the weight matrix of the observations.

It is important that the application of this principle does not require *a priori* knowledge of the distribution associated with the observations.

2.2 Criterion matrices

Criterion matrices are a substitute for non available real variance-covariance matrices.

In network design a criterion matrix is a target with special properties which should be realized by the designed observation scheme (Bill 1985). He stated that the criterion matrices have the most important role in planar geodetic networks:

a) the main range of their application is the optimization of geodetic networks, where a criterion matrix is substitute for the desired variance-covariance matrix. The designed network fulfils a best approximation in the sense of a general least-squares solution to the criterion matrix (Schmitt 1985a);

b) the second application of the criterion matrices is the quality description of geodetic networks (Van Mierlo 1981), where a criterion matrix defines a target which can be computed with reality by solving the general eigenvalue problem;

c) the third application is in network densification problems, where the theoretical model of dynamic network densification (Bähr 1973) can be approximated and restricted on the the local densification area by taking a criterion matrix as a substitute for the non available real variance-covariance matrix of the higher order network points.

The principle of criterion matrices was introduced in geodetic networks by Grafarend (1972) and Baarda (1973).

Grafarend (1972) showed that the covariance matrix of a random field of error vectors in a geodetic network has the so-called Taylor-Karman (T-K) structure, i.e. the correlations of the vector field are concentrated into distance-dependent functions, the longitudinal and the transversal correlation functions. So, all results depend on these functions. If we suppose a common correlation function in both directions, we get a special case of the T-K structure which is called the completely isotropic structure.

Baarda (1973) proposed a regression straight line as covariance function and introduced a completely isotropic model and also showed how to transform the regular criterion matrix into a singular one using a similarity-transformation to define a datum system.

In conclusion there is an intrinsic relationship between the principle of the least-squares and the principle of the criterion matrices which finally leads to acceptable results in the field of optimization of the geodetic networks.

3. Solution method

The approach for the solution of the S.O.D. problem is the general least-squares approximation of the inverse criterion matrix chosen by Schmitt (1985b). This approach is the best solution among the known three approaches: the direct approximation of the criterion matrix (approach i), the iterative approximation of the criterion matrix (approach ii), and our solution, the direct approximation of the inverse criterion matrix (approach iii).

In other words we call our approximation, after a modification according to Müller and Illner (1984) a modified direct approximation of the inverse criterion matrix. By this modification we get a better approximation of the criterion matrix itself.

This solution has high stability with respect to changes in the criterion matrix and its numerical efficiency.

This approach has been developed using the three-step-strategy of Karlsruhe (Müller and Schmitt 1985). The first step starts with an observation plan, including all possible observations and solving the S.O.D. problem with individual weights in order to see the effect of each single observation. In the second step the observation plan is reduced by using accuracy, reliability, and cost criteria after corresponding analyses. For example, all observations which result from the first step with low weights can be eliminated, and those observations which do not influence the controllability of other measurements and other observations which require high costs for setting up or beaconing stations can be eliminated, too. In the third step we find the final solution of the reduced observation plan, where group weights are now introduced, which is necessary for real measurements, for example, sets of directions or groups of distances with approximately equal lengths. Finally we ascertain the accuracy and the reliability.

Now I mathematically explain the principal equations of this approach as obtained from adjustment by the variation of co-ordinates.

The basic equation is:

$$\mathbf{N} = \mathbf{Q}^{-1}$$

or for free networks:

$$\mathbf{A}^t\mathbf{P}\mathbf{A} = \mathbf{Q}_x^+ , \tag{1}$$

where $\mathbf{A}$ is the design matrix, $\mathbf{P}$ is the unknown weight matrix of the observations, $\mathbf{Q}_x^+$ is the pseudoinverse of $\mathbf{Q}_x$.

By using the Khatri-Rao-product (Rao and Mitra 1971) we get a set of linear equations:

$$(\mathbf{A}^t \otimes \mathbf{A}^t)\mathrm{vec}\mathbf{dP} = \mathrm{vec}\mathbf{Q}_x^+ . \tag{2}$$

By solving it:

$$\mathrm{vec}\mathbf{dP} = (\mathbf{A}^t \otimes \mathbf{A}^t)^+\mathrm{vec}\mathbf{Q}_x^+ . \tag{3}$$

From Eq. (1) the matrix of the residuals $\mathbf{D}$:

$$\mathbf{D} = (\mathbf{A}^t\mathbf{P}\mathbf{A})^+ - \mathbf{Q}_x . \tag{4}$$

This matrix expresses the quality of the approximation and by supposing the square sum of the residuals — similarly to a least-squares adjustment — as global measure for the approximation quality one gets:

$$\mathbf{d}^t\mathbf{d}, \tag{5}$$

where $\mathbf{d} = \mathrm{vec}\mathbf{D}$.

By applying the least-squares principle:

$$\mathbf{d}^t\mathbf{d} \to \min \Rightarrow (\mathbf{A}\mathbf{A}^t * \mathbf{A}\mathbf{A}^t)\mathrm{vec}\mathbf{dP} - (\mathbf{A}^t \otimes \mathbf{A}^t)\mathrm{vec}\mathbf{Q}_x^+ = \mathbf{0} , \tag{6}$$

where $*$ is the Hadamard-product. To transform linearly this solution, the weight vector has to be multiplied by a factor λ found by Müller and Illner (1984):

$$\mathbf{P} = \mathrm{diag}(\lambda\mathrm{vec}\mathbf{dP}) . \tag{7}$$

4. Program system

According to Müller and Schmitt (1985) this system consists of three components. (The system was written in Fortran-77 and running on a mainframe computer of type Siemens 8771 under the operating system BS 3000.)

— KRIT: produces a criterion matrix of T-K-structure or a completely isotropic one and allows an arbitrary datum transformation. This program gives criterion matrices for the defects 2, 3 and 4 with one or two fixed points as S-base (for the defect 2 or 4) or with a restricted group of points respectively with all points as datum for free networks.

The input for KRIT is a list of co-ordinates, some code numbers for the desired type of the criterion matrix and the desired S-transformation and the characteristic distance for the corresponding correlation function must also be defined.

The main output of KRIT is the original or transformed criterion matrix.

— SODES2: computes the S.O.D.-solution for individual and group weights and analyses the resulting design with respect to its accuracy and reliability. This program is the central component for the S.O.D.-solution by the least-squares approximation of the inverse criterion matrix.

The program is only working to optimize free networks. This program relies on the 3-step-strategy of Karlsruhe. The following observation types are admitted: distances; directions; azimuths; distance-ratios; angles; distances and azimuths.

In this component of S.O.D. an internal calculation is only possible for a regular Taylor-Karman structure of the potential type or for a completely isotropic structure of the criterion matrix.

As input some code numbers are needed to distinguish the weight estimation, that means, individual or group weights, the unit weight to fix the size of error circles, a pair of numbers to compute the relative error ellipses and finally a list with the connection information of the planned observations. All these data and also the criterion matrix can be read from a file if it was computed with the program KRIT, or an internal calculation was made for a regular Taylor-Karman structure of potential type or a completely isotropic structure of the criterion matrices.

The output of the program is the list of co-ordinates, the postulated error and the relative error situations are computed from the criterion matrix, measures for the global quality of the approximation of the criterion matrix, a list of observations of the resulting design with weights and their repetition numbers, standard deviations related to the unit weight and reliability measures, a list of eliminated measurements from the original observation plan and point and relative error ellipses which express the accuracy of the final design. All these outputs are written on a file.

— NETZPLOT: produces a plot of the network design including postulated and realized error ellipses.

I have written this component in Turbo Pascal, by which a plot of the final network design can be drawn. As input the file is needed which was produced by SODES2. This input includes the point numbers with their co-ordinates, the postulated and effective error ellipses. The size or the scale of the drawing has to be defined by the user.

5. The chosen way to modify the components KRIT and SODES2

Because of the length of the main component SODES2 it was necessary to find an adequate compiler. Because the Microsoft's compiler did not solve it, I have finally found the compiler SVS Fortran-77 which has solved this problem.

Therefore at first I wrote the program system on a file, in the next step I added some modifications to the software by which the input data can be read from a file not to input them on the screen which is a time consuming way.

Because of the long component (SODES2) the whole memory of the personal computer is used.

As input for the component KRIT the correlation coefficients are needed that means, the longitudinals and the transversals of the modified Bessel function a program in Fortran-77 and in Turbo Pascal computes them.

To compute these coefficients I used the original relations of the modified Bessel function found at Abramowitz and Stegun (1970), and at Grafarend and Schaffrin (1979).

I also modified the software of SODES2 in a way by which I can previously define the accuracy of the instruments which are available in the project by the two factors (A, B) the relation of which define the accuracy of distance measurements, for example:

$$\mathbf{M}_{\mathrm{mm}} = (\mathbf{A}_{\mathrm{mm}} + \mathbf{B}_{\mathrm{mm}}/\mathbf{D}_{\mathrm{km}}) .$$

The same thing was done for the directions' accuracies, for example, the standard deviation of measuring one direction for one round is $m = 0.0006$ grad.

To check the program I have taken into consideration the same example used by Müller and Schmitt (1985). I got exactly the same results. I tested it by another example which is the PAKS power station network. In conclusion the program system is working on a personal computer as it was working in the past on a main frame.

According to the experience I got by this system it is quick enough, compared with other programs operated by linear discrete programing. It is obvious that this system is very effective, quick, and it can solve any network with a big size; for example, if the network has more than 60 points with more than 500 observations it takes one or two minutes.

6. Example

The first application of this software package is checking the PAKS power station network whether it is the optimal design of this network or not.

I got the data related to this network from a study which was made by the Geodetic Institute of the Technical University Budapest in February 1989 (Gáspár et al. 1989).

I have performed this study for the main network which consists of 37 points. I have changed the numbers of these points as Table I shows. This table contains the approximate co-ordinates of these points. I have used the observation plan of the observations' compaign which was done for distances and directions.

Table I. List of the main netpoints coordinates

New numbers	Old numbers	Y m	X m
1	1145	1033.5750	11187.5392
2	1146	0958.2458	11405.4795
3	1147	0937.6181	11868.5007
4	1148	0939.4330	12506.0272
5	1149	0977.6499	12753.9673
6	1150	1377.8610	13154.6916
7	1151	2063.2111	12950.7664
8	1152	1712.2962	12496.9146
9	1153	1578.4411	11729.1499
10	1154	1506.6724	11223.3617
11	1155	1358.9550	12505.9182
12	1156	1729.0945	12001.9681
13	1159	1107.6879	12493.0682
14	1160	1103.9740	12213.4754
15	1161	0936.2079	12206.6135
16	1162	1103.5523	11847.7414
17	1163	1207.3032	11780.6269
18	1164	1198.1075	11416.9458
19	1165	1229.2731	11195.9087
20	1166	1395.6985	11825.9102
21	1167	1500.3859	11983.4110
22	1168	1507.7809	12235.4741
23	1169	2323.4726	12772.1053
24	1170	1798.7417	13424.4322
25	1171	1347.7991	13543.0532
26	1172	0990.9982	13135.9463
27	1173	0616.7535	13800.4092
28	1174	0619.6484	13203.5699
29	1001	1667.7938	12413.4766
30	1002	1172.4094	11769.9558
31	1003	1329.3284	12475.2537
32	1227	1995.7990	12500.0331
33	1239	1853.7693	12225.0037
34	1203	0999.9988	11300.0037
35	1214	0999.9819	12399.9708
36	1214/1	0990.9888	12299.9827
37	35-3301/1	1072.5650	10900.9698

6.1 Studying of the network of 37 points

I will explain only one situation in which I used the highest accuracies of measuring distances and directions:

— the standard deviation of the unit weight $\sigma_0 = 1$ mm,

— the defect degree of the network is equal to 3, that means, the observations are distances and directions,

— the characteristic distance $d = 30$ m,

— the accuracy of measuring a distance is defined by the relation:

$$\mathbf{M}\ \text{mm} = 0.2\ \text{mm} + 0.2\ \text{mm/D}\ \text{km},$$

— the accuracy of measuring a direction is: $\mathbf{M}_p = 0.0006$ grad.

First step: Estimation of individual weights for all observations

– the number of distance observations: 176,

– the number of direction observations: 176.

At first the program automatically eliminates the linear related observations, after that the list of co-ordinates and the postulated error situation is obtained.

Because of the unsuitable determination of the criterion matrix and the observation plan according to Müller and Schmitt (1985), and also to Wimmer (1981), we obtained some negative weights, so we need additional constraints to get positive weights.

The program automatically eliminates these negative weight observations.

In a summary we have got negative weights for three distances and for three directions between the points: (30–29), (31–30), (31–29).

After that we get the global quality measures RR1, RR2 of the fitting with the inverse criterion matrix and respectively the criterion matrix before transforming the weights, where RR1, RR2 are the square sums of the residuals of the elements of difference matrix between the realized cofactor matrix and the inverse of the criterion matrix respectively of the criterion matrix itself. RR3, RR4 are also quality measures after transforming which have the same meaning as RR1, RR2. We note that RR1 < RR2, that means, the fitting with the inverse criterion matrix is better than the fitting with the criterion matrix itself. After linear transformation of the weights by a factor $\lambda = 2.6157$ we find RR3 < RR4, that means, also after the transformation the fitting is better with the inverse criterion matrix.

The weights sum before transformation is equal to 753.2397.

Second step: Reduction of the observation plan.

At first the linearly related and negative weight observations have to be eliminated. Secondly low observation weights which have insignificant influence on the precision and the reliability of the network must be eliminated. This elimination is done according to the standard deviations. For example, if for a distance the standard deviation is ten times larger than that of the most accurate distance, this distance can be eliminated. We applied the same method for eliminating directions with low weights, that means they were eliminated if the lateral deviation derived from the corresponding angular standard deviation is larger than the smallest lateral deviation by at least one order of magnitude.

According to this strategy we found that the minimum standard deviation of the distances was $\sigma_{\min} = 0.0005$ m. The distances which have standard deviations more than 0.005 m were eliminated. Such eliminated distances are (31–4). (29–4). As to directions, I found that the minimum lateral deviation is equal to $\Delta Q_{\min} = 0.0005$ m, thus the directions which have lateral deviations more than 0.005 m were eliminated. Such directions are (31–4) and (29–4) too.

At the end of the results of the individual weight situation we find the realized error ellipses with their axes and directions. Finally we see the geometric determination of the redundancy beside the sum of independent redundancies, which must be the same:

$$\begin{aligned} R &= N - U + D &&= 94 && \text{(geometric determination)} \\ R &= R(1) + \ldots + R(N) &&= 94.000 && \text{(the sum of independent redundancies).} \end{aligned}$$

In other words R is the number of redundant observations in the observation plan of the campaign. After these two steps the observation plan became as follows:

– the number of distance observations is: 83,

– the number of direction observations is: 83.

Third step: Estimation of group weights for the remaining observations.

The realization of individual weights by practical measurements is almost impossible. To solve this problem we must collect the directions and also the distances in groups, for example, all directions which start from the same point have to be collected in one group and get one common weight. By the same way we can divide the distances into five weight groups according to the accuracy of the electronic measurement instruments of distances.

At first we divided these distances into three sets (short-range, medium-range, long-range) because of the accuracy of electronic distance measurements depends essentially on the range. The long-range and medium-range groups are split into two sets each according to the measurements on one day or on different days. The distances were thus divided into five groups as follows:

- the first group: $S < 130$ m,
- the second group: $130 < S < 270$ m (measured on different days),
- the third group: $130 < S < 270$ m (measured on one day),
- the fourth group: $S > 270$ m (measured on different days),
- the fifth group: $S > 270$ m (measured on one day).

The number of directions in this step is 166.

The number of distances in this step is 83.

Because of the unsuitable determination of the criterion matrix with the observation plan, three groups of directions have negative weights. All these

Table II.

The case	For individual weights	For group weights
Number of total observations	166	232
Number of distance measurements	83	83
Number of direction measurements	83	149
Number of redundant observations	94	127
Degree of defects of the network	2	3
Standard deviation of the unit weight	1 mm	1mm
Number of network points	37	37

directions started from the points (30, 31, 34). In other words 17 observations are eliminated because of their negative weights.

We found that RR1 < RR2, that means, the fitting with the inverse criterion matrix is better than the fitting with the criterion matrix itself. We found that RR3 > RR4, that means, the fitting with the criterion matrix is better than the fitting with the inverse criterion matrix after transformation.

The geometric determination redundancy is $R = 127$.

Te sum of the independent redundancies is $R = 127.000$.

The final number of distance measurement is 83.

The final number of direction measurements is 149.

All these results are summarized in Table II.

Finally in Fig. 1 the plot of the points and of the postulated and realized ellipses can be seen for individual weigths and Fig. 2 for group weight.

7. Remarks

I made some experiments for this network using different types of criterion matrices and by using different accuracies of the instruments. By comparing the results to each other I found the following:

7.1 For individual weights

— In all cases when using criterion matrices negative weights appeared, but their numbers are different from one case to another. However these negative weights are restricted to the observations (30–29, 31–30, 31–29) both for distances or directions, or both together.

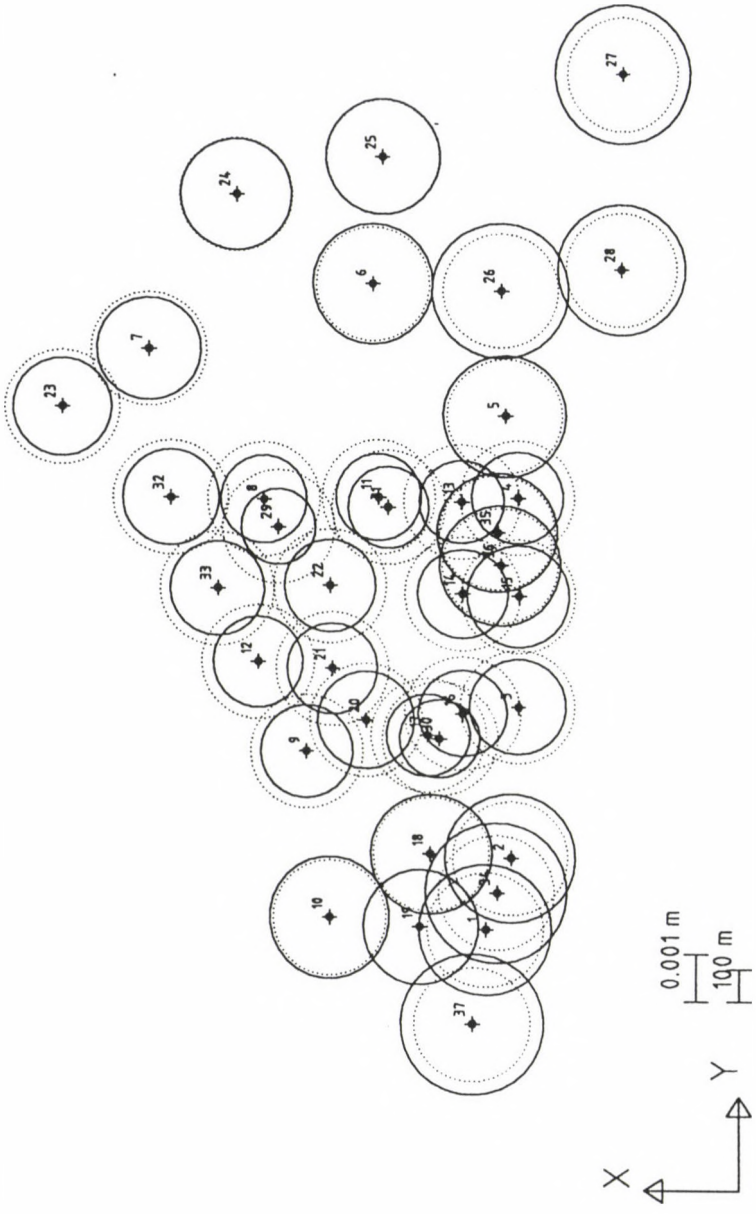


Fig. 1.

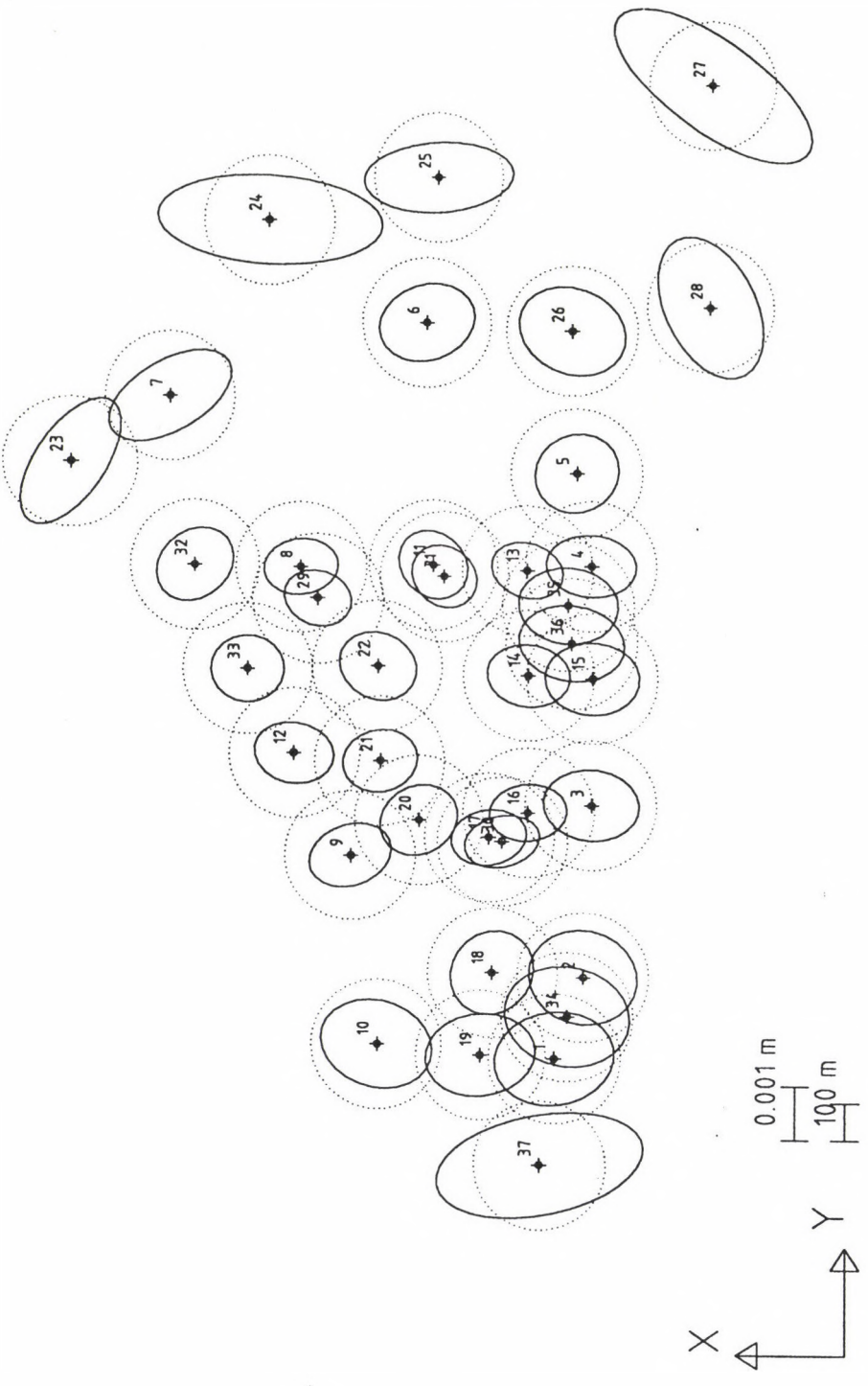


Fig. 2.

— There are also other observations which have low weights, by eliminating them, the precision, the reliability and the cost criteria are not influenced. These observations are (31–4, 29–4) both distances or directions or both together.

— In all cases of criterion matrices I found that the sum of the weights before transformation when using a singular (transformed) matrix is less than the same sum when using the regular criterion matrix.

— In all cases the weight factor λ is small when a regular criterion matrix is used instead of using a singular one.

— The criterion matrix which gives us the smallest sum of the weights before transformation is the singular Taylor-Karman structure of turbulence type.

— The criterion matrix which gives us the largest number of redundant observations is the singular identity matrix.

— The criterion matrix which gives us the smallest total number of observations is the regular or singular Taylor-Karman structured complete and isotropic matix.

— In all cases I have found that the global quality measures are as follows: RR1 < RR2, RR3 < RR4, that means the fitting is better for the inverse criterion matrix.

7.2 For weight groups

— All results obtained were the same for regular or singular criterion matrices except the identity matrix solution.

— The Taylor-Karman structure of turbulence type gives the lowest sum of weights as it was for individual weights.

— The Taylor-Karman structure of potential type gives the highest number of the redundant observations.

— The Taylor-Karman structure of turbulence type gives the lowest number of the total observations.

— In all cases I got negative weights but only for direction observations.

— In all cases the sets of directions from points (30, 31, 34) have negative weights, thus they are eliminated. In the case of Taylor-Karman structure of potential type there were only two negative directions (34–2, 34–1).

— According to the global quality measures RR1 < RR2, but RR3 > RR4, i.e. the fitting is better for the inverse criterion matrix before the weight transformation, and it is better for the criterion matrix itself after the transforming.

— We have three criterion matrices which give us good results for the designed network. These are:

  - the Taylor-Karman structure of potential type,
  - the Taylor-Karman structure of turbulence type,
  - the Taylor-Karman structure of completely isotropic case.

— I prefer to use criterion matrix T-K structure of completely isotropic type, because the weights are better to be realized then from the others.

The solution of the S.O.D. problem by the direct approximation of the inverse criterion matrices has high stability with respect to changes in the criterion matrices and its numerical efficiency (Schmitt 1985b).

8. Conclusions

1. There is a clear relationship between the appearance of negative weights and the structure of the criterion matrices, where I have noted that negative weights appeared in all cases both with individual or groups weights, and both for the singular or regular criterion matrices.

   Observations (directions, distances) which do not have negative weights, in some cases have low weights and may be eliminated because of their weak influence. Thus we must not be afraid of this appearance of the negative weights because of these observations have a weak effect on the accuracy, reliability and cost criteria of the network.

2. By using the regular criterion matrices I got better results especially from the economic point of view. Thus regular criterion matrices satisfy the cost criterion and also does not influence the other criteria, i.e. precision and reliability. Sometimes another problem arises: the optimization process yields in some points, unrealistic high weights and therewith accuracies which are not realizable and cannot be accepted (Schmitt 1977, Schmitt et al. 1978 and Fritsch 1982).

3. As I mentioned the best solution is the Taylor-Karman structured complete and isotropic matrix. In spite of the fact that this solution is the best, a lot of designed observations have high weights respectively high repetition numbers, that means, that the accuracies cannot be practically achieved. These observations were restricted in the direction measurements to the following points (5–6–7–8–9–10–18–26–32–24–25–27–28–37). Some other observations do not satisfy the reliability criterion, these observations are: the distance (10–19), and the directions (18–17), (18–19). The points in which the isotropy does not realize are: (7–23–24–25–27–28–37), as seen in Fig. 2 of the error ellipses. Thus these points are not determined sufficiently well, and the network has a shape defect which tells us that this network has to be studied once again according to the principles of the optimization of geodetic networks. My future work will be to design a new network for the PAKS power station.

4. I have done an experiment within which I calculated all effective relative error ellipses, and I found that the completely isotropic situation is realized by using individual weights, where the effective relative error ellipses are circles, but with weight groups, these relative ellipses are not circles, so the complete isotropic situation is missing. Thus it is impossible to realize the complete isotropic situation when the defect degree of the network is equal to 3, this is in concordance with the opinion of Bill (1985), Gáspár (1986) and Müller (1985).

5. Wimmer (1981) said that with the T-K-structured criterion matrices no essential results are to be expected, so he recommended to use the identity matrix as representation of the T-K-structured matrices. But Schmitt (1977) contradicted this opinion and stated that using the identity matrix as criterion matrix has some disadvantages, for example, correlations between the points of a net are not allowed. This statement is not correct as it leads effectively to the elimination of measurements between new points in the observation plan of point bisections. Network reliability and neighbouring accuracy are strongly neglected. Gáspár (1987) also advocated this opinion where he criticized the use of the identity matrix as a criterion matrix because it really suffices the conditions of perfect isotropy and homogeneity but the shape of network is not considered. Results were in concordance with these opinions as it was found that the T-K-structured turbulence type, potential type and the complete isotropic criterion matrices gave better results than the identity criterion matrix.

References

Abramowitz M, Stegun I A 1970: Handbook of mathematicl functions. Dover, New York

Baarda W 1973: S-transformations and criterion matrices. Netherlands Geodetic Commission, Vol. 5, No. 1.

Bähr H G 1973: Zum Ausgleichungsverfahren bei der Ergänzung bestehender geodätischer Netze. Deutsche Geodätische Kommission, München, Reihe C, No. 187.

Bill R 1985: Kriteriummatrizen ebener geodätischer Netze. Deutsche Geodätische Kommission, München, Reihe A, No. 102.

Fritsch D 1982: Second order design of geodetic networks: problems and examples. Deutsche Geodätische Kommission, München, Reihe B, No. 258.

Gáspár P 1986: *Periodica Politechnica Civil Engineering, Budapest*, 30, No. 1–2.

Gáspár P 1987: *Geodézia*, 95, 39–45.

Gáspár P, Schmitt G 1989: *Allgemeine Vermessungs-Nachrichten*, 217–227.

Gáspár P, Zoltán S, Károly Sz 1989: Common adjustment of the measurement results in the frame of the maintenance of the horizontal main network of basic points in connection with the 2 × 1000 MW enlargement of the Paks Nuclear Power Plant. (manuscript)

Grafarend E 1972: Genauigkeitsma"se geodätische Netze. Deutsche Geodätische Kommission, München, Reihe A, No. 73.

Grafarend E, Schaffrin B 1979: *Zeitschrift für Vermessungswesen*, 104, 133–149.

Kuang S L 1991: Optimization and design of deformation monitoring schemes. Ph.D. dissertation, Department of Surveying Engineering, University of New Brunswick

Kuang S L 1992: *Survey Review*, 31, 279–288.

Mikhail E M 1976: Observations and least-squares. IEP Dun-Donelly Publisher, New York

Müller H 1985: *Bulletin Geodesique*, 59, 316–331.

Müller H, Illner M 1984: *Allgemeine Vermessungs-Nachrichten*, 253–269.

Müller H, Schmitt G 1985: SODES2 – Ein Programm System zur Gewichtsoptimierung zweidimensionaler geodätischer Netze. Deutsche Geodätische Kommission, München, Reihe B, No. 276.

Rao C R, Mitra S K 1971: Generalized inverse of matrices and its applications. J. Wiley, New York

Schmitt G 1977: Experience with the second order design of geodetic networks. Proc. Int. Symp. of Design and Computation of Control Networks, Sopron, Hungary

Schmitt G 1985a: In: Optimization and design of geodetic networks. Springer Verlag, Berlin, 6–10.

Schmitt G 1985b: Optimization and design of geodetic networks. Springer Verlag, Berlin, 74–121.

Schmitt G, Grafarend E, Schaffrin B 1978: *Manuscripta Geodaetica*, 3, 1–22.

Van Mierlo J 1981: *Allgemeine Vermessungs-Nachrichten*, 88, 95–101.

Wimmer H 1981: In: Proceeding of the IAG Symposium on geodetic networks and computations, Sigle R ed., Deutsche Geodätische Kommission, München, Reiche C, 258/III, 112–127.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 43–52 (1997)

RECENT CRUSTAL MOVEMENTS AND SEISMICITY STUDIES AROUND ACTIVE FAULTS, ASWAN, EGYPT

ABD EL MONEM S MOHAMED[1]

[Manuscript received December 10, 1996]

Monitoring of recent crustal movements was carried out in Egypt for the first time in the Aswan region after the occurrence of the November 14, 1981 earthquake with a magnitude of 5.5 in the Kalabsha area 60 km southwest from the Aswan High Dam. This earthquake was considered a very important event as it was located not far from the Aswan High Dam. Due to its possible association with the impoundment of the Aswan Lake, especially the water activity in and around this lake.

The data used in this study consists of the geodetic data and the catalogue of earthquake determined by the Aswan seismic network. In this work, we are trying to give a clear picture about the order of movements and their role in earthquake occurrence in the Aswan region. The analysis of such measurements together with the frequency and intensity of seismic activity and the changes of the water level, can lead to understanding the cause of earthquake occurrence and its origin.

Keywords: computer programs; geodetic data; seismic data; water level variation data

Introduction

Monitoring recent crustal movements has become one of the major and important tasks facing geodesists, since it has vital consequences for human safety and economic achievements. There are a variety of geodetic and other techniques that are currently in use for monitoring such deformations on all local, regional and global scales. In fact, the recent crustal movements nowadays have several implications in the Egyptian territory. For instance, there have been earthquake activities in the Aswan region since 1981, especially in the northwest of the Aswan Lake. These earthquakes were considered very important events, as they were very close to the Aswan High Dam. The Aswan High Dam is a unique case in the world where a single dam affects the whole country. Therefore, several study programs were initiated. These programs include monitoring seismicity and recent crustal movement.

The whole area of the northwest of the Aswan Lake is characterized by many structural faults (Fig. 1). For this purpose it has become necessary to monitor crustal movements around the active faults in order to understand the geodynamics of the area.

In order to monitor horizontal and vertical crustal movements, using geodetic methods, around the northern part of the Aswan Lake to investigate their possible association with the earthquake activity and the water loading in the lake and to

[1]National Research Institute of Astronomy and Geophysics, Helwan, Cairo, Egypt

1217-8977/97/$ 5.00

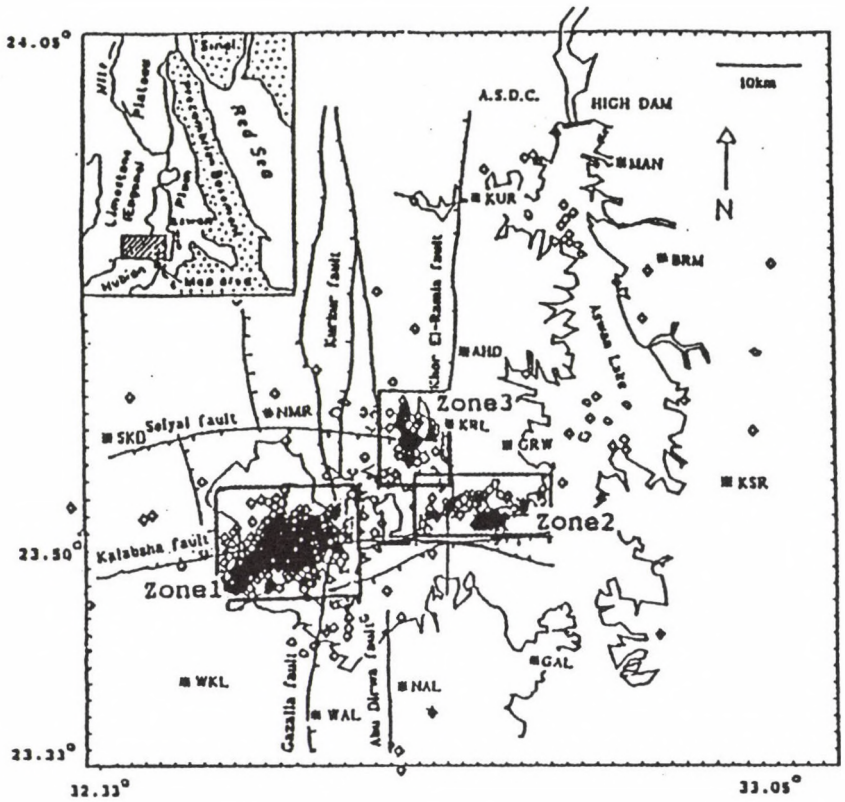


Fig. 1. Microearthquakes recorded by the Aswan network from 1982 to 1992. ⊥⊥⊥ faults (taken from map by Issawi (1969)), ◇ epicenters, * seismic station

understand the geodynamics of the area, a long term study program was initiated. This program includes establishment and measurement of local and sub-regional geodetic networks around potential faults.

The main purpose of this work is to investigate the relationship between the recent crustal movements and the earthquake activity around active faults in the Aswan region, in order to predict the possible earthquake activities and to diminish the seismic risk within the area of Aswan as well as safety of the Aswan High Dam and its economic resources.

Seismicity at Aswan region

The seismicity in the Aswan region was examined using microearthquake data recorded by the Aswan Radio Telemetry Network, surrounding the northern part of the lake between (1981–1992). Analysis of seismic data with the water level variation in the lake and monitoring the crustal movements around active faults can lead to determining how well it represents past seismicity and what it might

Table I. The extent of the three active zones of the specified area

Zones	Latitude extent		Longitude extent	
Zone 1	23.49°N	23.58°N	32.49°E	32.61°E
Zone 2	23.54°N	23.60°N	32.68°E	32.85°E
Zone 3	23.62°N	23.68°N	32.65°E	32.73°E

imply about future earthquakes, and to detect possible periodicity in total energy release during the periods of the geodetic measurements.

The seismicity northwest of the Aswan Lake is concentrated at Gebel Marawa, near the intersection of the easterly trending faults with a northern trend. The seismicity at Aswan is clustered in three main zones (Fig.1 and Table I):

1. Gebel Marawa,
2. east of Gebel Marawa, and
3. northeast of Marawa along the Khor El-Ramila N-S fault.

The most active zone (zone 1) is directly beneath Gebel Marawa on the Kalabsha fault, at a depth of 15 to 25 km. This deeper activity is taking place where the two faults sets intersect beneath Gebel Marawa. A second zone (zone 2) of much less activity is located farther east along the Kalabsha fault, at a depth of 5 to 10 km. The third zone (zone 3) is located east of the Seiyal fault and south of Wadi Kurkur, at a depth of 0 to 5 km. All activity outside the Marawa area is shallower, i.e. 0–10 km, with no activity below 10 km. These three seismic zones are located in the mainstream of the Nile between the High Dam and Wadi Kalabsha (Mohamed 1994, 1995).

Figures 2a, b and c show the energy release over three main zones in the Aswan region (Fig. 1). Figure 2a shows the energy release over the most active zone in the area (zone 1). The level of activity in this zone is gradually decreasing with time.

Figure 2b shows the energy release over the second zone. The activity level is concentrated between 1982 to 1988, with a low level at 1986 and after 1988.

Figure 2c shows the energy release over the third zone. The level of seismicity in this zone increased with time until 1989, and began to increase again in 1992.

From Figs 2a, b and c, it was found that, the level of seismic activity gradually migrated east and northeastward, (shallower zones), while the level of seismicity decreased in the deep zone.

Crustal movement monitoring around the active faults

The data which were used in this work, are taken from the Kalabsha and Seiyal geodetic networks. Eight repeated horizontal geodetic measurements for the Kalabsha network and six repeated horizontal measurements for the Seiyal network were adjusted and analysed. Computer programs for the adjustment and transformation of coordinates in different epochs were used in these computations. The displace-

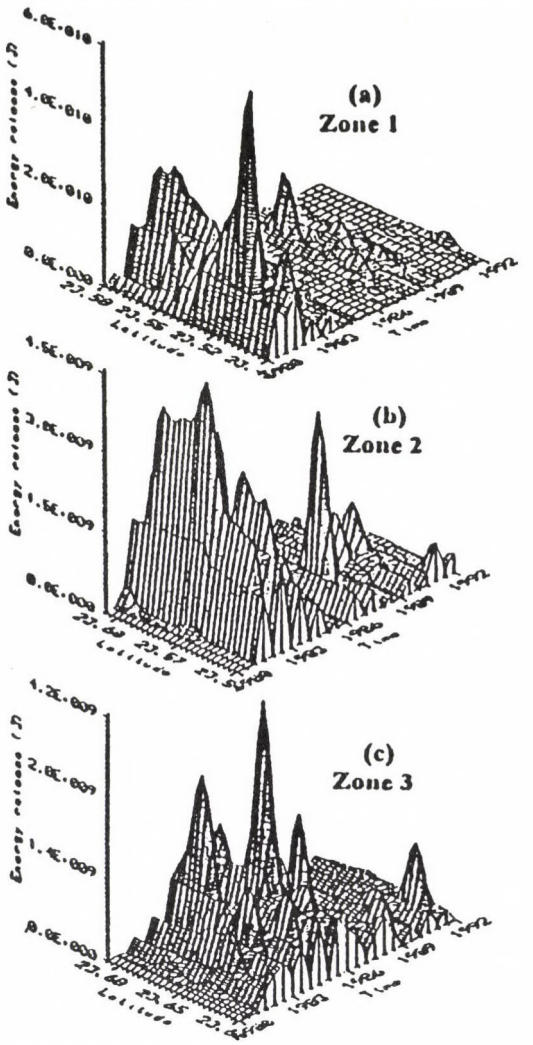


Fig. 2. Energy release by earthquakes in the three main zones (Aswan region) as a function of time and latitude for the period Dec. 1981 to Jun. 1992

ment vectors for each of measurements were calculated from the coordinate changes. Considering the confidence limit, most of these displacement vectors can be mainly attributed to the movement with the Kalabsha and Seiyal areas in these epochs of measurement. The displacements from coordinate differences were used as a tool for monitoring deformation and it gave a good idea of the deformation trend.

Crustal movements monitoring around the Kalabsha fault

The movements around the Kalabsha fault are determined for the different epochs of measurements. The results of these movements represent the form of the dynamic model for the deformations which occurred along the Kalabsha fault during the different epochs of measurements.

The minimalization of all stations in the southern part of the Kalabsha fault being covered by the geodetic network, and the northern part which becomes free of the movement as an example can be shown in Fig. 3.

Figure 4 gives a graphic display of the estimated deformation of the area which is covered by the Kalabsha network. In this model, block "A" contains the points of the network which lie north the fault, and block "B" contains the points which lie south the fault zone. The graphic display of the movements of the block "A" clearly shows the general direction of the relative displacements along the Kalabsha fault.

On the basis of the repeated geodetic measurements, seismicity of the area and geophysical as well as geological data, the present state of the geodynamic properties of the Kalabsha area are been studied. The relationship between the geodetic results, the seismicity in Kalabsha area and water level in the Aswan Lake are correlated as:

a) The period from 1982 until 1988, the water level in the lake decreased with

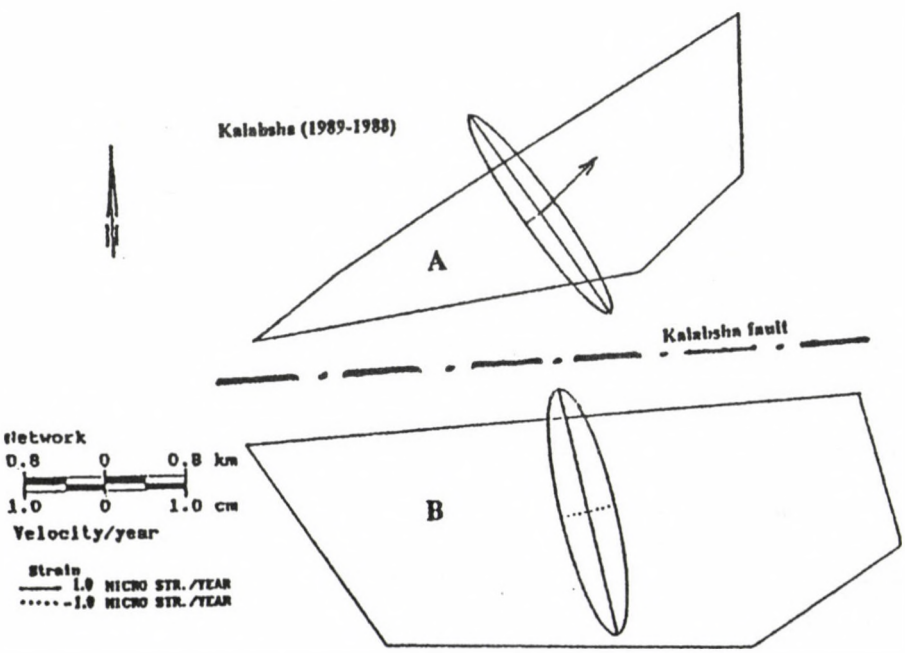


Fig. 3. Velocity of rigid block translation around the Kalabsha fault from Jan. 1988 to Feb. 1989

time. The seismicity also decreased with time. The magnitude of movements detected along the fault is variable for the different epochs of measurements in the same period.

b) The period from 1988 to 1989, the water level increased quickly and is reaching approximately 170 m as can be shown in Fig. 5. The seismic activity increased in this period (Fig. 2a). The magnitude of movements for all stations give significant changes and large movements as shown in Fig 6.

c) The period from 1989 to 1992, the water level in the lake is variable between 165 m and 170 m. The seismic activity increased only in the second and third zones (Fig. 2b and c), while the first zone is relativity stable in seismicity. The magnitude of movements are variable for these epochs.

Crustal movements monitoring around the Seiyal fault

The movements around the Seiyal fault are determined for the different epochs of measurements. The results of these movements represent the form of the preliminary dynamic model for the movements which occurred along the Seiyal fault plane during the different epochs of measurements. In comparison with the seismic

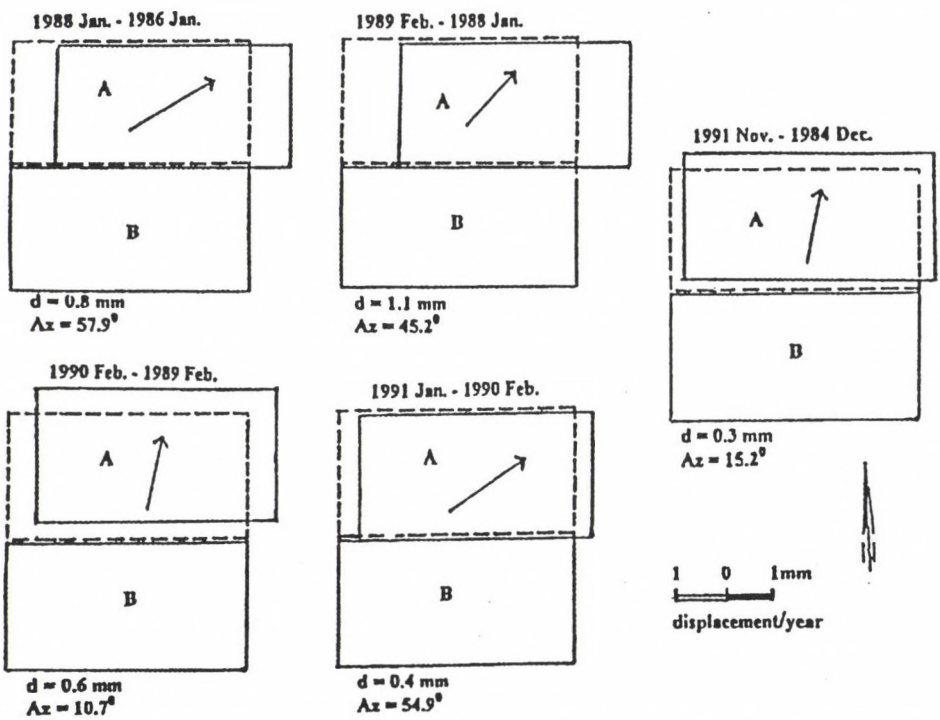


Fig. 4. Detected displacement of Block A with respect to remaining to Block B of the monitoring network (arrows indicate the azimuth angle)

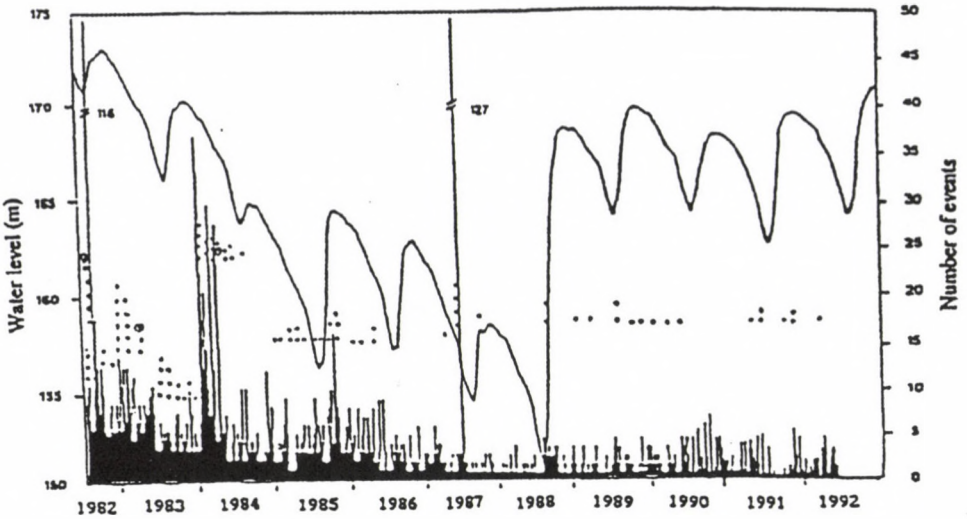


Fig. 5. Water level of the Aswan Lake representing the seasonal peaks from 1982 to the middle of 1992 and the number of events per day. o earthquakes with magnitude ≥ 4 and • are earthquakes with magnitude between 3 and 4

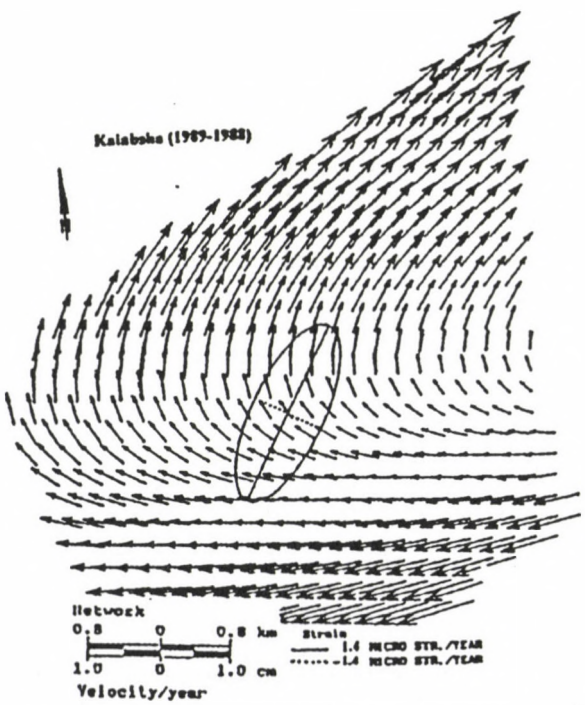


Fig. 6. Horizontal movements of the network for the epoch from Jan. 1988 to Feb. 1989

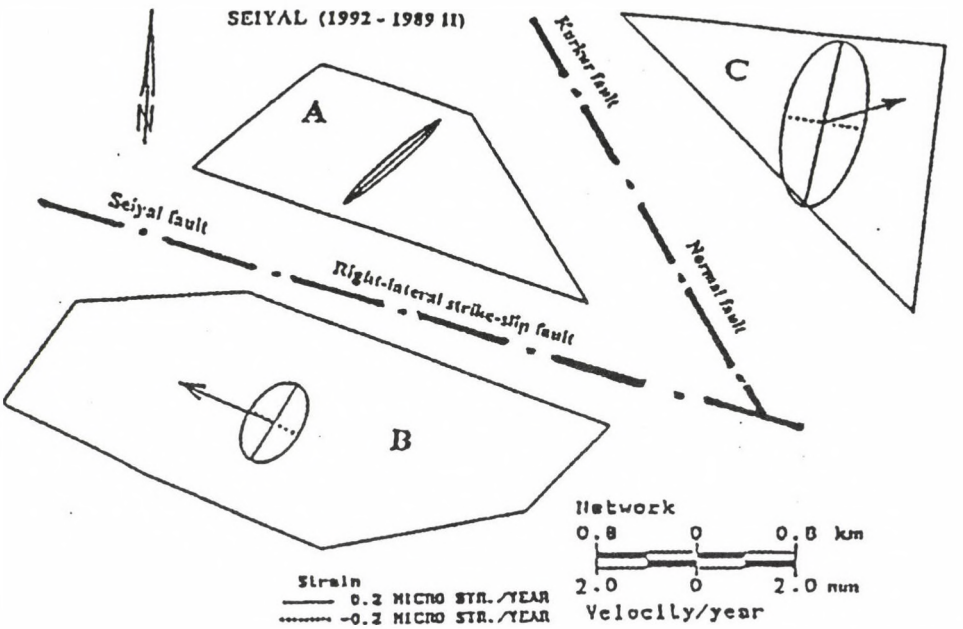


Fig. 7. Velocity of rigid block translation around the Seiyal and Kurkur faults from Nov. 1989 to Feb. 1992

events close to the Seiyal area, the movements along the Seiyal fault are correlated with results of the focal mechanism and geological evidence. It is noticed that the direction of horizontal movement agrees with the focal mechanism obtained from the seismic activity and with the direction of movement obtained from the geological evidence. For the different epochs of geodetic measurements, the magnitude of movement along the Seiyal fault is variable and increases with time.

The minimalization of all stations in the northwest part of the Seiyal network, and the south and east which became free of the movement as an example can be shown in Fig. 7. In this case the movement around the Seiyal and Kurkur faults can be obtained.

Although the seismic events close to the Seiyal area are relatively small, there is a movement along the Seiyal fault. This movement seems to be symmetric around the Seiyal fault, which is indicated from horizontal and levelling measurements analysis. According to the elastic rebound theory, the blocks on either side of the fault begin to move relative to each other, some time before an earthquake.

Conclusions

In comparison geodetic results with the seismic activity in the Kalabsha area are found to have a strong relationship between the accumulated-released sequences associated with the earthquake occurrences and the movement around the Kalabsha fault. The rate of accumulated strain in the area of the Kalabsha geodetic network

increases with the increasing of earthquake magnitude. This result verifies the strain accumulation principle accompanying earthquake occurrence and the visibility of crustal movements as an earthquake precursory. The geodetic results are in good agreement with seismic activity in the Kalabsha area.

The results of the analysis show that the Kalabsha fault is predominantly a normal fault with a right-lateral strike-slip component. The same conclusion was obtained from the focal mechanism of the seismic activity which was located in the Kalabsha area.

For the different epochs of geodetic measurements, the magnitudes of movement along the Seiyal fault are variable and increase with time. They indicate that the Seiyal fault is a right-lateral strike-slip fault and the Kurkur fault is a normal fault.

According to the stress and strain fields which were calculated for the different epochs of measurements, the northwestern area of the Aswan Lake suffers from stress and strain. The direction of the stress force is from SE to NW. The same conclusion was obtained from the focal mechanism solution of the earthquakes located in the Aswan region. This force may be due to two reasons:

— The first possible explanation (regional), results from the fact that the direction of this force is in accordance with the maximum horizontal compression induced by the approximately SE to NW drift of the African plate (Schandelmeier et al. 1987).

— The second explanation (local) would be due to the fact that, the direction of this force is perpendicular to the main axis of the lake, and the water load in the lake induces this force.

The tectonic movement for the faults is the main source to induce earthquakes in the area especially earthquakes with a magnitude ≥ 4. The role of reservoir water loading, as a supplementary source of earthquake events in the Aswan region, can not be neglected. Therefore, we can say that earthquake activity in the area was originated tectonically and the water factor works as an activating medium in triggering earthquakes.

From the study of the main seismic level through the Aswan area, it was found that a continuous decrease in the seismicity level happened between 1982 to 1992, although the water level (amount) in the Aswan Lake fluctuated during the same period. This means that water loading is only one factor from several factors, as an activating medium in triggering earthquakes. The common factors for all cases of induced seismicity seem to be the presence of specific geological conditions, the tectonic setting and water loading. And, also the decreasing of the seismicity with time may be related to the gradual stabilization of the area after the start up of seismicity in 1981 or receded the water from Marawa area.

References

Issawi B 1969: The geology of Kurkur-Dungul area. General Egyptian Organization for Geological Research and Mining; Geological Survey. Paper No. 46, Cairo, Egypt

Mohamed A S 1994: *Acta Geod. Geoph. Hung.*, 29, 5–18.

Mohamed A S 1995: *Acta Geod. Geoph. Hung.*, 30, 175–188.

Schandelmeier H, Klitzsch E, Hendriks F, Wycisk P 1987: *Berliner Geowiss. Abh.*, (A), 75, 5–24.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 53–71 (1997)

THE AQABA EARTQUAKE OF NOVEMBER 22, 1995 AND CO-SEISMIC DEFORMATION IN SINAI PENINSULA, DEDUCED FROM REPEATED GPS MEASUREMENTS

F KIMATA[1], A TEALEB[2], H MURAKAMI[3], N FURUKAWA[4], S MAHMOUD[2], H KHALIL[2], K O SAKR[2], A M HAMDY[2]

[Manuscript received December 10, 1996]

The new techniques of space geodesy, especially the Global Positioning System (GPS), allow the monitoring of ongoing crustal deformation and provide a powerful tool for the analysis of the regional seismotectonic regime and characterization of current strain accumulation and release.

Keywords: Aqaba earthquake; crustal deformation; GPS; seismicity; Sinai Peninsula

Recently, the northern region of the Red Sea, the southern part of the Gulf of Suez and the Gulf of Aqaba-Dead Sea Fault are tectonically seismoactive zones. A large earthquake occurred in March 31, 1969 at Shedwan island, to the south of the Gulf of Suez and north Red Sea, with a magnitude 6.9. The latest large earthquake occurred in November 22, 1995 in the Gulf of Aqaba with a magnitude 7.0.

In order to monitor the crustal deformation in Sinai Peninsula and around the regions of the northern Red Sea, Gulf of Suez and Gulf of Aqaba a GPS geodetic network consisting of 12 geodetic points was established in 1994. The initial geodetic measurements were performed in April 1994 using GPS receivers type Trimble 4000 SSEs. The geodetic measurements were repeated in April 1995 and April 1996 using the same type of GPS receivers. Twenty-eight baselines ranging from 46 to 166 km were determined from the six- to eight-hours observation using precise ephemerides.

A southwest displacement of 166 mm was recorded at Dahab geodetic point, 26 km southwest of the epicenter of the November earthquake, in the period April 1995–April 1996, and a northeast displacement of 32 mm was observed at Ras-Mohamed, 50 km southwest of the epicenter. These displacements show a left lateral fault and it is in agreement with the focal mechanism of the November earthquake (strike-slip mode with the tension axis of 65 degree from north), and aftershocks distribution.

Introduction

Territory of Egypt is not a major seismic zone, but earthquakes represent a significant hazard. In general, Egypt is a part of the north African Craton (Fig. 1), which occupies the northeastern corner of the African plate. The African plate is

[1]Research Center for Seismology and Volcanology, School of Science, Nogoya University, Japan
[2]National Research Institute of Astronomy and Geophysics, Helwan, Cairo, Egypt
[3]Makomani-Minomimachi, ku Sapporo, 005, Japan
[4]IISEE, Build. Res. Inst., Tyukuba-Shi, 305, Japan

1217-8977/97/$ 5.00

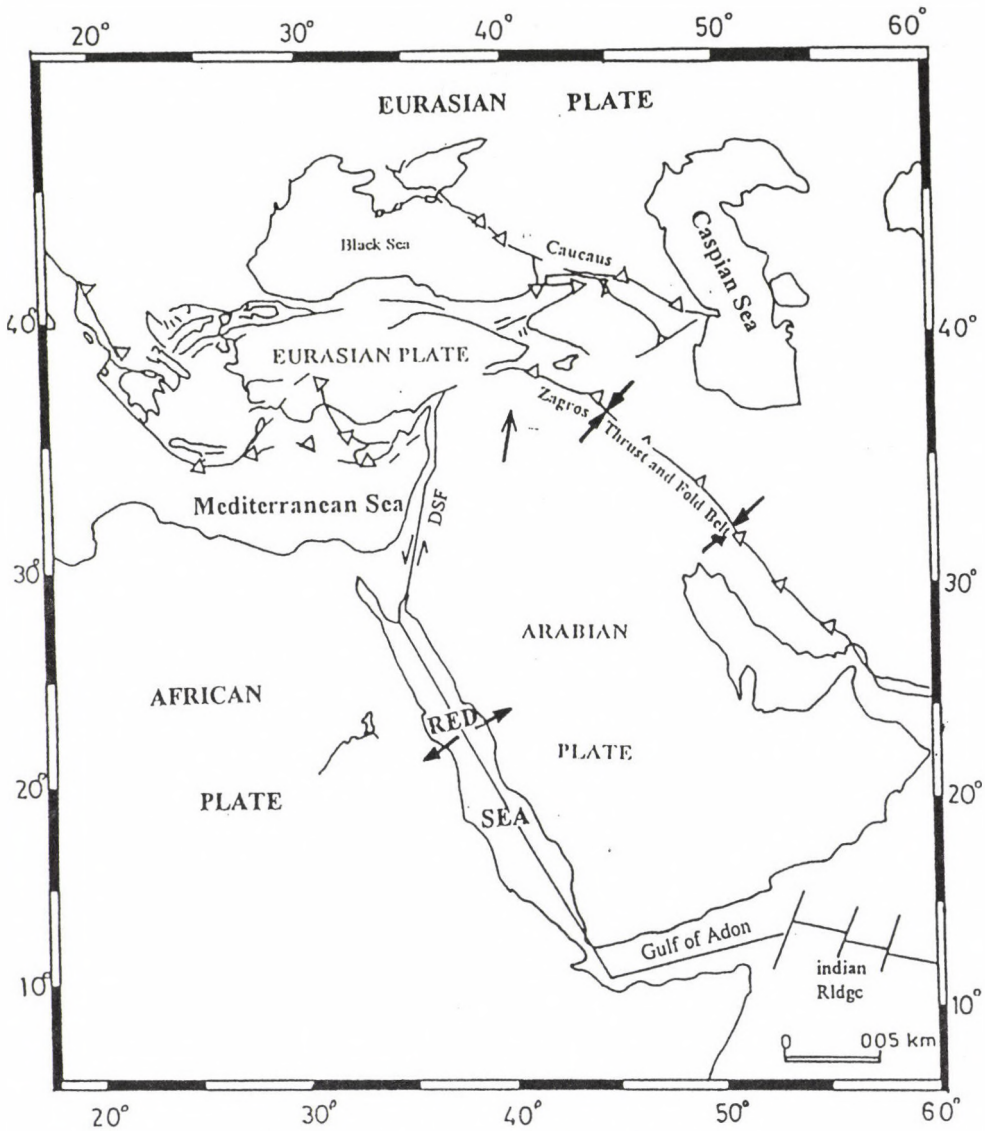


Fig. 1. Tectonic framework of the Middle East

surrounded from the east by the Arabian plate and from the north by the Anatolian/Eurasian plate. The Sinai subplate (Sinai Peninsula) occupies the area to the northeast corner of the African plate. The boundaries between these plates are in relative movement with respect to each other. The Arabian plate is moving northward, toward the Tourous-Zagros belt, causing the crustal spreading along the axis of the Red Sea and lateral slip along the Dead Sea transform zone producing the Gulf of Aqaba-Dead Sea-East Anatolian trend. The African plate is moving northward and colliding the Eurasian plate.

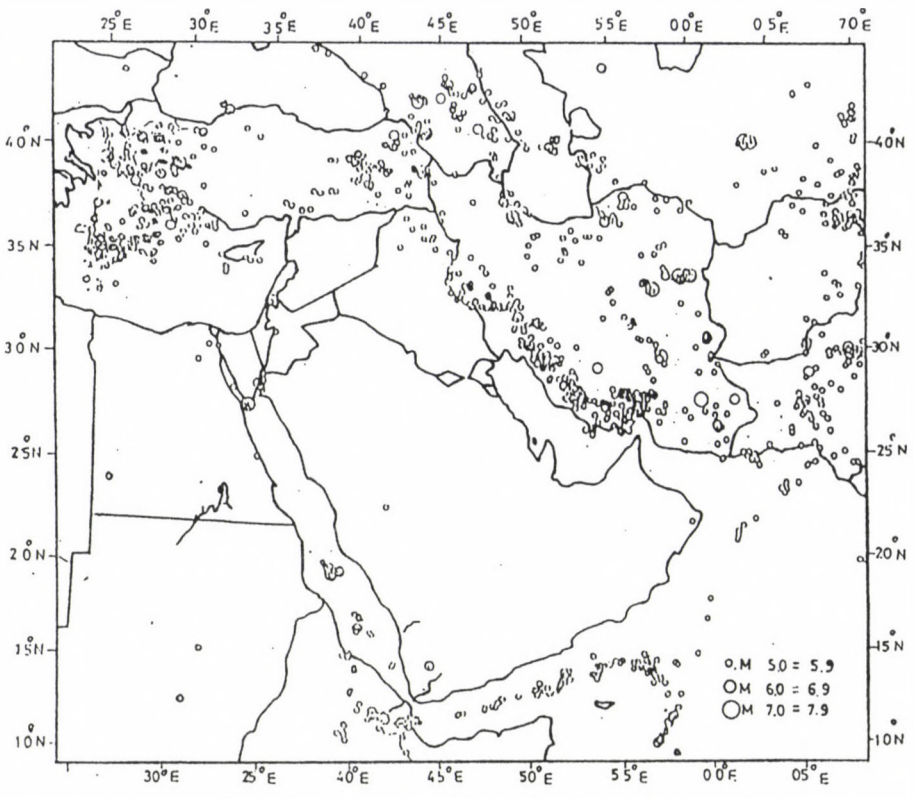


Fig. 2. Seismicity of the Middle East (1963–1994)

Most of the earthquake activities are associated with the tectonic features along the margins of these plates (Fig. 2). Most of the historical and recent seismic activity were recorded at the northern part of the Red Sea, southern part of the Gulf of Suez, the Gulf of Aqaba and Red Sea regions. Within the Egyptian territory (Fig. 3), seismic activities were recently recorded in the area southwest of Cairo, north Delta region, the area between Cairo-Ismaillia and Cairo-Suez roads, along the Red Sea coast, south Aswan and Gelf El-Kebir (Kebeasy 1990). These areas are strategic regions of the country.

Crustal deformation processes occur during the accumulation of the energy within the earth's crust. Its release produces earthquakes of certain magnitude therefore monitoring the rates of the crustal deformations are considered as valuable information about the seismic activity. The new techniques of space geodesy allow the monitoring of ongoing crustal deformation and provide a powerful tool for analysis of the regional seismotectonic regime and characterization of current strain

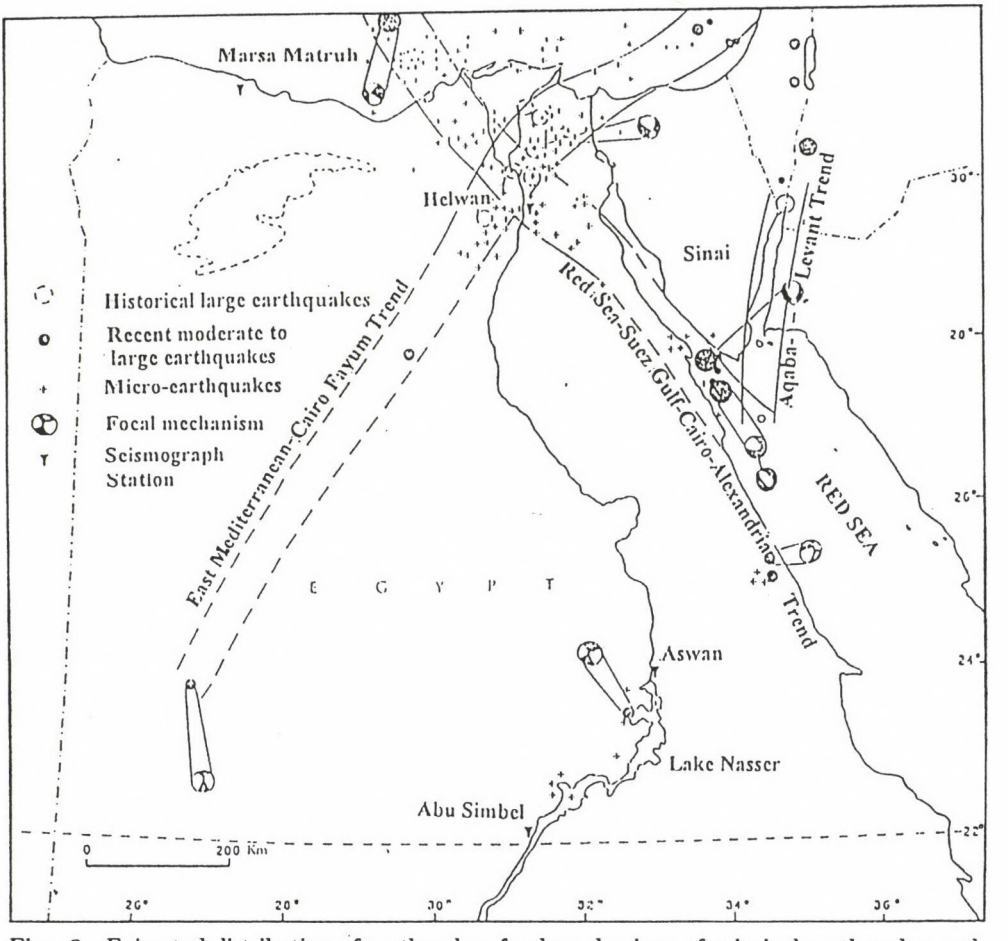


Fig. 3. Epicentral distribution of earthquakes, focal mechanisms of principal earthquakes and active seismic trends in Egypt (after Kebeasy 1990)

accumulation and release. The most applicable tool for monitoring crustal deformation along plate boundaries of regional and local scale is the Global Positioning System (GPS).

Both the techniques of terrestrial and space geodesy (GPS) were applied in several regions of Egypt (Tealeb 1994). The Egyptian program for monitoring crustal deformation was early initiated in 1983. The terrestrial geodetic techniques were applied in different local geodetic networks around active faults in the area to the northwest of lake Nasser (south Aswan city) and Aswan High Dam. The GPS technique was applied for measuring regional geodetic networks around the northern part of lake Nasser. Later on, GPS technique was used for measuring geodetic networks around greater Cairo and Gulf of Suez-Sinai Peninsula. It is proposed to establish a geodetic network at the Dead Sea Rift-Gulf of Aqaba-Red Sea system (Tealeb 1995).

A joint research project on the evaluation of seismic activities in the junction between the African and Arabian plates was started in 1993 (Kimata et al. 1995). The project was sponsored by the Japan International Cooperation Agency (JICA). The counterpart of the project is the National Research Institute of Astronomy and Geophysics (NRIAG), Helwan, Cairo, Egypt. The main scope of the project is the monitoring of the earthquake mechanism. A telemetered seismological network was established in 1994 around the Gulf of Suez and Sinai Peninsula (Figs 4 and 5). The data are recorded at the seismological center in Hurgada. The seismicity maps for the periods from August 1994 to December 1994 and from January 1995 to December 1995 are shown in Figs 4 and 5 respectively.

Under the same project Japan and Egypt were cooperating in the field of crustal deformation monitoring. A geodetic network for GPS measurements, consisting of 12 geodetic points, was established in 1994 around the Gulf of Suez-Sinai Peninsula and northern Red Sea. The initial geodetic measurements was performed in April 1994 using three GPS receivers type Trimble–4000 SSEs. The geodetic measurements were repeated in April 1995 and April 1996 using the same type of GPS receivers. After correcting, adjusting and analyzing the GPS data from the three epochs of measurements, the results are explained and discussed briefly in the present paper.

Regional geology, tectonics and seismicity of the region

Red Sea Rift System

The Red Sea rift system occupies an elongated escarpment bounded depression. This rift system was formed by separation of Arabia from Africa. The Red Sea depression was formed by the anticlockwise rotation of Arabia away from Africa around a pole of rotation in central or south-central Mediterranean (McKenzie 1970, Freund 1970).

At the northern end of the Red Sea the opening is split between the opening of the Gulf of Suez and the predominantly sinistral shear along the Gulf of Aqaba-Dead Sea rift system (Ben-Menahem et al. 1976). In general, it is accepted that the Gulf of Suez and Red Sea depressions were formed by the anticlockwise rotation of Arabia away from Africa (Cochran 1983). Both the Red Sea and Gulf of Suez rifting began in Oligocene time and developed in the Miocene, while the opening of the Gulf of Aqaba began in late Miocene and Pliocene times (Dietz and Hoden 1970), and resulted in lateral displacement along the Dead Sea-Aqaba rift. The northern Red Sea is almost entirely underlain by oceanic crust (McKenzie 1970, Girdler 1985).

Earthquake data indicate that the recent motion of the Gulf of Suez is sinistral oblique-slip. Motion along the Gulf of Aqaba on the other side of the Sinai subplate is approximately pure sinistral strike-slip (Ben-Menahem et al. 1979). The Sinai subplate moved relative to both Africa and Arabia.

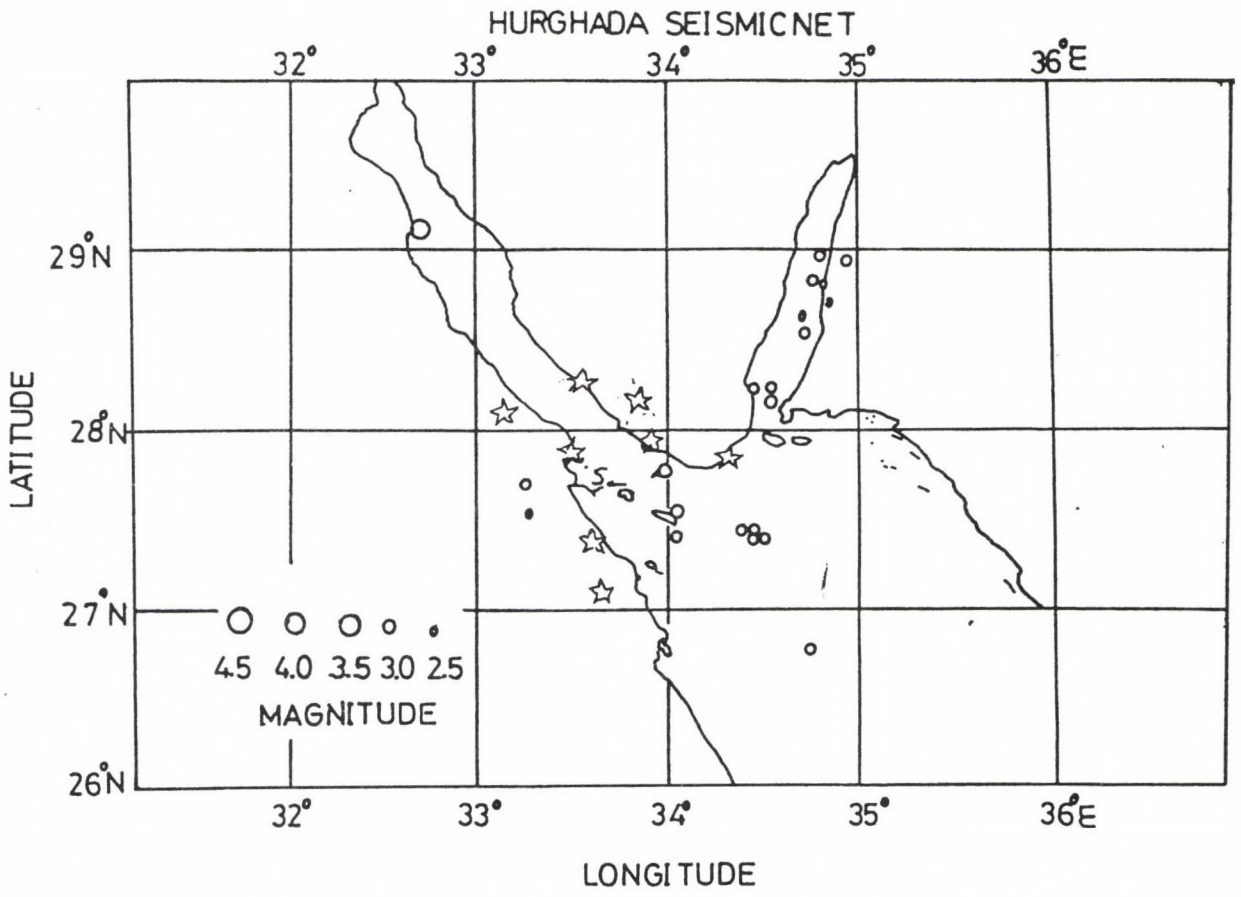


Fig. 4. Seismicity map of southern Sinai for the period from Aug. 1994 to Dec. 1994)

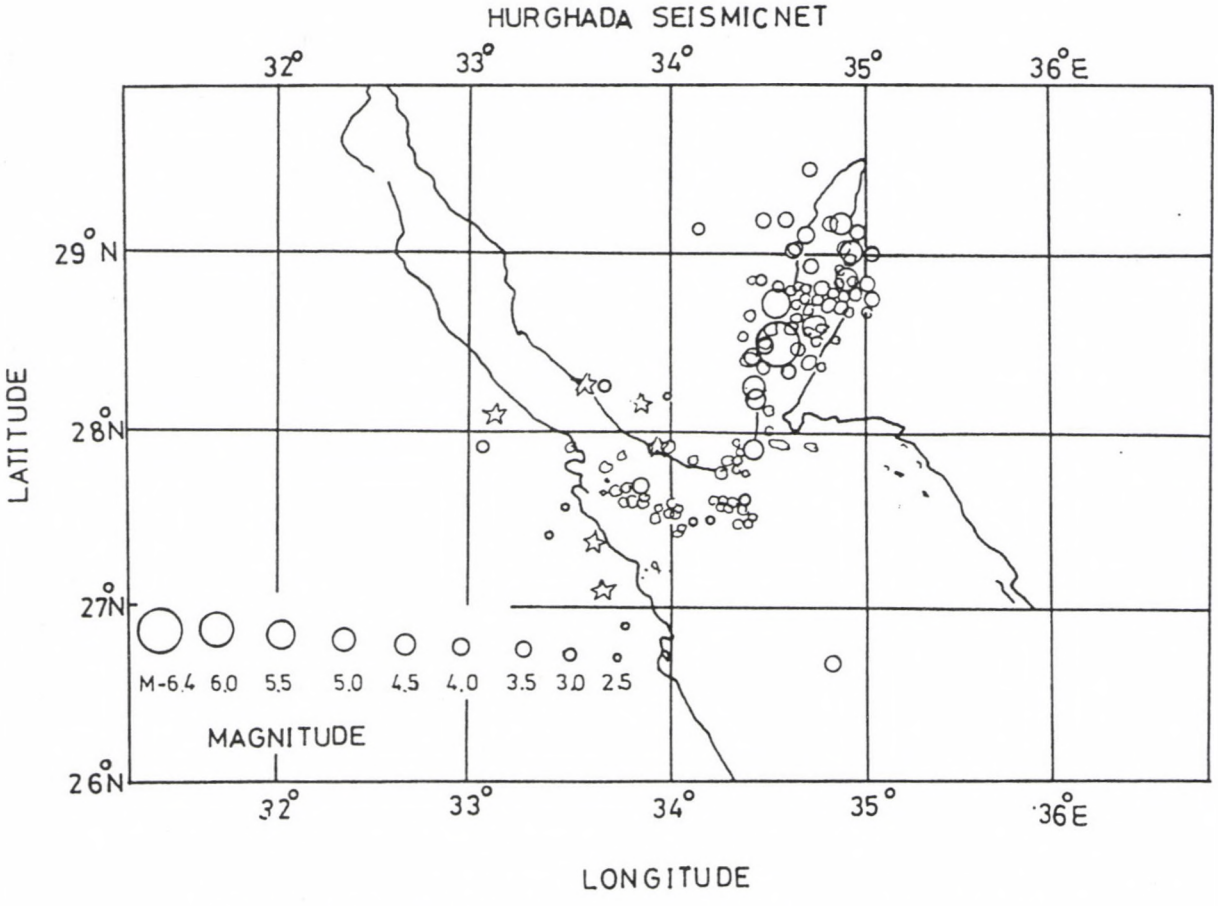


Fig. 5. Seismicity map of southern Sinai for the period from Aug. 1995 to Dec. 1995

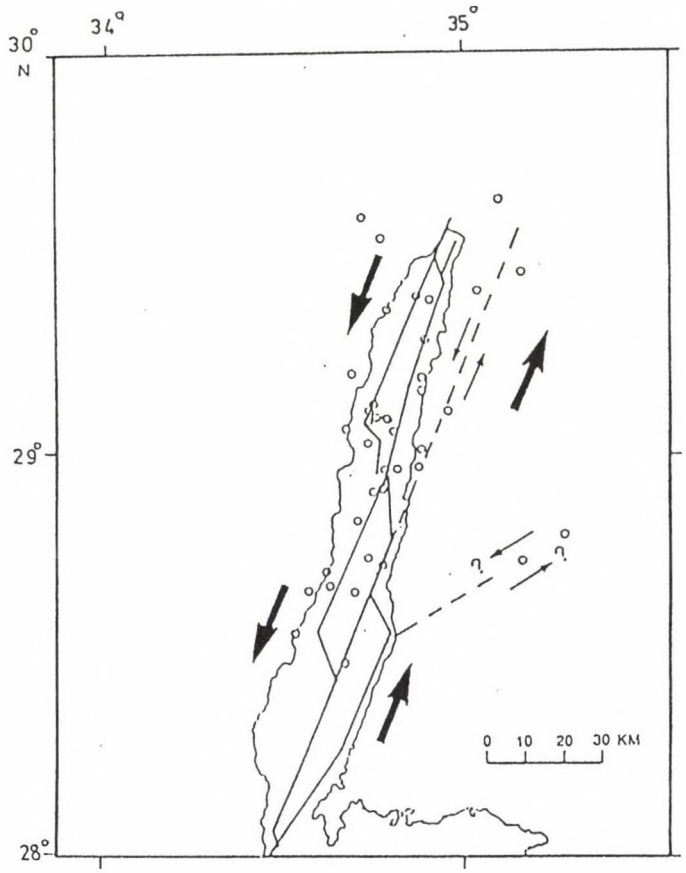


Fig. 6. Generalized model for structure of the Gulf Aqaba (after Ben-Avraham et al. 1979)

Gulf of Suez

The Gulf of Suez rift comprises a northwest trending intra-cratonic basin separated from the Red Sea by the Aqaba transform fault. The Gulf of Suez rift is bounded on the east by the Sinai Massif and on the west by the Red Sea Hills in the Eastern Desert. The Gulf of Suez is flat bottomed with a depth of 55–75 km, about 70–80 km wide, and becoming narrower towards the north (Girdler 1985). Structural sections show about 2–5 km of sediment (mainly Tertiary) overlaying the Pre-cambrian basement.

The Gulf of Suez is a tensional graben structure (Girdler 1985). Numerous faults are shown giving rise to several horsts and grabens within the downfaulted basement. There is no geological or geophysical evidence for an extension of the Gulf of Suez northwards under the Mediterranean (Girdler 1985).

Recently numerous moderate to large earthquakes occurred in the territory of the Gulf of Suez and northern part of the Red Sea. The largest earthquake occurred

Fig. 7. Ground fractures closed to Newabaa due to the November 22, 1995 eartquake

at Shedwan island, to the south of the Gulf of Suez, in March 31, 1969 with a magnitude 6.9.

Fig. 8. Damage of the Port of Newabaa due to the November 22, 1995 earthquake

Gulf of Aqaba

The Gulf of Aqaba is about 180 km long, 12–25 km wide and has irregular bottom topography of an average depth of 1.250 m. It is separated from the Red Sea by the sill of Tiron of about 250–260 m deep (outside this the bottom is at a depth more than 1000 m).

Along the beaches of the Gulf of Aqaba the mountains rise abruptly. The center of the Gulf is occupied by three deep and elongated basins separated by low sills (Hall and Ben-Avraham 1978).

The Gulf of Aqaba forms the northern segment of a major sinistral transform boundary of the Dead Sea rift. The structure of the Gulf of Aqaba is characterized by dominated en-echelon rhomb grabens (pull-apart basins) which delimit three elongated basins (Fig. 6) that strikes northwest (Ben-Avraham et al. 1979, Garfunkel 1981, Ben-Avraham 1985). The Dead Sea transform is a major boundary between Arabian and African plates. It connects the Red Sea spreading center in the south with the Tourus-Zagros collision belt in the north (Bahat et al. 1983, Ben-Avraham et al. 1979). The evolution of the Dead Sea rift system is comprised by an early stage of fracture initiation by extension, with a later development into a major left-lateral strike-slip fault (Bahat et al. 1983). The currently active faults in the Gulf of Aqaba area are two types: strike-slip and normal (Ben-Avraham 1985).

Archeological evidence indicates considerable earthquake activity and shear along

Fig. 9. Fracturing of a building closed to Newabaa city due the November 22, 1995 earthquake

the Dead Sea rift in historical times (Karcz et al. 1977). Evidence of seismic activity in the Gulf of Aqaba is documented from 747 A.D. to the present. Ben-Menahem (1981) indicated relatively lower seismic activity in the Gulf of Aqaba than in the Red Sea or Dead Sea regions. Evidence for three large earthquakes in northwestern Saudi Arabia, which were felt strongly around the Gulf of Aqaba in 873, 1068 and 1588 was documented by Ambrayses and Melville (1989). The earthquake in 1588 was strongly felt in Cairo and monastery at Saint Catherine collapsed.

In the Gulf of Aqaba region numerous moderate earthquakes with associated aftershock sequences and some earthquake swarms were experienced. An intensive earthquake swarm occurred during August 1993–February 1994 and was associated with more than 1200 events. The largest earthquake occurred in the Gulf of Aqaba in November 22, 1995 with a magnitude 7.0. Ground fractures, damage and collapse of buildings and liquefaction were observed in Aqaba (Jordan), Elat (Israel) and along the western coast of the Gulf of Aqaba in Sinai Peninsula. Figures 7, 8, 9 are examples from Sinai Peninsula close to the Gulf shore in Newabaa.

Sinai Peninsula

The Sinai Peninsula covers an area of about 61,000 km^2. It separated from mainland of Egypt by the Suez Canal and the Gulf of Suez.

Sinai Peninsula is divided into three different geomorphologic and structural units: northern-, central- and southern Sinai (Fig. 10). Northern Sinai is charac-

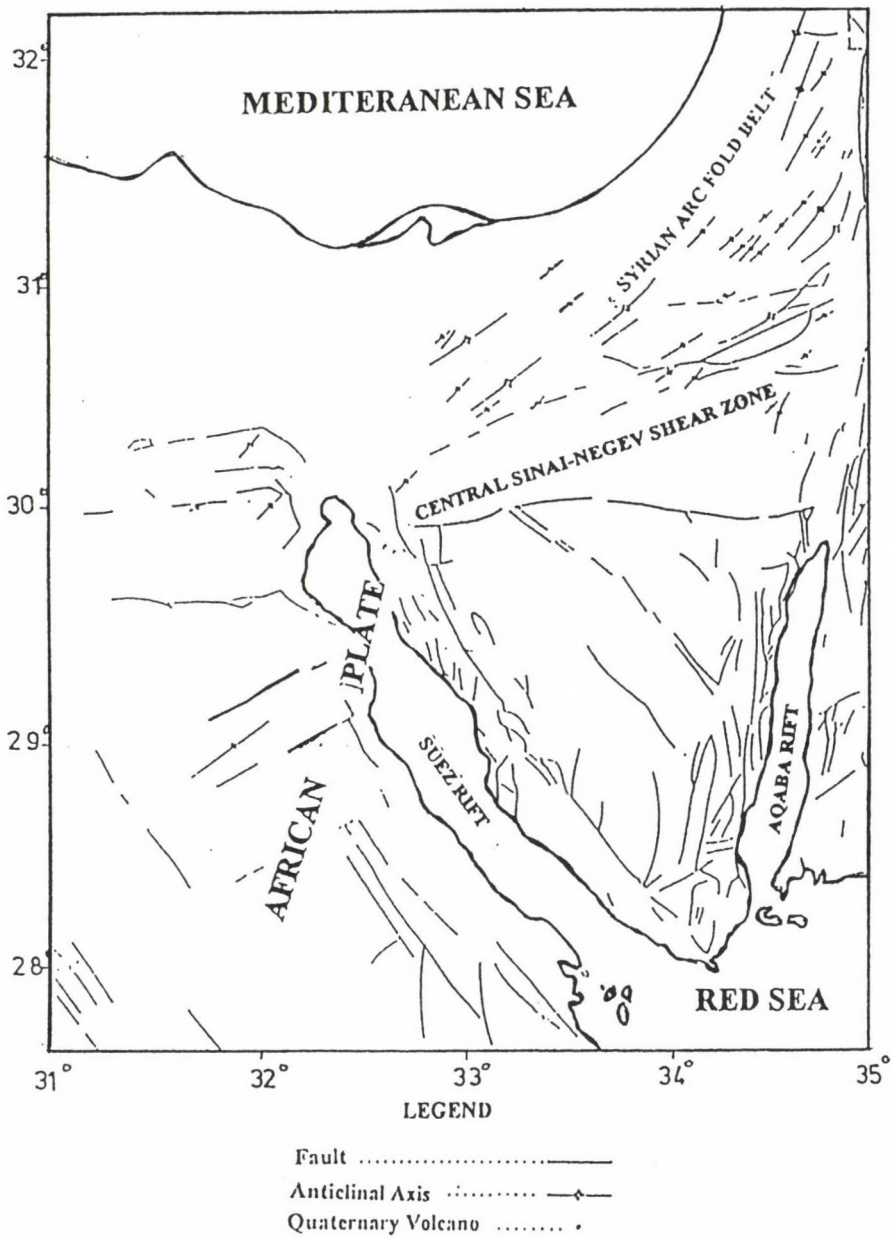


Fig. 10. Generalized tectonic map of Sinai

terized by three structural units (Shata 1956, El-Kerdany 1962). These units, from north to south are: the Mediterranean fore-shore area, north Sinai strongly-folded area (frontal folds) and north Sinai fractured area (Hinge belt). The main structural features of northern Sinai is the NE-SW elliptical anticlines and intervening syn-

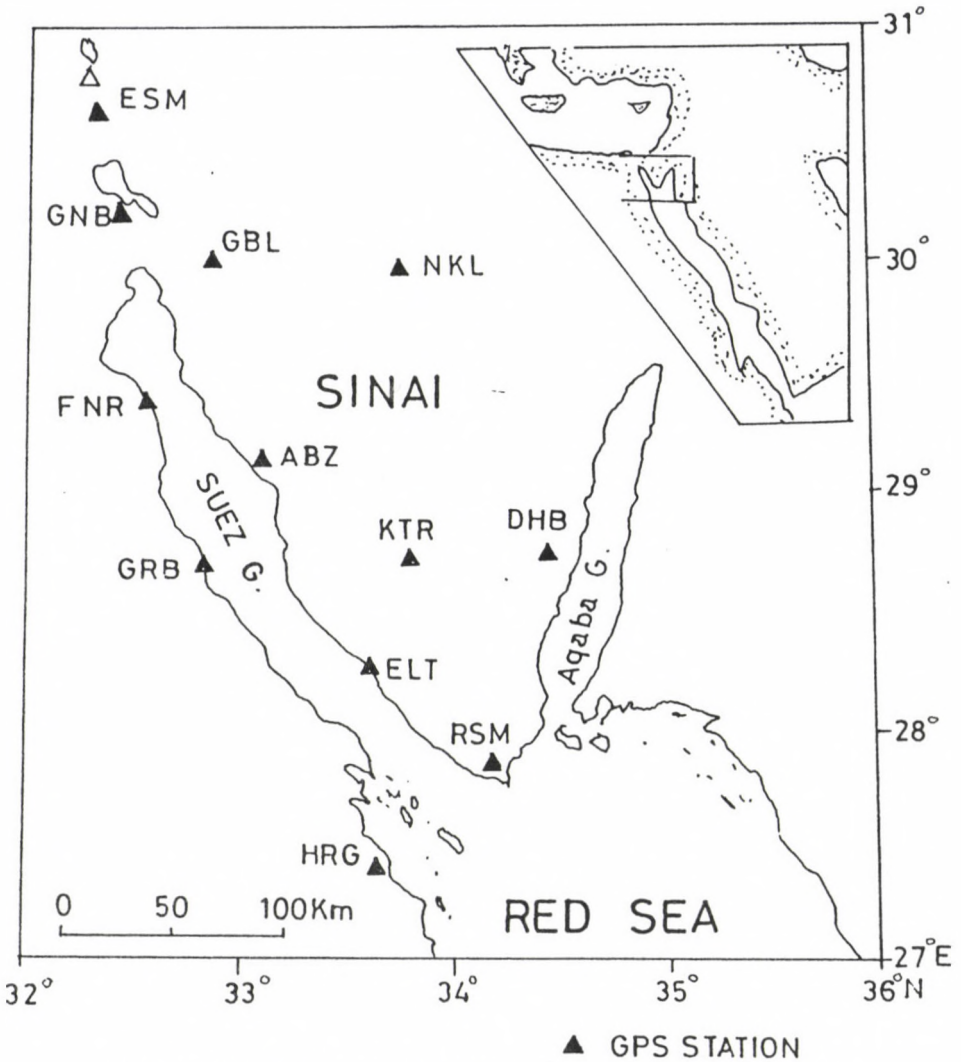


Fig. 11. GPS geodetic network in Sinai-Gulf of Suez-Northern Red Sea

clinal depressions in concordance with the regional trend of the Syrian Arc System (Shata 1956, Said 1962, El-Kerdany 1962, El-Shazly et al. 1974). The greater part of northern Sinai is drained by Wadi El-Arish which crosses a number of mountain blocks. In northern Sinai, the sedimentary section increases northwards. The total thickness exceeds 6 km close to the Mediterranean shore line.

The central part of Sinai is covered by subhorizontal strata overlapping the basement complex (Said 1962). The most pronounced part of central Sinai is El-Tih Plateau. It is bounded on its east, south and west sides by vertical scarps.

The southern part of Sinai consists of an intricate complex of high and very rugged igneous and metamorphic mountains (Said 1962, El-Shazly et al. 1974).

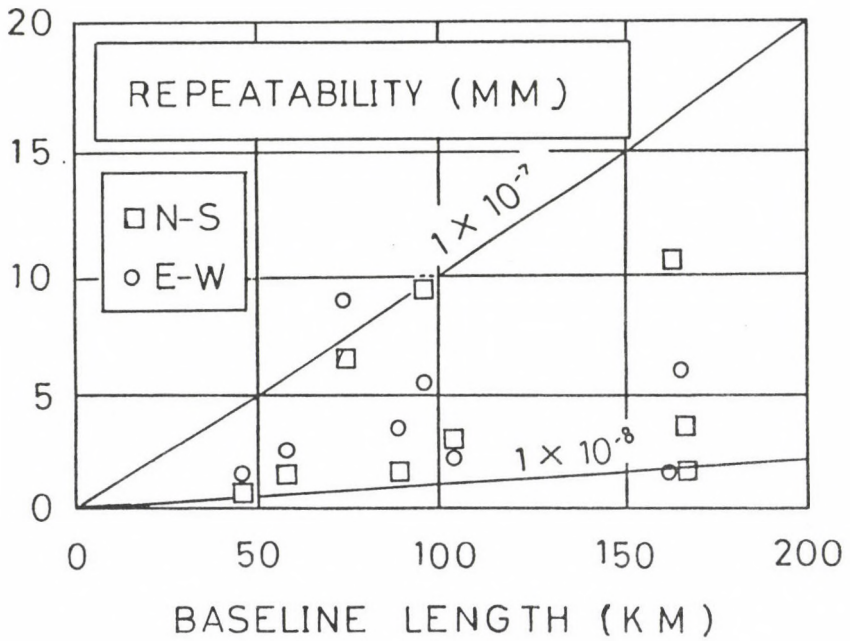


Fig. 12. Repeatability of horizontal vectors of baseline determination

The Moho depth in the central part of Sinai Peninsula is about 32 km. The Moho depth decreases toward the northern and southern regions of Sinai and reaches about 26 km near to the Mediterranean coast and at the Red Sea coast (Tealeb and Riad 1986).

GPS geodetic network and campaigns

In January 1994 a geodetic network of 12 GPS stations was established around the Gulf of Suez and in Sinai Peninsula (Fig. 11). The initial GPS measurements were performed in April 1994 using three Trimble 4000 SSEs receivers of Nagoya University, Japan. The GPS observations were carried out during a seven days campaign (two six-hour sessions a day). Twenty eight baselines — ranging from 46 to 166 km — were determined. Measurements were repeated for nine baselines and the deviation of baseline determination (horizontal vectors of N–S and E–W components) shows less than 10 mm (Fig. 12).

Analysis of the GPS data amd results

The GPS data were processed using precise ephemerids and the Trimble software. The coordinates at Catherine (KTR) site were determined relative to Helwan (HLW) site.

From the analysis the maximum displacement on the baselines were recorded close to the Gulf of Aqaba. The displacement at Ras-Mohamed (RSM) and Hurgada

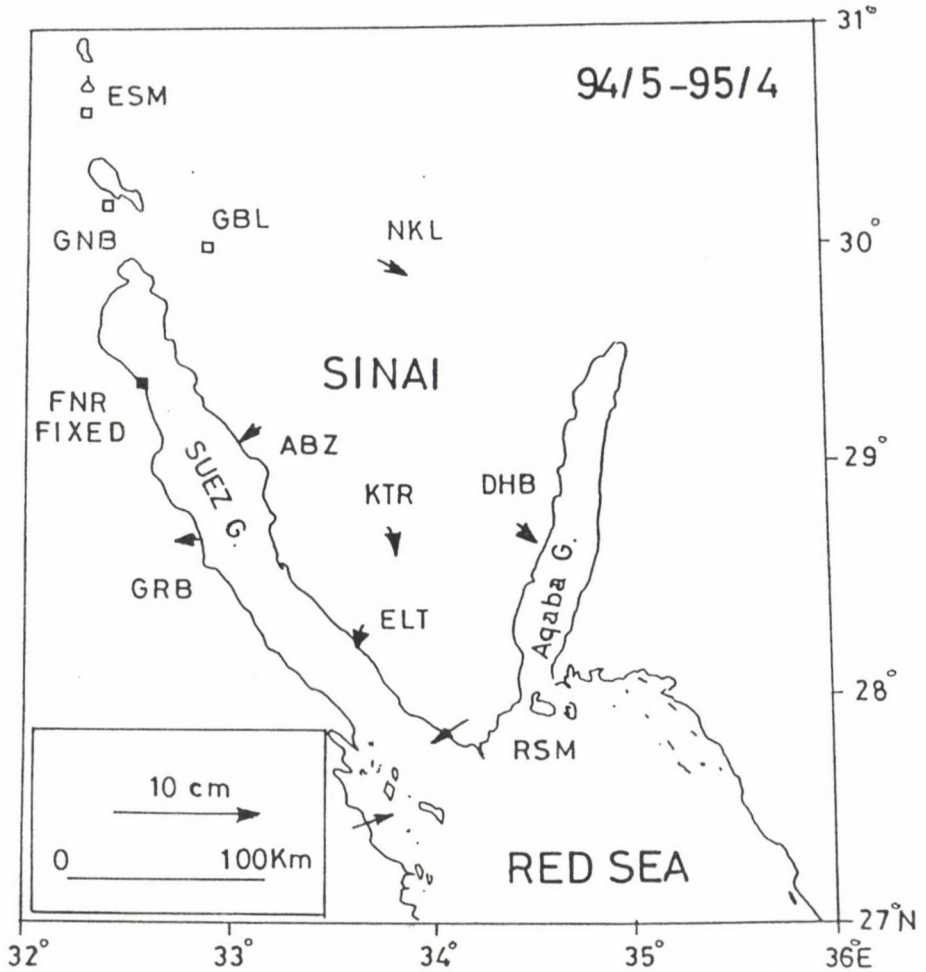


Fig. 13. Horizontal displacements in the period from April 1994 to April 1995

(HRG) are also remarkable (Fig. 13). The monitored displacement suggests a left lateral movement between these two stations.

After the November 22, 1995 earthquake in the Gulf of Aqaba region, the deformation parameters show also maximum displacement close to the western side of the Gulf of Aqaba (Fig. 14). A northeast (22° from the north) displacement of 32 mm was observed at Ras-Mohamed (RSM), 50 km southwest of the epicenter, and a southward (246° from north) displacement of 166 mm was recorded at Dahab (DHB), 26 km southwest of the epicenter of the November earthquake. The displacement vectors, for the periods 1994–1995 and 1995–1996, at the GPS stations are given in Table I. The results for the period 1995–1996 are the co-seismic displacement according to the Aqaba earthquake of November 22, 1995. These displacements suggest a left lateral movement in agreement with the focal mechanism

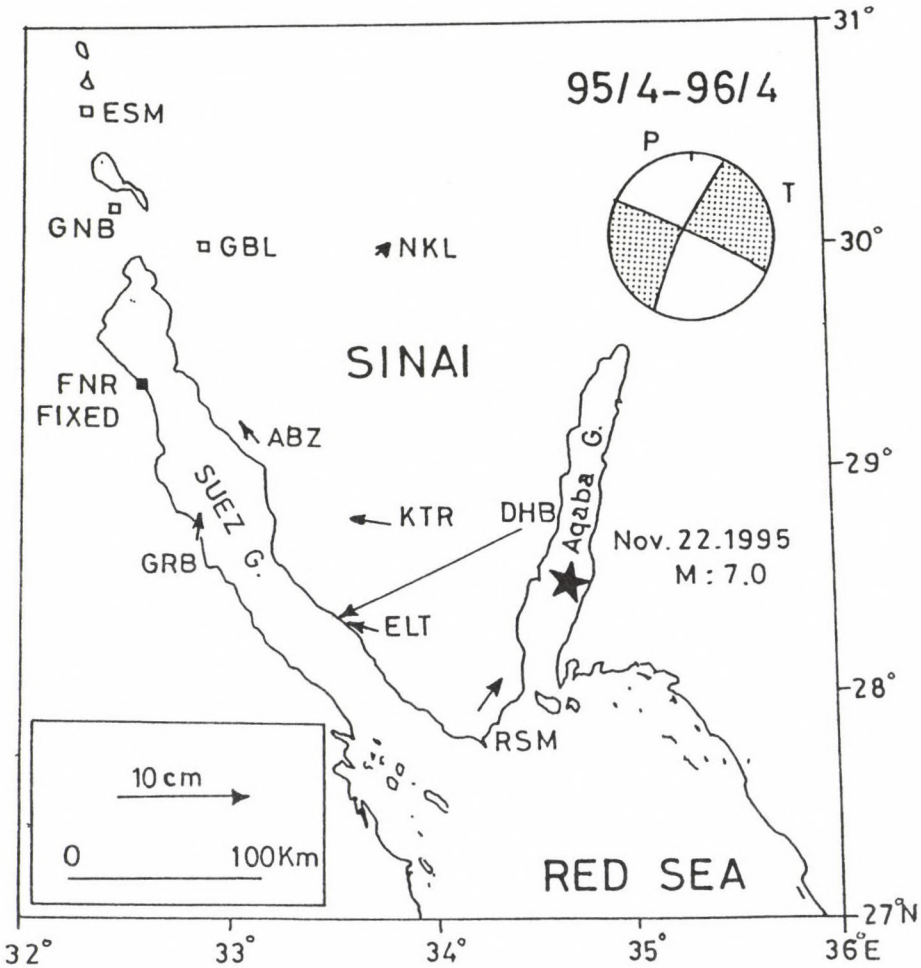


Fig. 14. Horizontal displacements in the period from April 1995 to April 1996

given by the US Geological Survey (strike-slip mode with the tension axis of 65 degree from north) for the November 22, 1995 earthquake.

Conclusions and recommendations

The GPS technique was used for monitoring the crustal deformations in Sinai Peninsula. GPS measurements were performed in 1994, 1995 and 1996. In the meantime the seismicity of the whole area was recorded in Hurgada earthquake center during the period 1994–1996. Also, the visible surface deformations resulted from the Aqaba earthquake of November 22, 1995 were studied during field trips in the area around the Gulf of Aqaba. According to the collected tectonic, surface, seismological and geodetic data the following conclusions and recommendations are highly considered.

Table I. GPS stations in Sinai-Gulf of Suez-Red Sea network and the horizontal displacement in the periods 1994–1995 and 1995–1996

No.	GPS station	Code	Station coordinate		Hight	Period							
			Longitude (N)	Latitude (E)		1994–1995				1995–1996			
			° ' "	° ' "	m	E–W mm	N–S mm	Vector (clockwise) mm	o	E–W mm	N–S mm	Vector (clockwise) mm	o
1	Hurgada	HRG	27 22 27	33 37 37	59	25	5	25	(79)	−8	30	31	(345)
2	Gharib	GRB	28 41 49	32 49 16	32	−12	−2	12	(260)	3	30	30	(5)
3	Fanar-Abu-El-Darag	FNR	29 22 43	32 33 57	40	-0	0	0	(0)	0	0	0	(0)
4	Abu-Zenima	ABZ	29 08 29	33 06 09	142	−1	−1	1	(225)	−11	−16	19	(224)
5	Catherine	KTR	28 43 21	33 47 25	913	1	−17	17	(176)	−34	7	35	(281)
6	El-Tur	ELT	28 16 10	33 35 47	128	−2	−12	12	(189)	−14	7	16	(296)
7	Ras-Mohamed	RSM	27 50 47	34 11 02	258	−30	−14	33	(245)	12	30	32	(22)
8	Dahab	DHB	28 44 53	34 28 02	556	8	−11	14	(144)	−151	−68	166	(246)
9	Nakhl	NKL	29 57 51	33 44 35	465	29	−17	34	(120)	14	15	20	(43)
10	Gebel-El-Rahma	GBL	29 59 25	32 52 10	320								
11	Esmaelia	ESM	30 36 57	32 19 08	44								
12	Geneiva	GNB	30 10 31	32 25 57	62								

1. From the historical documentations and recent observations, the area of the northern part of the Red Sea, the southern part of the Gulf of Suez and the Gulf of Aqaba-Dead Sea transform are considered as a seismoactive region.

2. The analysis of the GPS data from the geodetic network around the Gulf of Suez-Sinai Peninsula for the periods 1994, 1995 and 1996 reveals a remarkable displacement close to the Gulf of Aqaba. For the period 1995–1996, co-seismic displacement of 166 mm was found at Dahab, 26 km southwest of the epicenter of the November earthquake in the Gulf of Aqaba. This displacement suggests a left lateral movement in agreement with the focal mechanism of the main earthquake of November 22, 1995.

3. As the Gulf of Aqaba consists of three basins which resulted from en-echelon rhomb-shaped grabens produced by strike slip, the location of the earthquake epicenter and that of the monitored co-seismic displacement as well as the focal mechanism of the November 22, 1995 eartquake suggest a left lateral fault. The November 22, 1995 earthquake in the Aqaba region is supposed to have occurred in the NW fault boarding the central and southern basins in the Gulf of Aqaba.

4. The historical and recent seismicity of the Gulf of Aqaba as well as the present results led us to accept the suggestions of the establishment of a geodetic network in the Gulf of Aqaba for monitoring crustal deformations and modelling its present tectonics and a recent geodynamics.

5. The countries surrounding the Gulf of Aqaba and northern Red Sea region are invited to cooperate in a near future program to study the geodynamics of the Arabian and African plates and the relative movements of the Sinai subplate.

References

Ambrayses N N, Melville C P 1989: *Bull. Seis. Soc. Am.*, 79, 1279–1281.
Bahat, Dov, Rabinovich A 1983: *Journal of Geology*, 91, 317–332.
Ben-Avraham Z 1985: *J. Geoph. Res.*, 90, 703–726.
Ben-Avraham Z, Almagor G, Garfunkel Z 1979: *Sad. Geol.*, 23, 239–267.
Ben-Menahem A, 1981: *Bull. Geofis. Teorica o Applicata*, 23, 349–354.
Ben-Menahem A, Nur A, Vered M 1979: *Phys. Earth Planet. Int.*, 21, 1–50.
Cochran J R 1983: *Bull. Am. Assoc. Petrol. Geol.*, 67, 41–69.
Dietz R S, Hoden J C 1970: *J. Geophys.*, 75, 4939–4955.
El-Kerdany M I 1962: General study of north Sinai, Egypt. G.P.C., 210, Egypt
El-Shazly E M, Abdel Hady M A, El-Kassas I A, El-Shazly M M 1974: Geology of Sinai Peninsula from ERTS-1 satellite images. Rem. Sens. Center, Cairo, Egypt
Freund R 1970: *Nature*, 228, 453.
Garfunkel Z 1981: *Tectonophysics*, 80, 81–108.
Girdler R W 1985: *Tectonophysics*, 116, 109–122.
Hall M I K, Ben-Avraham Z 1978: Bathymetric chart of the Gulf of Elat. Geol Survey, Israel

Karcz I, Kafri U, Meshel Z 1977: *Natura*, 269, 234–235.
Kebeasy R M 1990: In: The Geology of Egypt, R Said ed., A A Balkema, Rotterdam, Brookfield, 51–59.
Kimata F, Tealeb A, Ibrahim E M, Yokoyama I, Furukawa N 1995: GPS measurements at Sinai Peninsula-Red Sea region in Egypt (preliminary report). Nagoya University, Japan
McKenzie D P 1970: *Nature*, 239–248.
Said R 1962: The geology of Egypt. Elsevier
Shata A 1956: *Bull. Inst. Desert Egypt*, 6, 117–157.
Tealeb A 1994: In: Proc. 1st Turkish Int. Symp. on deform., 872–898.
Tealeb A 1995: *Bull. Nat. Res. Inst. of Astr. and Geophys.*, V.XI, Ser. B, 139–147.
Tealeb A, Riad S 1986: In: EGS Proc. of the 5th Ann. Meeting, 18–49.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 73–86 (1997)

FOCAL DEPTH ESTIMATION FROM INTENSITY DISTRIBUTION

T Zsíros[1]

[Manuscript received January 7, 1997]

For the scarcity of data, isoseismal maps frequently cannot be drawn based on the intensity distributions of historic earthquakes. Therefore, the traditional focal depth estimations with isoseismals are out of question for such events. The method presented in this paper, instead of isoseismals, uses the discrete intensity distributions of earthquakes. The focal depth is determined from the observed intensities by the Kövesligethy model using a nonlinear least-squares method. The method is tested for 50 earthquakes and though the results are not satisfactory for intermediate Háromszék-Vrancea events and may be questionable to some crustal earthquakes, too, the method has the capability of depth estimation of historic earthquakes when isoseismals cannot be drawn. Even, when isoseismals are compiled, this method means a more objective solution, because of the subjective nature of isoseismals.

Keywords: earthquake; focal depth; isoseismal; intensity; Kövesligethy

Introduction

There are great efforts to compile macroseismic field effects of significant historic earthquakes, which are of crucial importance in hazard estimation. The intensity distribution of many historic events, however, consists of only some observations not enough for the compilation of isoseismal maps. Still, in common practice the isoseismal maps have been used for the estimation of focal depth of earthquakes. From these facts, it is understandable that the depth estimation is very uncertain or does not exist at all for most of the historic events. The method presented below does not need the compilation of isoseismals for depth estimation.

Method

The traditional intensity attenuation model is described by Kövesligethy (1906)

$$I_o - I_k = 3 \cdot \log\left(\frac{D_k}{h}\right) + 3 \cdot \alpha \cdot \log(e) \cdot (D_k - h)\,, \tag{1}$$

where

I_o – epicentral intensity
I_k – intensity at K_k hypocentral distance
$D_k^2 = r_k^2 + h^2$
r_k – focal depth (km)
α – absorption coefficient (km^{-1})
$\log(e) = 0.4343$.

[1]Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences, H-1112 Budapest, Meredek u. 18, Hungary, E-mail: tibor@sas.seismology.hu

1217-8977/97/$ 5.00

This model assumes that the energy of seismic waves declines owing to the geometrical spreading and to the absorption of the traversed media. In the original form the Eq. (1) referred to isoseismals, but the individual intensity data can also be used in the above formula. Based on this idea the unknown h and α parameters in Eq. (1) can be estimated using a nonlinear least-squares method (Bevington 1969). This applies the conjugate gradients method (Press et al. 1989) for function minimization and in addition to the estimated parameters it also calculates the variance-covariance matrix of the solution. In the actual calculations BECS Fortran computer program (Tarcsai 1991) was used.

From the observed intensities of a given earthquake the maximum felt intensity is chosen as the epicentral intensity I_o if it is not given in the source literature, otherwise I_o is adopted from the study. The mean values of the co-ordinates of maximum felt (epicentral) intensities are chosen as the co-ordinated of the epicentre. Then the epicentral distance R can be calculated from the co-ordinates of the epicentre and the location with felt intensity in question.

A similar method has already been used for French earthquakes (Levret et al. 1994) showing the estimated attenuation curves versus all available local intensities and the average distances by degree and half-degree of intensity for one selected event, as an example.

Results and conclusions

For testing the method 50 earthquakes were arbitrarily selected including intermediate Háromszék-Vrancea events. Both strong and weak earthquakes are used, so the number of observed intensities (input data) vary in wide range, from a few to over one thousand. The results of macroseismic depth estimations of the selected earthquakes are summarized in Table I. In these results there are both very shallow, between 1–5 km (e.g. Palombara Sabina 1901, Belice 1968, West Wyalong 1982) and deeper, between 30–60 km (e.g. New Madrid 1811, Háromszék-Vrancea 1790, 1838, 1940, Alaska 1979, Saguenay 1988) focal depths. The larger Háromszék-Vrancea earthquakes however are known as intermediate events but our estimates are definitely lower than the usual 80–120 km depth values. The negative absorption results ($\alpha < 0$), furthermore, clearly indicate that the observed intensities are determined not only by the geometrical spreading of the earthquake energy and the absorption of the media as the attenuation model (Eq. 1) assumes. For depth estimation we would need intensity observations without any site effects, but there is no such method which can separate the influence of geometrical and absorption effects, on the one hand, and site effects, on the other, on the observed intensities. In this way we have to use all the data oberved. The Kövesligethy model also assumes a circular intensity attenuation but in reality there are significantly non-circular macroseismic fields (e.g. Háromszék-Vrancea events). The unexpected results obtained for some events may be accounted for largely by these difficulties. The fit between the obtained intensity attenuation curves and the input (observed intensity) data are shown for every selected event in Figs 1–7.

Though the instrumental and macroseismic focuses do not necessarily coincide,

Table I.

Location/Region	Date	Epicentre		I_o	h (km)	α (km^{-1})	n	S
Villach	1348 Jan. 25	46.6N	13.8E	11	8.3±1.7	−0.0001±.0009	45	1
Basel[1]	1356 Oct. 18	47.5N	07.6E	10	5.7±0.9	−0.006±.0009	32	2
Basel[2]	1356 Oct 18	47.5N	07.6E	9	11.5±1.6	−0.006±.0008	32	2
Catalonia	1428 Feb. 02	42.4N	02.2E	9–10	8.1±1.0	−0.001±.002	39	3
Zólyomlipcse	1443 Jun. 05	48.6N	18.9E	8	13.9±4.6	−0.004±.002	13	4
Modena	1501 Jun. 05	44.5N	10.9E	9	4.8±1.6	0.007±.008	13	1
Ferrara	1570 Nov. 17	44.8N	11.6E	8	5.4±0.8	−0.008±.003	38	1
Wien	1590 Sep. 15	48.2N	16.1E	9	8.0±0.8	−0.002±.0005	59	5
Rimini	1672 Apr. 14	44.1N	12.6E	8	5.4±2.7	−0.0006±.008	16	1
Cotignola	1688 Apr. 11	44.4N	11.9E	9	4.8±1.6	0.01±.009	17	1
Irpinia	1702 Mar. 14	41.1N	14.8E	10	12.0±6.1	0.008±.02	17	1
Foggia	1731 Mar. 20	41.5N	15.6E	10	8.1±2.7	0.003±.003	20	1
Irpinia	1732 Nov. 29	41.0N	15.1E	10	9.5±0.7	0.004±.002	108	1
Háromszék-Vrancea	1790 Apr. 06	45.7N	26.6E	8	46.7±13.7	0.002±.001	23	6
New Madrid	1811 Dec. 16	35.4N	90.1W	10–11	32.3±6.9	0.0009±.0003	39	7
New Madrid	1812 Jan. 23	36.6N	89.6W	9	36.6±31.4	0.0004±.0008	11	7
New Madrid	1812 Febr. 07	36.6N	89.6W	10–11	34.9±10.6	0.0008±.0003	11	7
Háromszék-Vrancea	1838 Jan. 23	45.6N	27.0E	9	50.6±11.8	0.0009±.001	36	6
Marlborough	1848 Oct. 16	41.4S	173.4E	10	21.4±8.0	0.002±.002	17	8
Gayndah	1883 Aug. 28	25.7S	151.6E	7	9.4±1.6	−0.004±.0009	49	9
Berridale	1885 Jan. 21	36.3S	149.2E	5	6.8±1.7	−0.01±.004	8	9
Yass	1886 Nov. 29	34.9S	148.8E	6–7	9.7±1.2	−0.005±.001	41	9
Palombara Sabina	1901 Apr. 24	42.1N	12.7E	8	1.3±0.3	0.008±.007	27	1
Háromszék-Vrancea	1938 Jul. 13	46.4N	26.8E	5–6	23.4±4.5	−0.003±.0007	28	6
Háromszék-Vrancea	1940 Oct. 22	46.4N	27.7E	7–8	58.3±9.2	0.002±.0005	135	6

[1]intensity estimated by Bureau de Recherches Géologiques et Minieres in Orleans
[2]intensity estimated by Schweizerischer Erdbebendienst in Zürich

Table I. (contd.)

Location/Region	Date	Epicentre		I_o	h (km)	α (km^{-1})	n	S
Háromszék-Vrancea	1955 May. 01	46.0N	27.4E	5	19.4±5.2	−0.005±.002	25	6
Irpinia	1962 Aug. 21	41.2N	14.9E	9	9.6±1.0	0.01±.001	192	1
Háromszék-Vrancea	1965 Jan. 10	45.8N	26.6E	6	12.6±1.3	−0.006±.0006	190	6
Belice	1968 Jan. 15	37.7N	13.0E	10	3.5±0.2	0.006±.0009	154	1
Alaska	1979 Feb. 28	60.0N	141.7W	7	43.0±8.7	−0.0001±.0005	55	10
Imperial Valley	1979 Oct. 15	32.8N	115.5W	7	10.5±1.1	−0.003±.0005	208	11
Irpinia	1980 Nov. 23	40.8N	15.3E	10	3.6±0.1	0.002±.0001	1131	1
Perugino	1982 Oct. 17	43.2N	12.6E	7	4.8±0.7	0.002±.004	44	12
West Wyalong	1982 Nov. 26	33.9S	147.2E	6	1.3±0.2	−0.06±.01	13	9
Fenouilledes	1984 Dec. 02	42.5N	02.3E	5	1.9±0.5	−0.05±.04	15	13
Provence	1985 May 28	43.2N	05.3E	5	2.1±0.9	0.03±.08	13	13
Bearn	1985 Aug. 27	43.1N	00.6W	4	2.9±0.6	−0.07±.01	11	13
Friuli	1986 Jan. 12	46.1N	13.4E	5	4.3±1.1	−0.04±.02	10	14
Ferrarese	1986 Dec. 06	44.9N	11.4E	6	4.0±0.9	0.01±.02	14	14
Colli Albani	1987 Apr. 11	41.7N	12.7E	6	4.5±0.6	−0.008±.004	52	15
Regio nell'Emilia	1987 Apr. 24	44.8N	10.8E	6	6.4±0.9	−0.01±.003	43	15
Modenese	1987 May 08	44.8N	11.1E	6	2.6±0.5	−0.02±.01	21	15
Lago di Garda	1987 May 24	45.7N	10.7E	6	3.6±0.6	−0.02±.005	22	15
Montefeltro	1987 Jul. 05	43.8N	12.2E	6	4.1±0.4	−0.02±.003	79	15
Montefeltro	1987 Jul. 08	43.7N	12.2E	5–6	2.9±0.8	−0.04±.01	13	15
Porto San Giorgio	1987 Sep. 04	43.1N	13.8E	6	4.2±0.4	−0.01±.003	67	15
Friuli	1988 Feb. 01	46.2N	13.0E	6–7	2.6±0.4	−0.02±.005	103	16
Saguenay	1988 Nov. 25	47.9N	70.6W	7	58.9±4.4	−0.0001±.0001	250	17
Uluru	1989 May 28	25.3S	130.7E	7	23.1±10.3	0.003±.003	11	18
Loma Prieta	1989 Oct. 18	37.1N	122.0W	8	14.9±0.8	0.002±.0003	514	19

h – focal depth
α – absorption coefficient
n – number of data used
S – source code of intensities

Table II.

Date	Time (UT)	Focal depth (km)		Magnitude
		Network determination	Macroseismic	
1968 Jan. 15	02:01	3 (ISC), 3 (MOS), 3.8±0.8 (ISC)	3.5±0.2	M5.4 (ISC)
1979 Feb. 28	21:27	15 (NEIC), 18 (OTT), 17±2.5 (ISC), 22±1.8 (ISC)	43.0±8.7	M_b6.2 (ISC)
1979 Oct. 15	23:17	12 (NEIC), 30±4 (ISC)	10.5±1.1	M_b5.6 (ISC)
1980 Nov. 23	18:35	18 (ROM)	3.6±0.1	M_b6.0 (ISC)
1982 Oct. 17	06:45	14 (NEIC), 9 (ROM), 6±5.8 (ISC)	4.8±0.7	M_b4.6 (ISC)
1982 Nov. 26	00:11	4 (AUST), 4 (ISC)	1.3±0.2	M_l4.6 (AUST)
1984 Dec. 02	21:49	7 (LDG)	1.9±0.5	M_l3.9 (LDG)
1985 May 28	00:50	2 (LDG), 3±5.2 (ISC)	2.1±0.9	M_l3.2 (LDG)
1985 Aug. 27	23:23	2 (LDG), 11±9.4 (ISC)	2.9±0.6	M_l2.9 (LDG)
1986 Jan. 12	21:04	7 (ROM), 7 (ISC)	4.3±1.1	M_l3.5 (LDG)
1986 Dec. 06	17:07	2 (LDG), 15 (TRI)	4.0±0.9	M_l4.1 (TRI)
1987 Apr. 11	02:26	18 (NEIC), 8±9.0 (ISC)	4.5±0.6	M_g3.7 (ISC)
1987 Apr. 24	02:30	13 (TRI), 13 (CSEM), 32 (NEIC), 20±6.1 (ISC)	6.4±0.9	M_b4.4 (ISC)
1987 May 08	11:10	34 (TRI), 3±5.3 (ISC)	2.6±0.5	M_l4.1 (LDG)
1987 May 24	10:23	21 (NEIC), 4 (TRI), 1±4.2 (ISC)	3.6±0.6	M_b4.5 (ISC)
1987 Jul. 05	13:12	11 (NEIC), 12±7.3	4.1±0.4	M_b4.7 (ISC)
1987 Jul. 08	04:29	12 (NEIC), 12 (ISC)	2.9±0.8	M_l3.7 (LDG)
1987 Sep. 04	16:43	22 (TTG), 19 (NEIC), 11±5.6 (ISC)	4.2±0.4	M_b5.0 (ISC)
1988 Feb. 01	14:21	8 (NEIC), 3±3.0 (ISC)	2.6±0.4	M_b4.5 (ISC)
1988 Nov. 25	23:46	29 (NEIC), 20 (NEIC), 28 (OTT), 28 (ISC), 24 (ISC), 28 (HRVD)	58.9±4.4	M_b5.9 (ISC)
1989 May 28	02:55	30±1.3 (ISC)	23.1±10.3	M_b5.4 (ISC)
1989 Oct. 18	00:04	19 (NEIC), 8 (NEIC), 8±3 (ISC), 17±2 (ISC), 19 (HRVD)	14.9±0.8	M_b6.2 (ISC)

the comparison of them may be interesting. The focal depths of the 22 latest events (since 1968, when instrumental results became more available) with magnitude of 2.9 to 6.2 reported by different seismological agencies are summarised in Table II together with our macroseismic estimates. The differences between the instrumental and macroseismic depth values are sometimes rather large, but it is to be noted that the instrumental values themselves are not very conclusive for a number of earthquakes. In general there is a tendency for the macroseismic depth values to be shallower than the instrumental ones. However, future comparisons for larger data sets should be made to get a more reliable conclusion for the possible agreement or disagreement between macroseismic and instrumental depth estimates.

In spite of the model and/or data problems mentioned above, this method is thought to be useful as it has the capability of depth estimation of historic earthquakes when isoseismals cannot be drawn. Even, when isoseismals are compiled, this method means a more objective and economic solution, because of the subjective and consuming nature of producing isoseismals for depth estimation.

Acknowledgements

I thank Dr. Gy Tarcsai (Eötvös University, Budapest) for his continuous encouragement and for allowing me to use his BECS program in the computations. I also thank Dr. A Tertulliani (Instituto Nazionale di Geofisica, Roma) for providing me recent Italian macroseismic bulletins.

Appendix

Estimated intensity attenuation curves based on the Kövesligethy model and the distribution of the observed (input) intensity data for the selected 50 earthquakes (Figs 1–7)

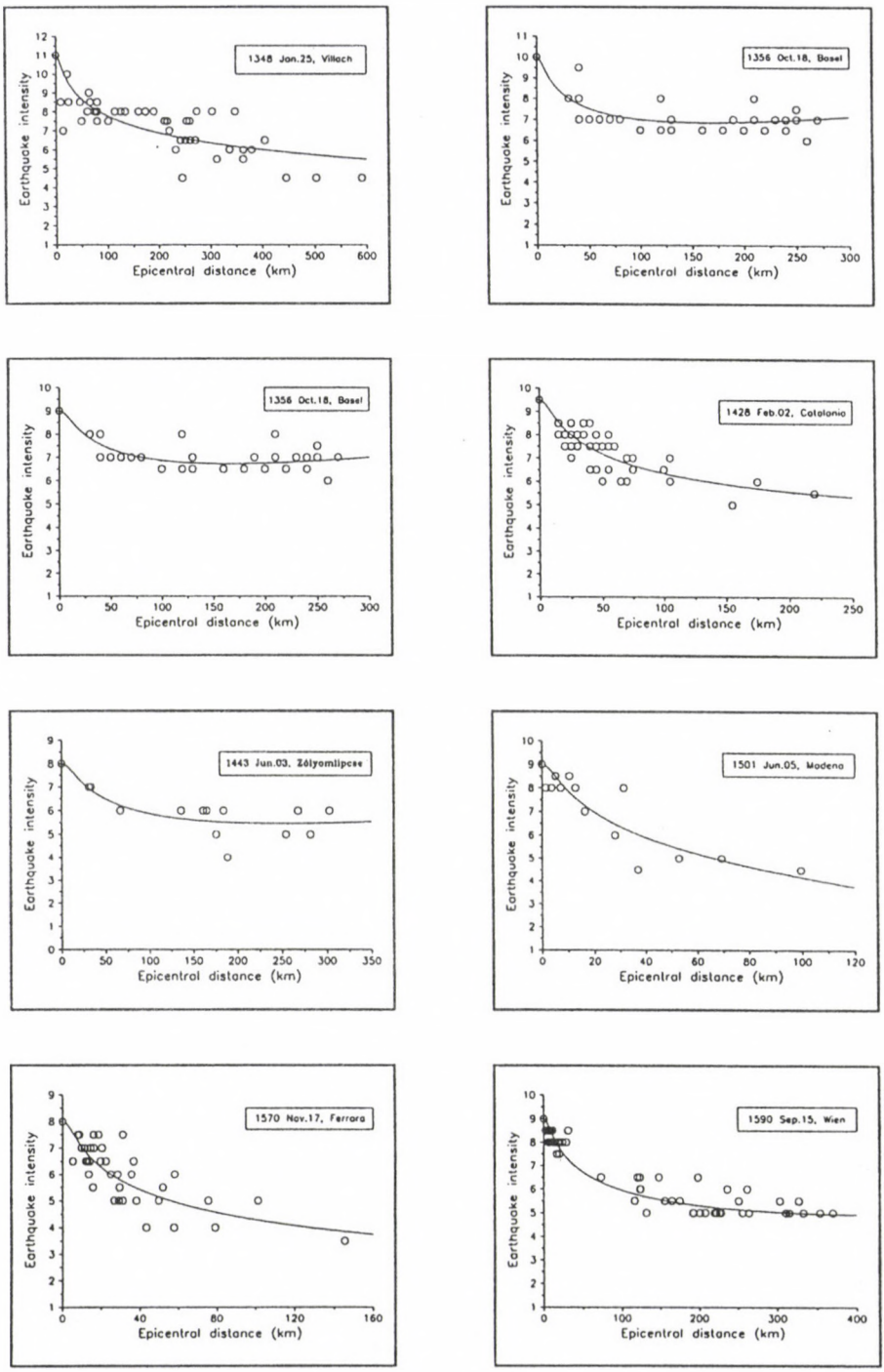
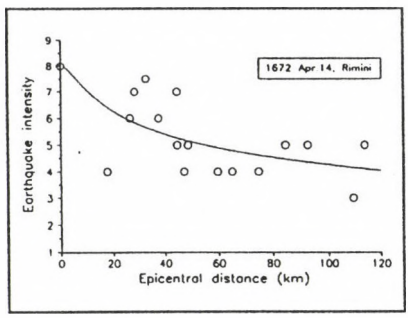
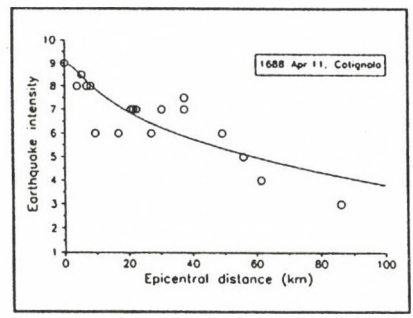
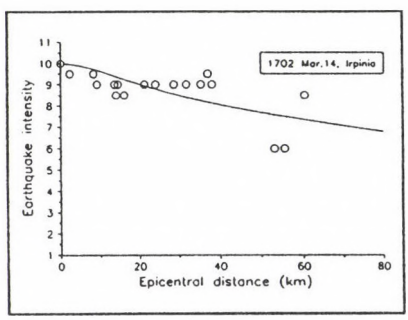
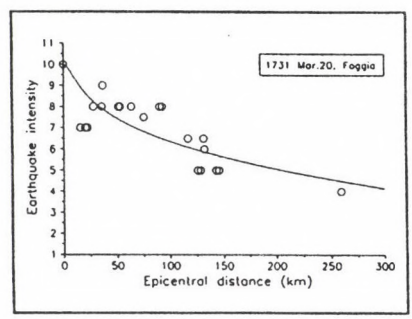

1348 Jan.25, Villach
1356 Oct.18, Basel
1356 Oct.18, Basel
1428 Feb.02, Catalonia
1443 Jun.03, Zólyomlipcse
1501 Jun.05, Modena
1570 Nov.17, Ferrara
1590 Sep.15, Wien
Earthquake intensity
Epicentral distance (km)

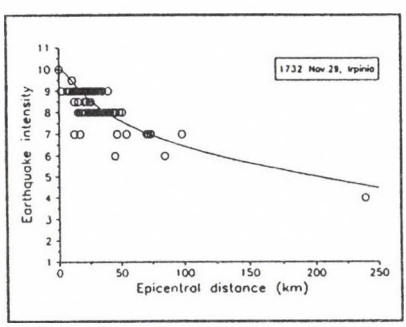
Fig. 1.

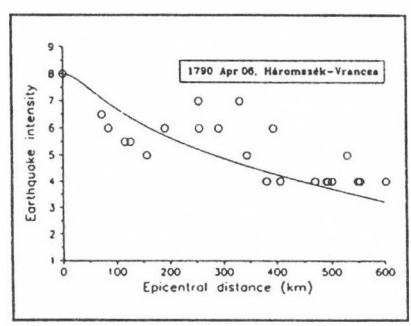

1672 Apr 14, Rimini
Earthquake intensity
Epicentral distance (km)

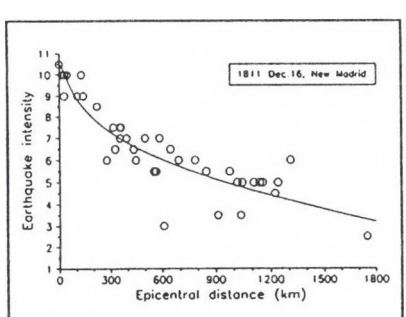

1688 Apr 11, Cotignola
Earthquake intensity
Epicentral distance (km)


1702 Mar.14, Irpinia
Earthquake intensity
Epicentral distance (km)


1731 Mar.20, Foggia
Earthquake intensity
Epicentral distance (km)


1732 Nov 29, Irpinia
Earthquake intensity
Epicentral distance (km)


1790 Apr 06, Háromszék-Vrancea
Earthquake intensity
Epicentral distance (km)


1811 Dec 16, New Madrid
Earthquake intensity
Epicentral distance (km)

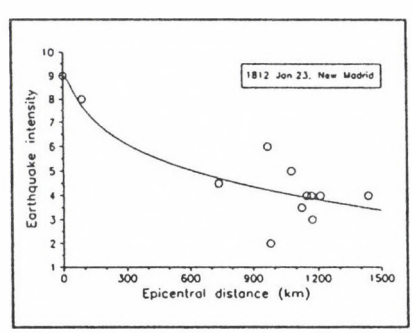
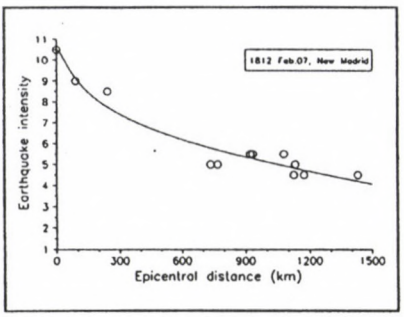
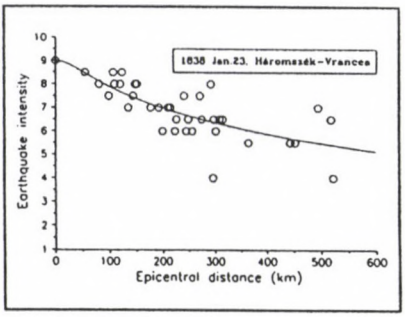
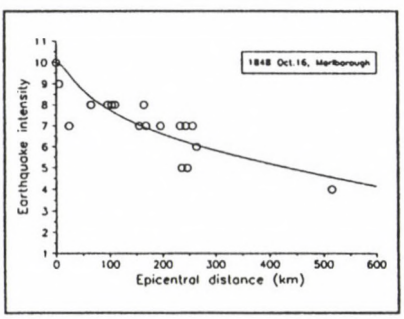
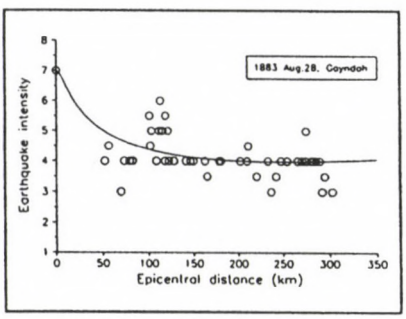

1812 Jan 23, New Madrid
Earthquake intensity
Epicentral distance (km)

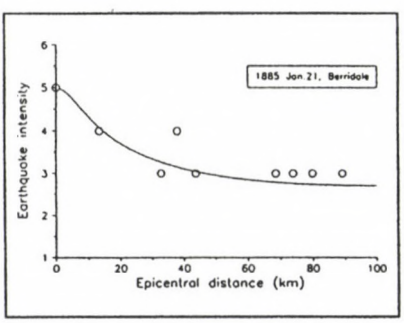
Fig. 2.

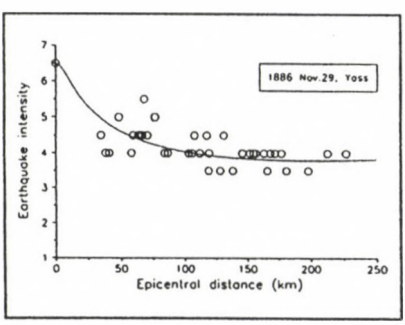

1812 Feb.07, New Madrid
Earthquake intensity
Epicentral distance (km)

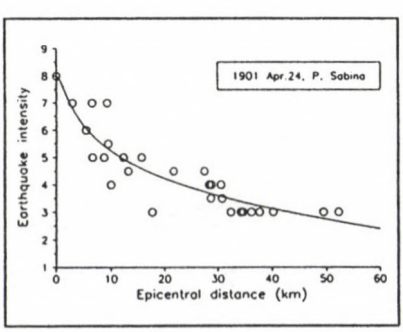

1838 Jan.23, Háromszék-Vrancea
Earthquake intensity
Epicentral distance (km)


1848 Oct.16, Marlborough
Earthquake intensity
Epicentral distance (km)


1883 Aug.28, Gayndah
Earthquake intensity
Epicentral distance (km)


1885 Jan.21, Berridale
Earthquake intensity
Epicentral distance (km)


1886 Nov.29, Yass
Earthquake intensity
Epicentral distance (km)


1901 Apr.24, P. Sabina
Earthquake intensity
Epicentral distance (km)

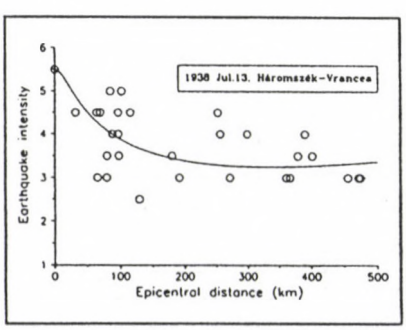

1938 Jul.13, Háromszék-Vrancea
Earthquake intensity
Epicentral distance (km)

Fig. 3.

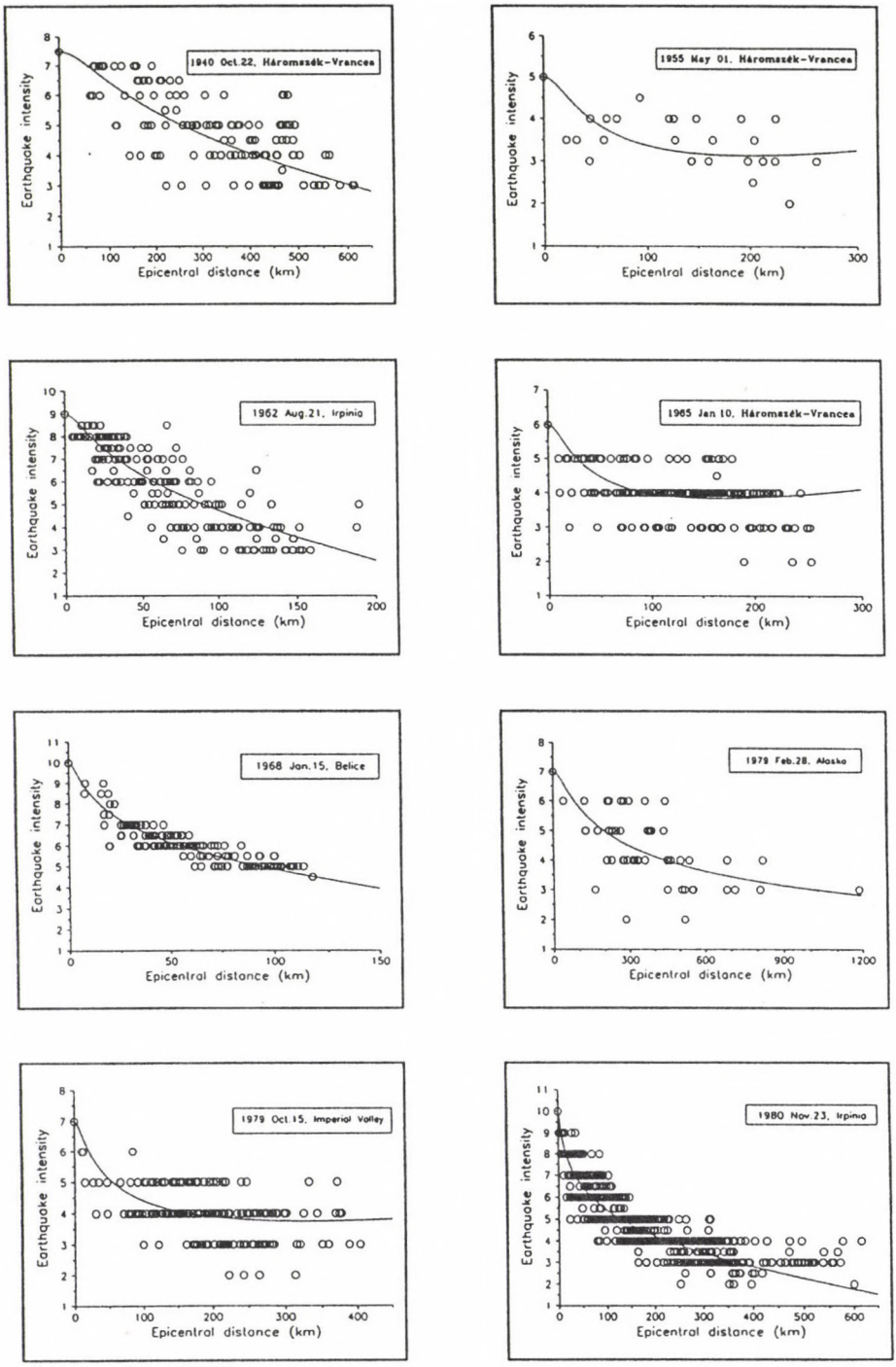

1940 Oct.22, Háromszék-Vrancea
1955 May 01, Háromszék-Vrancea
1962 Aug.21, Irpinia
1965 Jan 10, Háromszék-Vrancea
1968 Jan.15, Belice
1979 Feb.28, Alaska
1979 Oct.15, Imperial Valley
1980 Nov.23, Irpinia
Earthquake intensity
Epicentral distance (km)

Fig. 4.

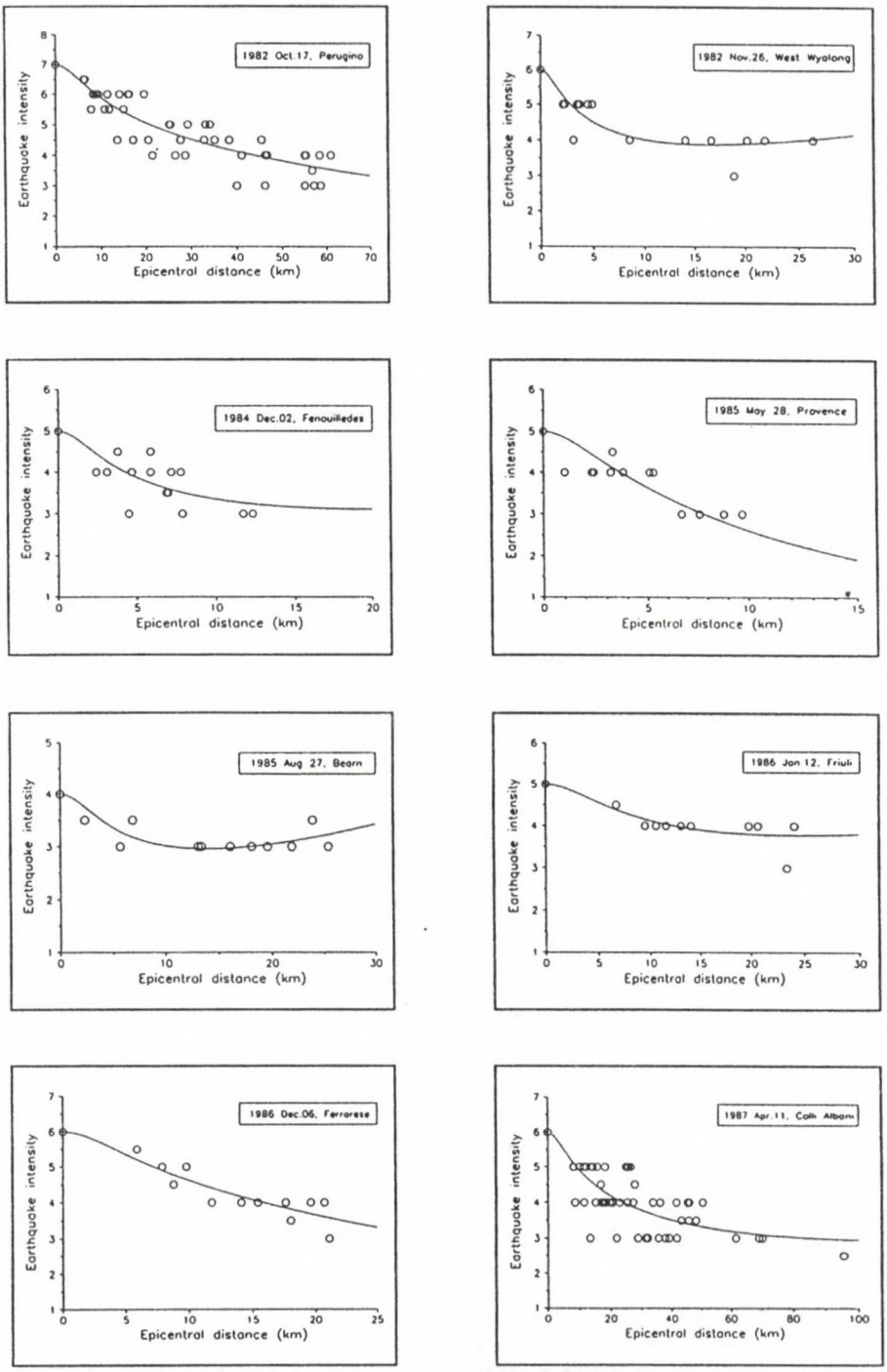

1982 Oct.17, Perugino
1982 Nov.26, West Wyalong
1984 Dec.02, Fenouilledes
1985 May 28, Provence
1985 Aug 27, Bearn
1986 Jan 12, Friuli
1986 Dec.06, Ferrarese
1987 Apr.11, Colli Albani
Earthquake intensity
Epicentral distance (km)

Fig. 5.

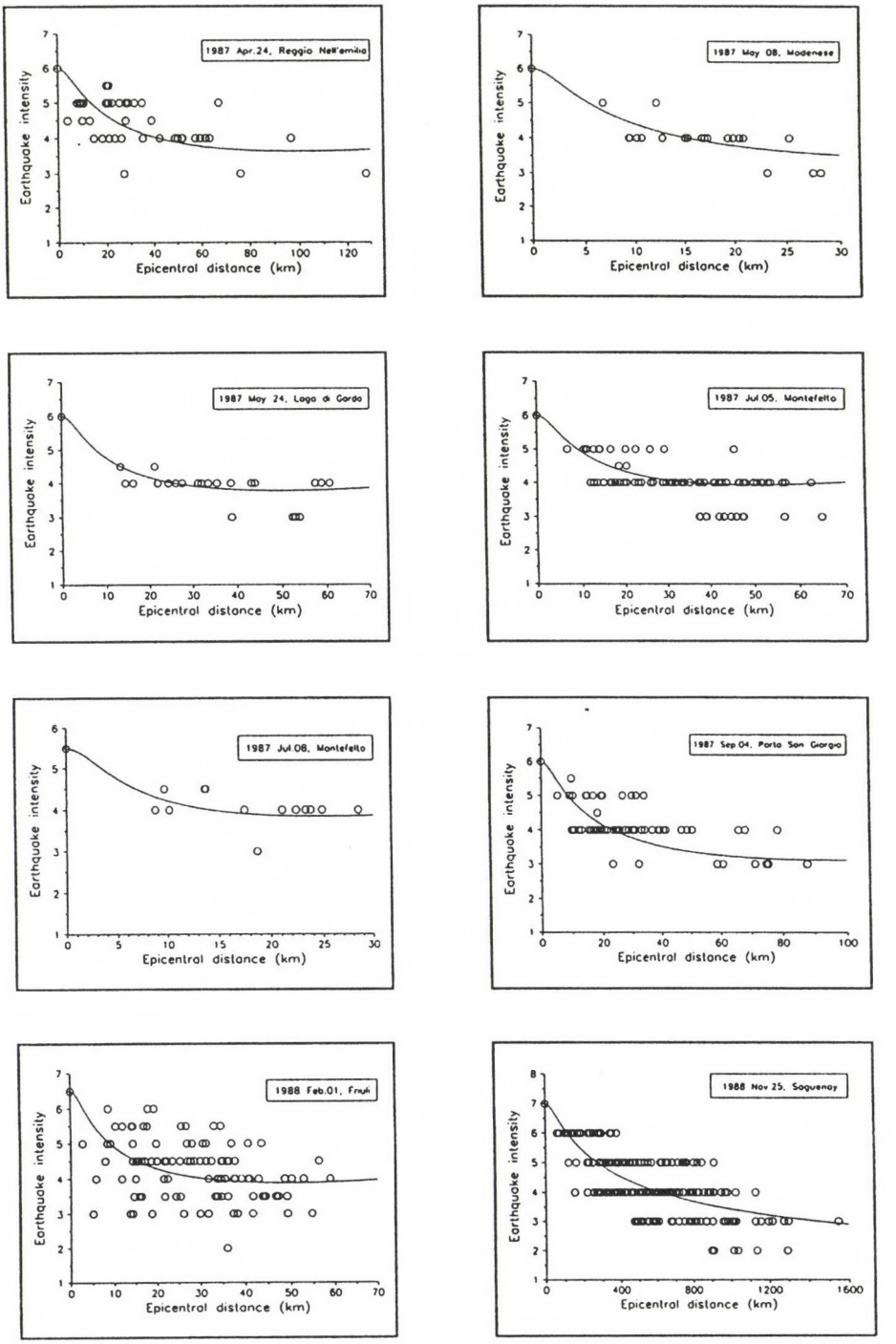

1987 Apr.24, Reggio Nell'emilia
1987 May 08, Modenese
1987 May 24, Lago di Garda
1987 Jul 05, Montefeltro
1987 Jul 08, Montefeltro
1987 Sep 04, Porta San Giorgio
1988 Feb.01, Friuli
1988 Nov 25, Saguenay
Earthquake intensity
Epicentral distance (km)

Fig. 6.

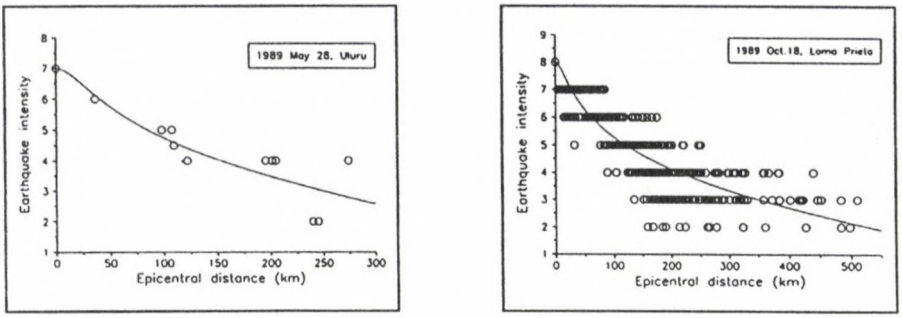


Fig. 7.

References

Bevington P R 1969: Data reduction and error analysis for the physical sciences. McGraw-Hill, New York

Kövesligethy R 1906: *Mathematikai és Természettudományi Értesítő* (in Hungarian), 24, 349–368.

Levret A, Backe J C, Cushing M 1994: *Natural Hazards*, 10, 19–46.

Press W H, Flannery B P, Teukolsky S A, Vetterling W T 1989: Numerical recipes. The art of scientific computing (Fortran version). Cambridge University Press, Cambridge

Tarcsai Gy 1991: BECS fortran computer program for parameter estimation of non-linear equations. Eötvös University, Budapest

References in Table I

1. Pospischl D ed. 1985: Atlas of isoseismal maps of Italian earthquakes. CNR, Bologna, pp. 164.
2. Mayer-Rosa D, Cadiot B 1979: A review of the Basel 1356 earthquake: Basic data. *Tectonophysics*, 53, 325–333.
3. Banda E, Correig A M 1984: The Catalan Earthquake of February 2, 1428. *Eng. Geol.*, 20, 89–97.
4. Procházkova D, Kárnik V eds 1978: Atlas of isoseismal maps, Central and Eastern Europe. Geophysical Institute, Prague, pp. 135.
5. Gutdeutsch R, Hammerl Chr, Mayer I, Vocelka K 1987: Erdbeben als historisches Ereignis. Die Rekonstruktion des Bebens von 1590 in Niederösterreich. Springer, Wien, pp. 222.
6. Shebalin N V ed. 1974: Atlas of isoseismal maps, Balkan Region. UNDP/UNESCO, Skopje, pp. 275.
7. Nuttli O W 1973: The Mississipi Valley earthquakes of 1811 and 1812: intensities, ground motion and magnitudes. *Bull. Seism. Soc. Am.*, 63, 227–248.
8. Eiby G A 1980: The Marlborough earthquake of 1848. DSIR Bulletin 225, Wellington, pp. 82.
9. Rynn J M W, Denham D, Greenhalgh S, Jones T, Gregson P J, McCue K F, Smith R S 1987: Atlas of isoseismal maps of Australian Earthquakes. Part 2, BMR Bulletin 222, Canberra
10. Stover C W, Reagor B G, Wetmiller R J 1980: Intensities and isoseismal map for the St. Elias earthquake of February 28, 1979. *Bull. Seism. Soc. Am.*, 70, 1635–1649.

11. Rojahn C ed. 1980: Selected papers on the Imperial Valley, California, earthquake of October 15, 1979. U.S. Geological Survey Open-File Report 80–1094, 1–16.
12. ING 1983: Bollettino Macroseismico 1982. Instituto Nazionale di Geofisica, Roma
13. BCSF 1990: Observations Sismologiques, Sismicité de la France en 1984, 1985 et 1986. Bureau Central Sismologique Français, Strasbourg, pp. 229.
14. ING 1987: Bollettino Macroseismico 1986. Instituto Nazionale di Geofisica, Roma
15. ING 1988: Bollettino Macroseismico 1987. Instituto Nazionale di Geofisica, Roma
16. ING 1989: Bollettino Macroseismico 1988. Instituto Nazionale di Geofisica, Roma
17. Drysdale T A, Catka M G 1989: Intensity distribution of the 1988 M6 Saguenay earthquake (abstract). *Seism. Res. Letters*, 60, 71–73.
18. Michael-Leiba M, Love D, McCue K, Gibson G 1994: The Uluru (Ayers Rock), Australia, earthquake of 28 May 1989. *Bull. Siesm. Soc. Am.*, 84, 209–214.
19. Stover C W, Reagor B G, Baldwin F W, Brewer L R 1990: Preliminary isoseismal map for the Santa Cruz (Loma Prieta), California, earthquake of October 18, UTC. U.S. Geological Survey Open-File Report 90–0018, pp. 24.

Acta Geod. Geoph. Hung., Vol. 32(1-2), pp. 87–103 (1997)

THE OPTIMAL DESIGN OF THE MONITORING NETWORK OF A GROUND FAULT IN SYRIA

SAAD YAZJI[1]

[Manuscript received February 4, 1997]

The optimal design is given of the monitoring network of the movements in the fault area beside a river and a dam in Syria.

The optimal design is sought depending on the optimality criteria used to solve the second order design (S.O.D.), that means the precision, the reliability and the economy. A model is proposed to control the vertical fault movements according to the geodetic methodology and its instrumentations in the sense of analysing the sensitivity concept of the geodetic networks with respect to specified deformation models.

Keywords: criterion matrix; crustal movements; deformation analysis; fault; graben; network design; S.O.D.

1. Introduction

Usually to study deformation problems we use either geodetic or non-geodetic methods containing their special instrumentation and methodologies. The geodetic methods include terrestrial geodetic procedures, space techniques, and photogrammetric solutions, while the non-geodetic ones use the special geotechnical and structural instrumentation such as strain meters, tiltmeters, and laser alignment. Terrestrial geodetic methods are useful in determining the global geometrical status of a deformable body. Therefore, in any case, a terrestrial geodetic network always forms part of a deformation monitoring scheme. Thus how to obtain an optimal design of the network configuration and observation plan so that it can serve the task in the most economical way is always a challenge to geodesists.

Although a lot of work has been made to develop numerical solution methods for required measurement accuracies (weights), geodesists are still using computer simulation or the "trial and error" method to optimise the configuration of a network. In spite of that the computer simulation method requires a very large amount of work, in most cases it never gives the optimum network.

Conventionally, it was usual to advocate that for engineering and monitoring networks the relatively small changes in configuration have no significance for the improvement of the network accuracy and therefore, there is no need for the optimization of the configuration. Thus we find the most important role of the second order design (S.O.D.) in the optimal design of geodetic networks, especially for monitoring purposes.

[1]Technical University Budapest, Department of Surveying, H–1111 Budapest, Műegyetem rkp. 3, Hungary

1217-8977/97/$ 5.00

2. Geometrical deformation analysis

In general, geodetic monitoring networks, whether they are established for monitoring engineering structures or ground subsidence in mining areas or tectonic movements, are divided into relative networks and reference networks (Chrzanowski et al. 1983). The relative geodetic networks provide the change in relative positions of the points with the identification of the deformation model being the main purpose, while the reference networks serve as a reference frame to which the displacements of objective points are referred.

Geometrical analysis seeks the values for the deformation parameters which characterise the behaviour of the deformable body. In practice, these may be obtained directly from repeated observations. Recently all the problems of geometrical deformation analysis have been solved. For details see, for example, Heck et al. (1982), Chrzanowski et al. (1983), Chen (1983, 1986), and Gaspary (1987).

In what follows I will give a brief review of the basic principle of the geometrical deformation analysis based on Kuang et al. (1991):

Let ℓ_i be n_i by 1 vector of observations, with weight matrix P_i in epoch $i = 0, 1, 2, \ldots, k$ (n_i is the number of observations at epoch i). The deformation model ($\mathbf{B}\,\mathbf{e}$) is related to the observables through the null hypothesis:

$$\mathbf{H}_0 : \mathbf{E}(\ell_i) = \mathbf{E}(\ell_0) + \mathbf{A}_i\mathbf{B}_i\mathbf{e}\,, \tag{1}$$

where $\mathbf{A}_i$ is the configuration matrix which relates the observables to deformation model; $\mathbf{B}_i$ is the coefficient matrix of the deformation model and it is a function of position and time; $\mathbf{e}$ is the vector of deformation parameters. The parameters $\mathbf{e}$ may be estimated from the following mathematical model:

$$\begin{pmatrix} \mathbf{I}_0 \\ \ell_1 \\ \cdot \\ \cdot \\ \ell_k \end{pmatrix} + \begin{pmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \cdot \\ \cdot \\ \mathbf{v}_k \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{A}_1\mathbf{B}_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \mathbf{I} & \mathbf{A}_k\mathbf{B}_k \end{pmatrix} \begin{pmatrix} \xi \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{e} \end{pmatrix} \tag{2}$$

with ξ being a vector of nuisance parameters. Assuming there is no correlation between epochs, the weight matrix can be written as:

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_0 & 0 & \cdot & \cdot & 0 \\ 0 & \mathbf{P}_0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \mathbf{P}_k \end{pmatrix}. \tag{3}$$

Applying the least-squares criterion to the above model and eliminating ξ allow the vector of deformation parameter $\mathbf{e}$ and its accuracy to be calculated from Chen (1983):

$$\mathbf{e} = \left(\sum_1^k \mathbf{B}_i^t\mathbf{A}_i^t\mathbf{P}_i\mathbf{B}_i - \sum_1^k \mathbf{B}_i^t\mathbf{A}_i^t\mathbf{P}_i \left(\sum_0^k \mathbf{P}_i \right)^{-1} \sum_1^k \mathbf{P}_i\mathbf{A}_i\mathbf{B}_i \right)^{-1} \cdot$$

$$\cdot \left(\sum_{1}^{k} \mathbf{B}_i^t \mathbf{A}_i^t \mathbf{P}_i \ell_i - \sum_{1}^{k} \mathbf{B}_i^t \mathbf{A}_i^t \mathbf{P}_i \left(\sum_{0}^{k} \mathbf{P}_i \right)^{-1} \sum_{1}^{k} \mathbf{P}_i \ell_i \right) \tag{4}$$

$$\mathbf{Q}_e = \left(\sum_{1}^{k} \mathbf{B}_i^t \mathbf{A}_i^t \mathbf{P}_i \mathbf{A}_i \mathbf{B}_i - \sum_{1}^{k} \mathbf{B}_i^t \mathbf{A}_i^t \mathbf{P}_i \left(\sum_{0}^{k} \mathbf{P}_i \right)^{-1} \sum_{1}^{k} \mathbf{P}_i \mathbf{A}_i \mathbf{B}_i \right)^{-1} . \tag{5}$$

In the design phase, it is justified to assume that the observation schemes and deformation models are the same for all the epochs (i.e. $\mathbf{A}_i = \mathbf{A}$, $\mathbf{P}_i = \mathbf{P}$, $\mathbf{B}_i = \mathbf{B}$ for all i).

Therefore the expression for the estimation of the deformation parameters and their accuracies can be obtained considering only two epochs by the following way:

$$\mathbf{e} = (\mathbf{B}^t \mathbf{A}^t \mathbf{P} \mathbf{A} \mathbf{B})^{-1} \mathbf{B}^t \mathbf{A}^t \mathbf{P} (\ell_2 - \ell_1) \tag{6}$$

$$\mathbf{Q}_e = 2(\mathbf{B}^t \mathbf{A}^t \mathbf{P} \mathbf{A} \mathbf{B})^{-1} . \tag{7}$$

Since at the design stage the deformation model ($\mathbf{B}$ $\mathbf{e}$) must be assumed to be known, the design problem involves how to find the best configuration matrix ($\mathbf{A}$) and the weight matrix ($\mathbf{P}$) in order to attain required accuracies of the deformation parameters $\mathbf{e}$ in the most economical way.

3. Definition of the optimality criteria

The quality of a monitoring network is characterised by precision, reliability, sensitivity, and economy. Precision is a measure of the variance and covariance of the estimated parameters, reliability means detection, localisation and elimination of outlying observations; sensitivity refers to the detection of the minimum magnitude of the parameters; and finally, economy is a measure of the cost and benefit of the project. So the problem is to design a precise, reliable, and sensitive monitoring network which can be realised in an economical way.

It must be mentioned here the important role of using criterion matrices as precision criteria for geodetic networks, in which an isotropic and homogeneous error structure, i.e. the Taylor-Karman structure or its special form of completely isotropic structure (Grafarend 1972, Baarda 1973), is usually sought. Construction of criterion matrices for detection of displacements has been investigated by Sprinsky (1978), Wimmer (1982), and Crosilla (1983).

4. The solution of the design problem

As it was mentioned for engineering and monitoring networks the relatively small changes in configuration have no significance for the improvement of the network accuracy and, therefore, there is no need for optimization of the configuration. Thus I will solve the design problem for a monitoring network by using the analytical S.O.D. methods.

To realise an approximation to solve the S.O.D. problem, there are three basic approaches, two of them try to approximate the criterion matrix itself, the third approximates its inverse. In order to include constraint as well as free networks and regular as well as singular criterion matrices, the inverses of the normal equations and the criterion matrices are written in the following as pseudo-inverses, in general.

4.1 Direct approximation of the criterion matrix, approach (i)

According to a proposal of Schaffrin (1983) the connection between the normal equations and the criterion matrix as ideal variance-covariance matrix is:

$$\mathbf{N} = \mathbf{A}^t\mathbf{P}\mathbf{A} = \mathbf{Q}_x^{-1}$$

or

$$\mathbf{A}^t\mathbf{P}\mathbf{A} = \mathbf{Q}_x^{+} . \tag{8}$$

The general solution can be written as:

$$\mathbf{vec}\,\mathbf{dP} = (\mathbf{K} \otimes \mathbf{K})^{+}\mathbf{vec}\,\mathbf{Q}_x , \tag{9}$$

where $\mathbf{vec}\,\mathbf{dP}$ is the vector of the diagonal weight matrix, $\mathbf{vec}\,\mathbf{Q}_x$ is the vector of the criterion matrix, $\otimes$ means the Khatri-Rao product, $\mathbf{K} = \mathbf{Q}_x\mathbf{A}^t$, and $\mathbf{A}$ is the design matrix.

4.2 Iterative approximation of the criterion matrix, approach (ii)

Wimmer (1981, 1982) applies the law of error propagation upon the universal solution of an adjustment by variation of co-ordinates:

$$\mathbf{X} = (\mathbf{A}^t\mathbf{P}\mathbf{A})^{+}\mathbf{A}^t\mathbf{P}\mathbf{L} \tag{10}$$

which leads to the variance-covariance matrix of x:

$$\mathbf{Q}_x = (\mathbf{A}^t\mathbf{P}\mathbf{A})^{+}\mathbf{A}^t\mathbf{P}\mathbf{Q}_l\mathbf{P}\mathbf{A}(\mathbf{A}^t\mathbf{P}\mathbf{A})^{+} . \tag{11}$$

Introducing:

$$\mathbf{H}^{(i)} = (\mathbf{A}^t\mathbf{P}^{(i)}\mathbf{A})^{+}\mathbf{A}^t\mathbf{P}^{(i)} \tag{12}$$

and

$$\mathbf{P}^{+} = \mathbf{Q}_l . \tag{13}$$

The basic formula of this approach:

$$\mathbf{H}\mathbf{P}^{+}\mathbf{H}^t = \mathbf{Q}_x \tag{14}$$

and the solution:

$$(\mathbf{vec}\,\mathbf{dP})^{i+1} = (\mathbf{H}^{(i)} \otimes \mathbf{H}^{(i)})^{+}\mathbf{vec}\,\mathbf{Q}_x . \tag{15}$$

This solution is iterative, because of $\mathbf{H}$ contains an actual weight matrix $\mathbf{P}$ which is updated to the new $\mathbf{P}^{+}$.

4.3 Direct approximation of the inverse criterion matrix, approach (iii)

Matrix Eq. (8) is converted to a set of linear equation:

$$(\mathbf{A}^t \otimes \mathbf{A}^t)\mathbf{vec}\,\mathbf{dP} = \mathbf{vec}\,\mathbf{Q}_x \tag{16}$$

the solution is

$$\mathbf{vec}\,\mathbf{dP} = (\mathbf{A}^t \otimes \mathbf{A}^t)^+ \mathbf{vec}\,\mathbf{Q}_x^+ . \tag{17}$$

Thus this approach approximates the inverse of the criterion matrix and not the criterion matrix itself. Numerical examples show indeed (Schmitt 1985), that the approximation of the inverse of the criterion matrix is very good, but the global measures $\mathbf{d}^t\mathbf{d}$ for the approximation of $\mathbf{Q}_x$ are much larger than in the other two approaches, where $\mathbf{d}$ is the diference vector between the effective cofactor matrix and the given criterion matrix:

$$\mathbf{D} = (\mathbf{A}^t\mathbf{PA})^+ - \mathbf{Q}_x \tag{18}$$

$$\mathbf{d} = \mathbf{vec}\,\mathbf{D}\,,$$

where $\mathbf{D}$ is called the matrix of residuals and by proposing the square sum of the residuals — similarly to a least-squares adjustment — as a global measure for the approximation quality:

$$\mathbf{d}^t\mathbf{d}, \qquad \mathbf{d} = \mathbf{vecD}\,.$$

In order to get a better approximation of the criterion matrix and to make a direct comparison with other methods possible, solution (iii) can be modified by a linear transformation of the weight vector with a factor λ (Müller and Illner 1984). This factor must be estimated such that the square sum of the residuals $\mathbf{d}^t\mathbf{d}$ is minimum for:

$$\mathbf{P} = \mathbf{diag}(\lambda \cdot \mathbf{vec}\,\mathbf{dP})\,. \tag{19}$$

The square sum of the residuals can be written as a function of λ:

$$\mathbf{f}(\lambda) = \mathbf{d}^t\mathbf{d} = \mathbf{e}^t \left\{ \left[1/\lambda(\mathbf{A}^t\mathbf{PA})^+ - \mathbf{Q}_x\right] * \left[1/\lambda(\mathbf{A}^t\mathbf{PA})^+ - \mathbf{Q}_x\right] \right\} \mathbf{e}\,, \tag{20}$$

where $\mathbf{e}^t = (\mathbf{1}, \mathbf{1}, \ldots, \mathbf{1})$ is the summation vector, $*$ is the element-wise Hadamard product and the result of the multiplication is:

$$\mathbf{f}(\lambda) = 1/\lambda^2 \left[(\mathbf{A}^t\mathbf{PA})^+ * (\mathbf{A}^t\mathbf{PA})^+\right] - 2/\lambda \left[(\mathbf{A}^t\mathbf{PA})^+ * \mathbf{Q}_x\right] \mathbf{Q}_x * \mathbf{Q}_x\,. \tag{21}$$

The best approximation can now be achieved for:

$$\frac{\mathbf{df}}{\mathbf{d}\lambda} = \mathbf{e}^t \left\{ \frac{2}{\lambda^2} \left[(\mathbf{A}^t\mathbf{PA})^+ * \mathbf{Q}_x\right] - \frac{2}{\lambda^3} \left[(\mathbf{A}^t\mathbf{PA})^+ * (\mathbf{A}^t\mathbf{PA})^+\right] \right\} \mathbf{e} = 0\,. \tag{22}$$

As a result λ can be computed by:

$$\lambda = \frac{\mathbf{e}^t \left[(\mathbf{A}^t\mathbf{PA})^+ * (\mathbf{A}^t\mathbf{PA})^+\right] \mathbf{e}}{\mathbf{e}^t \left[(\mathbf{A}^t\mathbf{PA})^+ * \mathbf{Q}_x\right] \mathbf{e}}\,. \tag{23}$$

Or by using the general matrix product and the operator trace, λ will be in the form:

$$\lambda = \frac{\mathbf{trace}\left[(\mathbf{A}^t\mathbf{PA})^+ (\mathbf{A}^t\mathbf{PA})^+\right]}{\mathbf{trace}\left[(\mathbf{A}^t\mathbf{PA})^+ \mathbf{Q}_x\right]}\,. \tag{24}$$

5. Solution methods

— Least-squares solution

— Linear programming

— Non-linear programming

In this article I will chose the first solution method, that is, the least-squares solution.

For our three approaches which lead to generally redundant systems, we can derive least-squares solutions where the global measure for the approximation quality is entering directly in the first two approaches:

$$(\mathbf{K}\otimes\mathbf{K})^t(\mathbf{K}\otimes\mathbf{K})\mathrm{vec}\,\mathrm{d}\mathbf{P} - (\mathbf{K}\otimes\mathbf{K})^t\mathrm{vec}\,\mathbf{Q}_x = (\mathbf{K}^t\mathbf{K} * \mathbf{K}^t\mathbf{K})\mathrm{vec}\,\mathrm{d}\mathbf{P} - (\mathbf{K}\otimes\mathbf{K})^t\mathrm{vec}\,\mathbf{Q}_x = 0$$

$$\mathbf{d}^t\mathbf{d} \rightarrow \mathbf{min}\,, \tag{25}$$

$$(\mathbf{H}^t\mathbf{H} * \mathbf{H}^t\mathbf{H})\mathrm{vec}\,\mathbf{dP} - (\mathbf{H}\otimes\mathbf{H})^t\mathrm{vec}\,\mathbf{Q}_x = 0\,, \qquad \mathbf{d}^t\mathbf{d} \rightarrow \mathbf{min}\,, \tag{26}$$

$$(\mathbf{A}\mathbf{A}^t * \mathbf{A}\mathbf{A}^t)\mathrm{vec}\,\mathbf{dP} - (\mathbf{A}^t\otimes\mathbf{A})^t\mathrm{vec}\,\mathbf{Q}_x = 0\,, \qquad \mathbf{d}^t\mathbf{d} \rightarrow \mathbf{min}\,. \tag{27}$$

Müller and Illner (1984) have studied comparatively the three solutions concerning different criteria, such as global and local approximation quality, resulting accuracy of the observations and computing time. They found that the third solution, approach (iii), yields the most homogeneous and symmetric weight distribution, and it will run faster than other solutions. Finally Schmitt (1985) found the decisive result which says that the approximation of the inverse criterion matrix, approach (iii), is very robust and stable against the criterion matrix, it produces useful solutions in all cases.

This is the main reason why I follow this approach forward to an operational strategy. For this goal Müller and Schmitt (1985) developed a program system for the second order design (S.O.D.) of two-dimensional networks, which was written in FORTRAN 77. The system allows the weight optimization within the 3-step-strategy, developed in Karlsruhe. The mathematical basis is the modified least-squares approximation of an inverse criterion matrix, approach (iii), which was restricted to non-negative weights. The following types of observations are permitted and can be mixed to some useful combinations: azimuths, directions, angles, distances and distance ratios.

The system works in the three main components:

— KRIT: produces a criterion matrix of Taylor-Karman or completely isotropic and allows an arbitrary datum transformation;

— SODES2: computes the S.O.D.-solution for individual and for group weights and analyses the resulting design with respect to its accuracy and reliability;

— NETZPLOT: produces a plot of the network design including postulated and realised error ellipses.

The problem of this system was that it was written for a mainframe computer of type Siemens 7881. Thus it is impossible to use it on personal computers, but I could implement this software program system on microcomputers, and its use became easy and very quick to get the solutions of the S.O.D. problem (Yazji 1997).

6. The actual problem

Our problem is the design of a horizontal network for monitoring movements of the fault at graben situated in Syria. This graben is beside a river and nowadays a dam is being built in this area. Since we decided to build this dam, attention was paid to study this fault and to observe any movements of this graben before starting the construction of the dam and during its construction. Until now the observations done showed that this graben, which has a width 30 m, doesn't move, but these observations were designed according to the principle of alignment. This non-geodetic method was made by taking four points, with one of them is situated inside the fault, but in my opinion this way is not enough for monitoring the movements of the crust because of constructing of the dam and because of the mutual influence between the fault and the dam and the sea which will be formed in front of the dam in the future and to ensure enough safety for the dam and finally for our economy. Thus I decided to study this problem and to design the desired network from the geodetic point of view by using the plans that I have for this area. I will also, in another work, design a horizontal network for monitoring the deformations of the dam itself.

To design this network I used approach (iii) mentioned above, and the program system that I have mentioned previously. Because of the topography of the graben area and because of the river and the dam I have chosen net points in such a way to get a good shape for the network, and secondly to cover enough area around the fault and beside the dam. So I have chosen 16 points, four of them are inside the graben. By this design the shape conditions are satisfied, where the network shape is approximately rectangular. Because of the topography of the area some observations are not possible what affects the homogeneity and the isotropy of the network, this was clear for the point number five which has a small number of observations in comparison with other points. I tried to design two networks, one of them consists of 16 points (Fig. 1), and in the second network point number five is excluded (Fig. 3), that means the network consists of 15 points only. The design is also done for the two cases, for trilateration design and for triangulation design, that is, by using only the distance measurements in the first case, and in the second the measurements of distances and directions are used. I have used the most accurate electronic measurement instruments, for example the instrument which has the accuracy relation of measuring a distance:

$$\sigma_m = (0.2\ \text{mm} + 0.2\ \text{mm}/D\text{km})\,.$$

The accuracy of measuring a direction is: M = 2 sec.

With respect to the characteristic distance **d** of the correlation functions, systematic studies about a suitable determination of **d** are practically missing, one can

use the result of Wimmer (1981) who derived from schematic networks the interval:

$$\frac{\mathbf{r}_{max}}{5} \geq \mathbf{d} \geq \frac{\mathbf{r}_{max}}{10} .$$

$\mathbf{r}_{max}$ is the maximum distance between two net points, or one can use the very widely accepted empirical value: $d \leq \mathbf{r}_{min}$, where $\mathbf{r}_{min}$ is the minimum distance between two net points.

For our network we have: $\mathbf{r}_{max} = 287.70$ m, $\mathbf{r}_{min} = 27.48$ m, so the first condition gives: 57.54 m $\geq \mathbf{d} \geq$ 28.77 m, but the empirical value is: $\mathbf{d} \leq 27.48$ m, finally I decided to consider $\mathbf{d} = 25$ m.

I chose the Taylor-Karman structure of completely isotropic type as a criterion matrix, that is the point and relative errors' ellipses are circles. Such an ideal error situation is coming up for example if the variance-covariance matrix of the network degenerates to identity matrix. But the identity matrix is unsuited to be introduced as criterion matrix in the second order design, because the postulation of zero-covariances between al co-ordinates is unrealisable. Therefore it is necessary to look for a general stochastic concept, in which an overall ideal covariance situation can be formulated. This concept is found in the interpretation of a geodetic point field as a stochastic process with the expectation values equal to zero and the covariance function with Taylor-Karman structure of the absolute co-ordinates, which is a homogeneous and isotropic structure. Thus I have chosen the T-K completely isotropic structure as a criterion matrix structure.

To save the original properties homogeneity and isotropy the criterion matrix has not been transformed by any datum transformation.

The postulate for the point accuracy is proposed to be less than 1.4 mm, so the standard deviation of the unit weight: $\sigma_0 = 1$ mm.

7. The model proposed to control the vertical movements of the fault

With respect to the design of a network to control the vertical movements I will give only a first idea of an optimal design of levelling networks for determination of *a priori* deformation models.

To get a first insight into the efficiency of different network designs Rohde (1980) carried out numerical simulation studies. As criterion for this optimality comparison Niemeier and Rohde (1982) used the sensitivity of a net with respect to specified deformation models.

According to these studies Niemeier and Rohde (1982) came to the following result with respect to a graben or dislocation. For a dislocation it seems to be most important to have as much as possible of the levelling lines crossing the fault line. This requirement seems to be obvious, but it implies an exact *a priori* knowledge of the geological fracture zones. The absolute number of points seems to be not important, there are only few differences in the scale factors between the inner loop variates (see page 155 of Niemeier and Rohde 1982). But the addition of diagonals crossing the fault line (dashed lines in Fig. 8) would give improvement of about 30 percent, while the total length of all levelling lines and by this the costs would

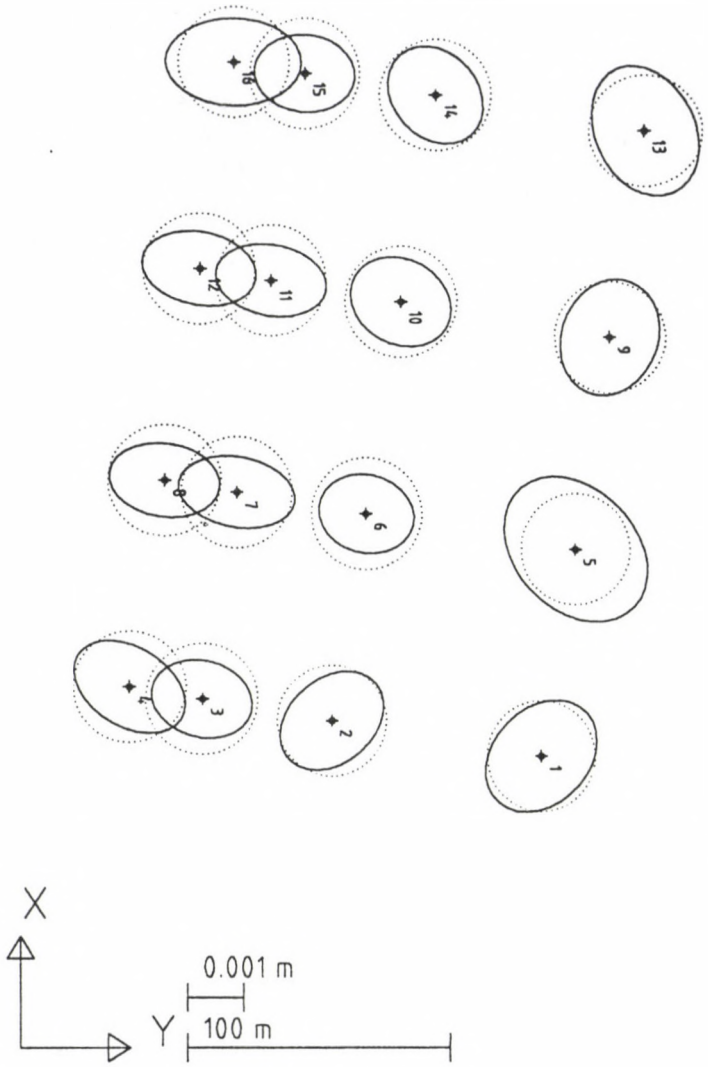


Fig. 1.

increase only by about 23 percent. Further on the absolute extension of the net is of low influence. Similar results are found for a graben fracture, which can be characterised with some simplification as a double dislocation.

Thus according to these results I suggest the scheme in Fig. 8 to control the vertical crustal movements of my problem, in which some points of the designed monitoring network are common between the two networks, that is, the design realises the cost criterion.

8. Results and discussion

For the designed horizontal network with 16 points, shown in Fig. 1 for distance measurements, it can be showed that the network is not so bad but the isotropy is missing, approximately for all points, especially for point number five, that is because of the small number of observations taken from this point to others. If we look at Fig. 2 in which the distance and direction measurements are taken, we clearly find the important role of directions to realise the isotropy of the network, so it can be said that the use of the mixed networks gives better results than using

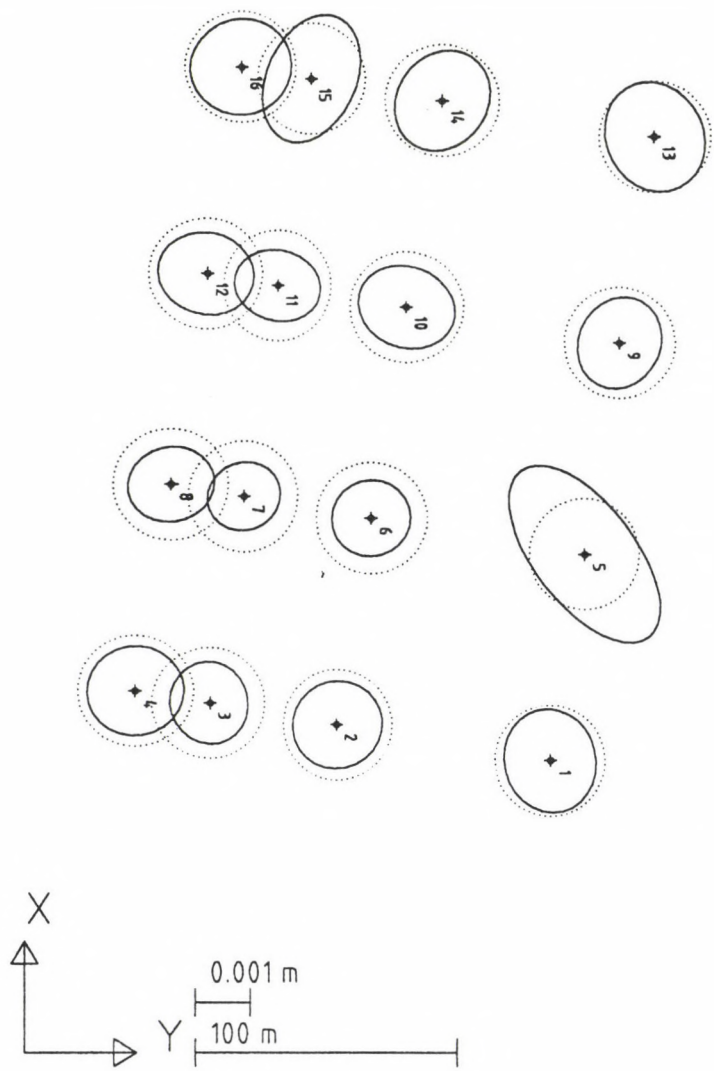


Fig. 2.

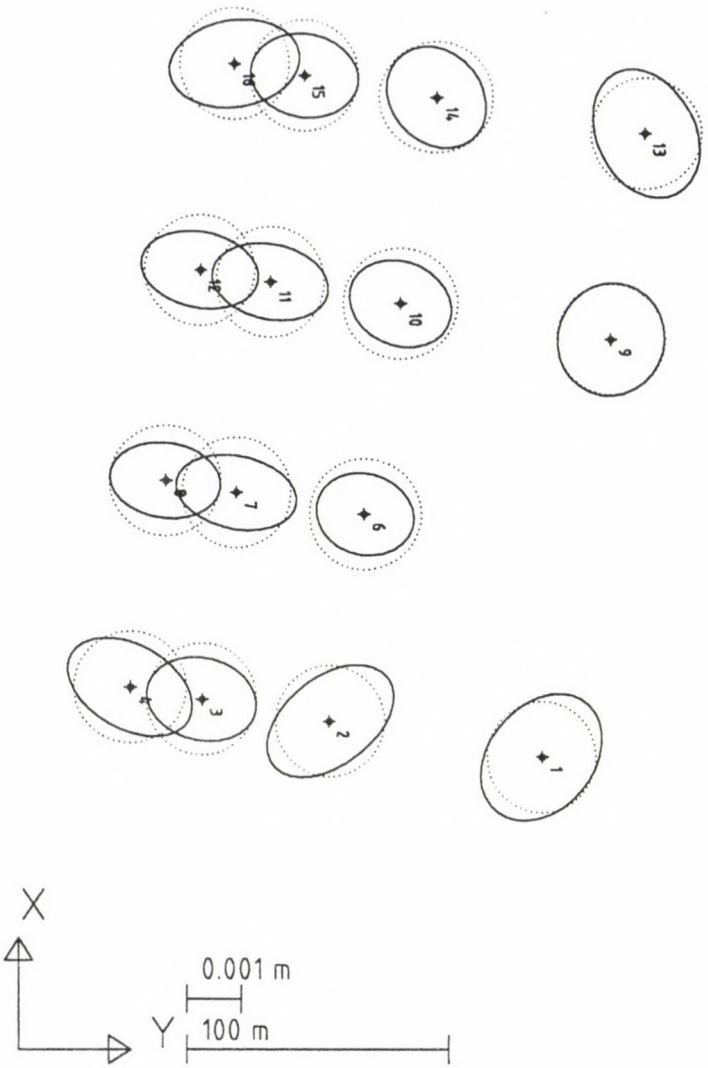


Fig. 3.

only trilateration networks. But the isotropy in point number five is worse than in the first case.

Because of these results, I omitted point number five, as it is shown in Fig. 3, which represents a trilateration network, and Fig. 4 for a triangulation network. In this case of a trilateration network, Fig. 3, the isotropy is missing for approximately all points, especially for points number one and thirteen. In Fig. 4 the situation is better because of using directions in addition to distances, where the isotropy is realised, but the precision is missing for points number one and thirteen.

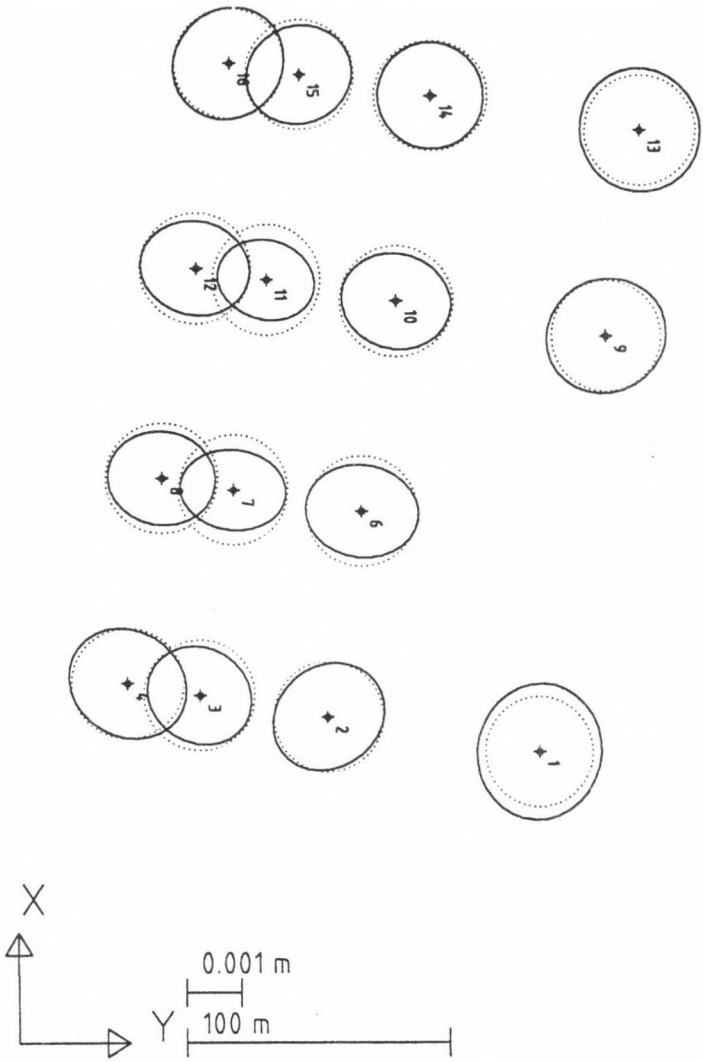


Fig. 4.

So, until now the best results are represented in Figs 2 and 4, but these results yet are not decisive to choose the only design of the network. I found another possibility to get a decision, where I took point number seventeen instead of point number five, and point number eighteen instead of point number four. I did that for two reasons; the first is that point (17) is close to point (5) and the point (18) is close to point (4), while the second reason is that points (17) and (18) are embedded in the project ground, so it is economical way to do this. The results for this design are represented in Fig. 5 for a characteristic distance equal to 25 m, and in Fig. 7

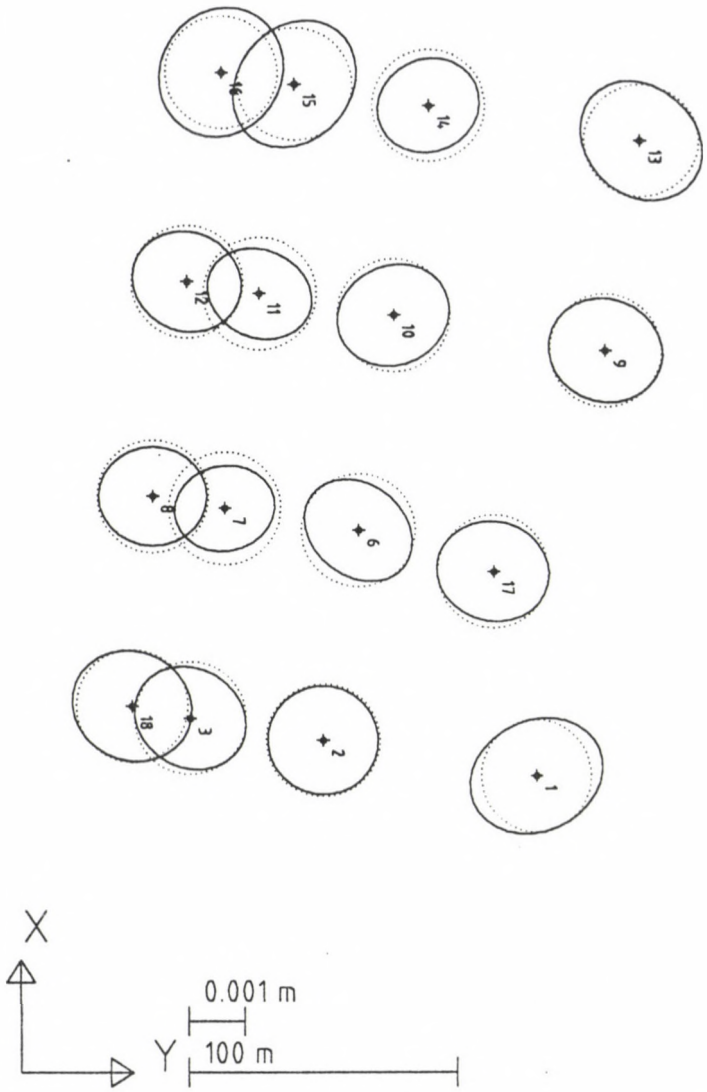


Fig. 5.

for a characteristic distance equal to 20 m. By comparing these results, I can say that the design number five is the best, which realises the homogeneity and isotropy of the network and a part of the economical situation by using the embedded points number (17) and (18).

In Table I, I have summarised some results of all these cases.

From this table I can say that for the optimal network, i.e. design number five, the original plan of observations is reduced by 36 percent, being a ratio which supports the efficiency of the solution method.

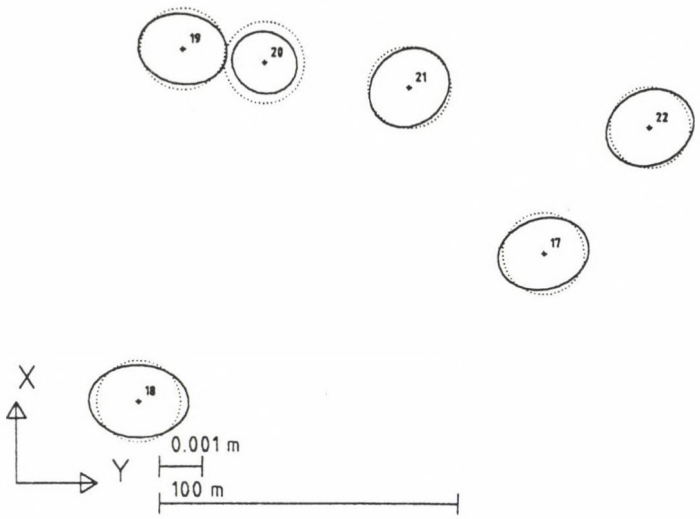


Fig. 6.

Table I.

Design number	1	2	3	4	5	6	7
The original number of distances	106	95	101	90	103	15	115
The original number of directions		210		202	232	30	232
The total number of observations	106	305	101	292	335	45	347
The final number of distances	99	58	95	57	57	14	89
The final number of directions		148		169	158	24	208
The total number of observations	99	206	95	226	215	38	297
The reduction ratio of the original observation plan	7%	32%	6%	23%	36%	16%	14%
The number of the redundant observations	70	165	68	184	174	24	252

Finally I have designed a monitoring network for the fault by using some points which now are embedded in the ground, so this case realises the economy criterion, especially that it has only 6 points. The results are represented in Fig. 6, the error situation is good, but only disadvantage in this network is that it covers a small area.

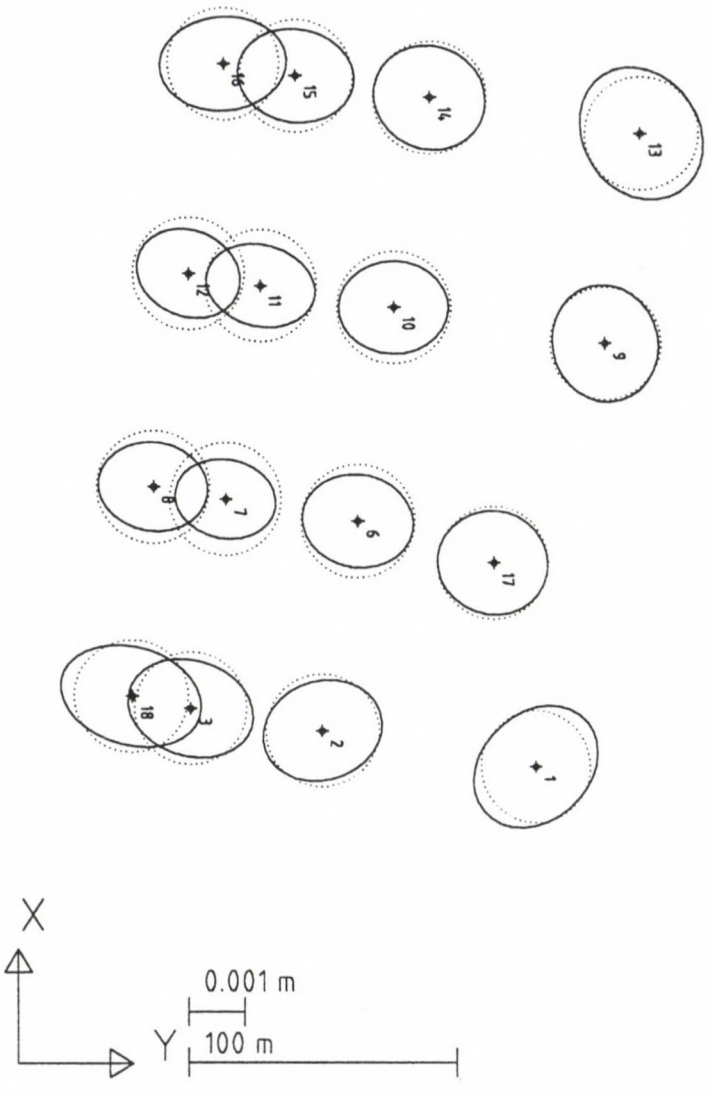


Fig. 7.

9. Conclusions

I have reached the optimal design of the monitoring network of the movements of the fault, which is represented in design number (5), i.e. Fig. 5. The design is realising the precision, the reliability, and the economy criteria.

From the results I found the way how the direction measurements support the distance measurements for realising the homogeneity and isotropy of the network, that is, the accuracy of the network, and these both types of observations control themselves concerning the reliability.

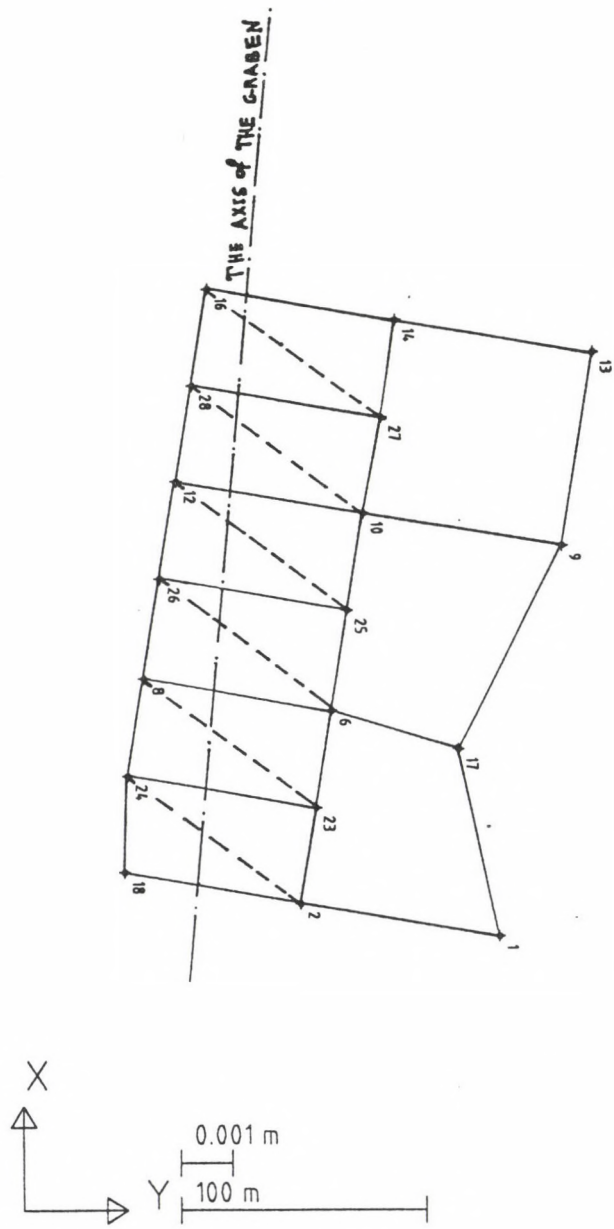


Fig. 8.

The model proposed to control the vertical fault movements uses some points from the designed horizontal network, that is, the same points serve for a horizontal monitoring network and also for the vertical network in the same time, which means an economical design.

Finally I advocate the most important result, which supports my choice of my solution method which is the modified least-squares approximation of an inverse criterion matrix, where I really obtained good and optimal design of my problem.

References

Baarda W 1973: S-transformations and criterion matrices. Netherlands Geodetic Commission, Vol. 5, No. 1.

Chen Y Q 1983: Analysis of deformation surveys – A generalised method. Department of Surveying Engineering, Technical Report No. 94, University of New Brunswick, Fredericton, Canada

Chrzanowski A, Chen Y Q, Secord J M 1983: *Tectonophysics*, 97, 297–315.

Chrzanowski A, Chen Y Q, Romero P, Secord J M 1986: *Tectonophysics*, 130, 369–383.

Grafarend E W 1972: Genauigkeitsmasse geodätischer Netze. Deutsche Geodätischer Kommission, Reihe A, Nr. 73.

Crosilla F 1983: *Bulletin Geodésique*, 57, 226–239.

Gaspary W 1987: Concepts of networks and deformation analysis. School of Surveying Monograph 11, University of New South Wales, Kensington, Australia

Heck B, Kock J J, Welsch W M, Baumer R, Chrzanowski A, Chen Y Q, Secord J M 1982: In: Report of the FIG working group on the analysis of deformation measurements. Proc. of the III International Symposium on Deformation Measurements. Publishing House of the Hungarian Academy of Sciences, Budapest, 373–415.

Kuang S L, Chrzanowski A, Chen Y Q 1991: *Manuscripta Geodaetica*, 16, 376–383.

Müller H, Illner M 1984: *Allg. Vermessungs-Nachrichten*, 253–269.

Müller H, Schmitt G 1985: SODES2 – Ein Programm System zur Gewichtsoptimierung zweidimensionaler geodätischer Netze. Deutsche Geodätische Kommission, München, Reihe B, No. 276.

Niemeier W, Rohde G 1982: On the optimization of levelling networks with respect to the determination of crustal movements. Deutsche Geodätische Kommission, Report, München, B–258/V, 148–160.

Rohde G 1980: Zur optimalen Gestaltung von Nivellementsnetzen für die Erfassung rezenter vertikaler Erdkrustenbewegungen. Universität Hannover, Diplomarbeit No. 911 (unpublished)

Schaffrin B 1983: *Geodesy*, 79, 43–56.

Schmitt G 1985: In: Optimization and design of geodetic networks. E Grafarend and Sanso eds, Springer Verlag, Berlin, 74–121.

Sprinsky W H 1978: *Bulletin Géodésique*, 52, 269–279.

Wimmer H 1981: Ein Beitrag zur Gewichtsoptimierung geodätischer Netze. Deutsche Geodätische Kommission, München, Rep. C 269.

Wimmer H 1982: In: Proc. of the International Symposium on Geodetic Networks and Computations of the International Association of Geodesy, München, 1981, 112–127.

Yazji S 1997: *Acta Geod. Geoph. Hung.* (present issue)

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 105–120 (1997)

A COMPUTER PROGRAM TO STUDY THE EFFECT OF TRANSITION LAYER PARAMETERS ON THE SCHLUMBERGER APPARENT RESISTIVITY

M BANERJEE[1], S K SINGH[1], NAGENDRA P SINGH[1], T LAL[1]

[Manuscript received February 12, 1997]

A computer program has been developed for the preparation of master curves of Schlumberger apparent resistivity function (ϱ_{as}/ϱ_1) versus the various values of half current electrode spacings over a horizontally stratified three-layered earth model having the intermediate interstratum as the transition layer. The interstratum has two types of conductivity variation: one, the linear variation, and second, the exponential variation of conductivity. The substratum has been taken as either highly resistive or conductive. The conductivity of the top layer merges with that of the substratum having a gradual variation of conductivity with depth in transition layer.

Keywords: apparent resistivity; conductivity variation; exponential variation; linear variation; transition layer

1. Introduction

In recent years some realistic representative earth models have been considered by a number of investigators (Mallick and Roy 1968, Lal 1970, Jain 1972, Sri Niwas and Upadhyay 1974, Mallick and Jain 1979, and Koefoed 1979), in which linear variation of conductivity with depth is analyzed. The geoelectrical situations where the conductivity of the interstratum varies exponentially with depth have been treated by many workers (e.g. Lal 1970, Stoyer and Wait 1977, Banerjee et al. 1980a, 1980b, and others). Then there are the situations where geological conditions do not permit the use of a single type of conductivity variation with a single transition layer and under such circumstances the geophysicists preferred to model with a combination of either uniform, or linear, or power law, or exponential law of variations (viz. Patella 1977, Raghuvanshi and Singh 1986, and others) over a multi-layered earth system.

In the present paper, a computer program based on Lal (1970) formulation for transition layer problem to compute the apparent resistivity master curves over three-layer earth with an inhomogeneous interstratum for the Schlumberger configuration with the maximum half current electrode spacing (AB/2) as 1200 metres, has been developed. Several models with the varying thickness of transition layer, and insulating and conducting substratum have been considered.

[1]Department of Geophysics, Banaras Hindu University, Varanasi–221005, India

1217-8977/97/$ 5.00

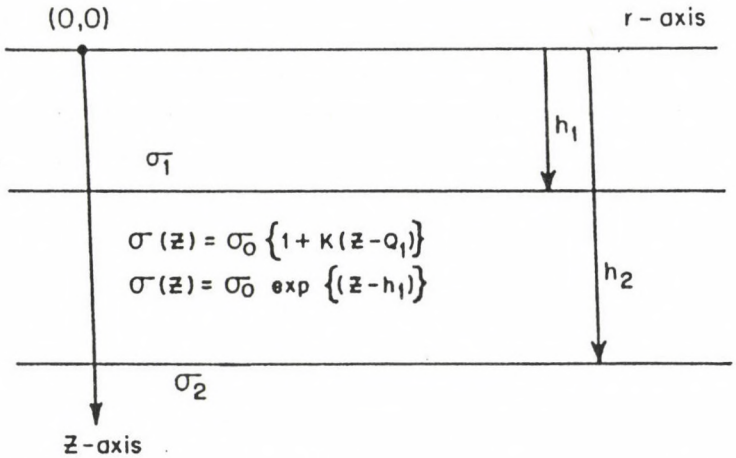


Fig. 1. The conductivity variation of the interstratum transition layer

2. Apparent resistivity for layered earth model with transition interstratum

Lal (1970) has considered the case of three layered earth, in more general way, by taking into account the different types of inhomogeneities in the conductivity of the interstratum, e.g. linear variation and exponential variation. The geometry of the system under investigation is shown in Fig. 1. The conductivity of the top layer, σ_1 and bottom layer, σ_2 are assumed to be constant, while the conductivity $\sigma(z)$ of the interstratum, which acts as transition layer is assumed to vary in accordance with the linear or exponential variation with depth, z, defined by the relations:

$$\sigma(z) = \sigma_1\{1 + \alpha(z - h_1)\} \tag{1}$$

and

$$\sigma(z) = \sigma_1 \exp\{\alpha(z - h_1)\}\,, \tag{2}$$

where α is the grade constant. The conductivity of the top layer gradually merges with that of the substratum through the interstratum transition layer.

2.1 Computer algorithm

The algorithms for the computation of Schlumberger apparent resistivity over a layered model with the transition layer having linear and exponential variations have been presented here.

A) *Linear variation*: The integration increment $\Delta\lambda$ has been assigned to `XX`; `TK` is the grade constant; `H1` and `Z` are the depths to the top and bottom boundaries of the transition layer as shown in Fig. 1.

```
XM = ABS(XX/TK)
XMK = XM*[1.0 + TK*(Z-H1)]
```

```
UU = SIGMAOD/SIGMA2
U = - (BESSKO(XMK) - BESSK1(XMK)*UU)/(BESSIO(XMK) +
* BESSK1(XMK)*UU)
SS = (BESSKO(XM) + U*BESSIO(XM)/(-BESSK1(XM) +   U*BESSI1(XM))
S = SIGMA1*SS/SIGMAO
BOD1 = (S + 1.0)/(S - 1.0)
D1 = BOD1*EXP(-2.0 + XX*H1)
FUNK = XX*BESSJ1(XX*RADR)*D1/(1.0 - D1)
GRND(K) = Σ (FUNC(K))
          K
```

The apparent resistivity is given as:

```
APPR(K) = 1.0 + 2.0*RADR(K)*RADR(K)*GRND(K)
```

where K is the particular value of half-current electrode distance, AB/2. SIGMAOD, SIGMA1 and SIGMA2 are the conductivities of the transition, top and bottom layers respectively.

B) *Exponential variation*: The integration increment, XX = $\Delta\lambda$.

```
TKE = (1.0/(Z-H1))*ALOG(SIGMAS/SIGMAO)
FSQ = (TKE*TKE + 4.0*XX*XX)**0.5
P1 = ( - TKE + FSQ)/2.0
P2 = ( - TKE - FSQ)/2.0
F1NM = (XX + P1*SIGMAS/SIGMA2)/(XX + P2*SIGMAS/SIGMA2)
CON10 = SIGMA1/SIGMA0
FEXP1 = EXP(P1*H1)
FEXP2 = EXP(P2*Z)
FEXP12 = EXP(Z*(P1-P2))
TNUM = FEXP1 - F1NM*FEXP2*FEXP12
DENOM = P1*FEXP1 - P2*F1NM*FEXP2*FEXP12
BETA = CON10*TNUM/DENOM
BOD1 = ((XX*BETA + 1.0)/(XX*BETA - 1.0)
D1 = BOD1*EXP(-2.0*XX*H1)
FUNC = (XX*BESSJ1(XX*RADR)*D1)/(1.0 - D1)
GRND(K) = Σ (FUNC(K))
          K
```

The apparent resistivity for a three-layered earth with a transition layer having exponential variation is given as:

```
APPR(K) = 1.0 + 2.0*RADR(K)* RADR(K)* GRND(K)
```

2.2 Computer program

The computer program, RESTRAN has been developed to compute the theoretical Schlumberger master curves to study the effects of the transition layer parameters. The flow chart for the linear or exponential variations have been shown in Fig. 2. The RESTRAN selects one of the these two variations, after initialization of the input parameters, through the logical IF statement, i.e. when NXX assumes

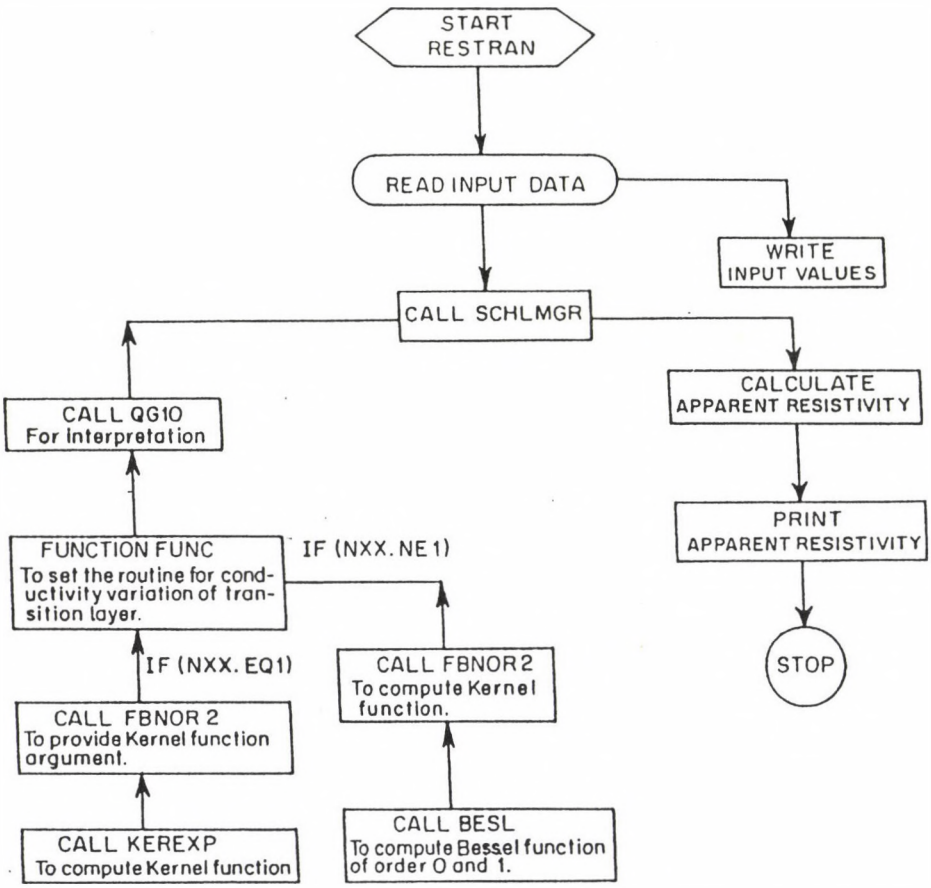


Fig. 2. Flow chart of RESTRAN for the computation of Schlumberger resistivity master curves for linear or exponential variation of conductivity

an integer value other than 1 it operates for linear variation and when **NXX = 1**, it operates for exponential variation. RESTRAN is comprising the subroutines as **SCHLMGR**, **QG10**, **BESL**, **FBNOR1**, **FBNOR2**, **KEREXP** and a function routine, **FUNC**. In addition, the program calls the standard subroutine to compute Bessel functions **BESSJ0**, **BESSJ1**, **BESSK0** and **BESSK1**. A listing of RESTRAN is provided in the Appendix.

3. Numerical results and discussions

The Schlumberger master curves for the three-layered earth with the intermediate transitional layer have been presented for the following four different cases:

a) Linear variation with highly

(i) insulating substratum, and

(ii) conducting substratum; and

b) Exponential variation with

(i) insulating substratum, and

(ii) conducting substratum.

3.1 Linear variation with insulating substratum

In this case the intermediate layer is an inhomogeneous transitional medium obeying the variational function of conductivity as $\sigma(z) = \sigma_1\{1+\alpha(z-h_1)\}$; where α is constant. For the preparation of the master curves, the values chosen for the various parameters are: the top layer resistivity, $\varrho_1 = 20$ ohmm; the bottom layer resistivity, $\varrho_2 = 1000$ ohmm; $h_1 = 20$ m, and $m = (z - h_1)/h_1 = 1$ to 20 as indicated in Fig. 3. In the figure, the function ϱ_{as}/ϱ_1 is plotted against the half-current electrode spacing, AB/2. The set of master curves in Fig. 3 indicates that as the thickness of the transition layer increases, the apparent resistivity decreases and does not attain the resistivity of substratum even for a very large value of the half-current electrode spacing, AB/2, i.e. 1200 m.

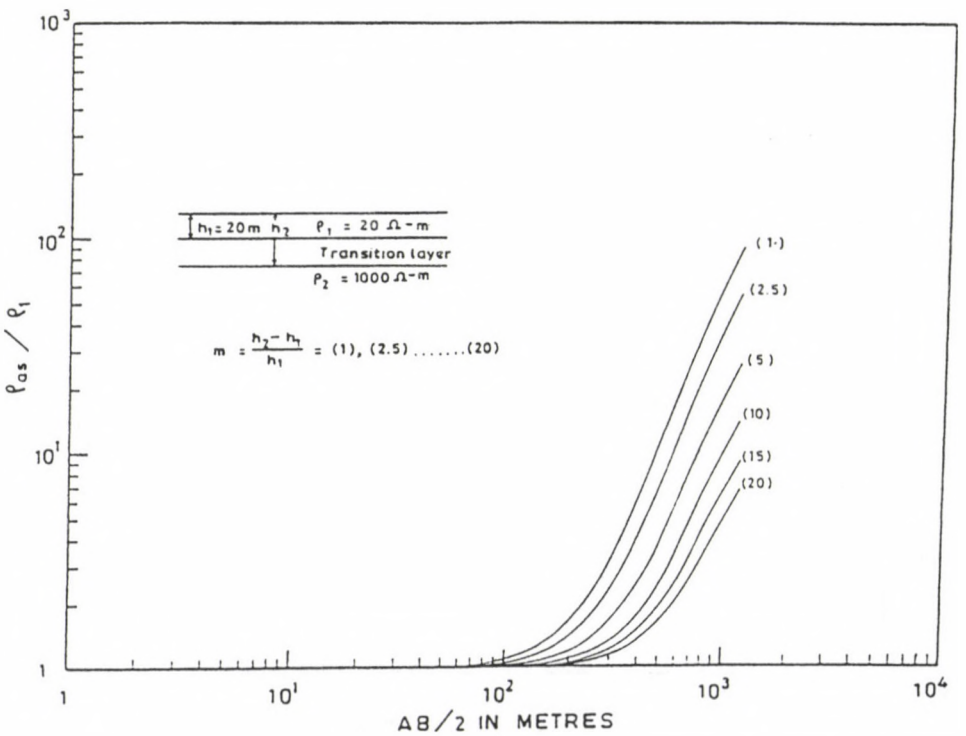


Fig. 3. The Schlumberger master curves with linear variation of conductivity of transition layer for insulating substratum ($\varrho_2 = 1000$ ohmm)

3.2 Linear variation with conducting substratum

In this case also the intermediate transitional layer conductivity variation follows the relation $\sigma(z) = \sigma_1\{1 + \alpha K(z - h_1)\}$; with a constant value of α. The Schlumberger master curves have been prepared for the model with the parameters: $\varrho_1 = 20$ ohmm; $\varrho_2 = 0.1$ ohmm; $h_1 = 20$ m, and $m = 0, 2.5, 5, 10, 15, 20$ and 25 which have been presented in Fig. 4. The figure shows a rather completely different picture, whereupon the curves have a tendency to diverge after their emanation and then again converge at a very large value of AB/2. It is noticed that the curve with $m = 0$ (i.e., in the absence of intermediate transition layer), the apparent resistivity values attain a constant value beyond AB/2 = 250 m; whereas the apparent resistivity master curve with $m = 250$ m (i.e. when the transition layer thickness is 500 m) falls off gradually to a very low value of resistivity only to approach a point of convergence (see Fig. 4).

3.3 Exponential variation with insulating substratum

In this case, the three-layer earth model is having the intermediate substratum as the transitional medium. The transitional layer is obeying the law of conductivity variation as $\sigma(z) = \sigma_1 \exp\{\alpha(z - h_1)\}$; where α is constant and h_1 and z are the

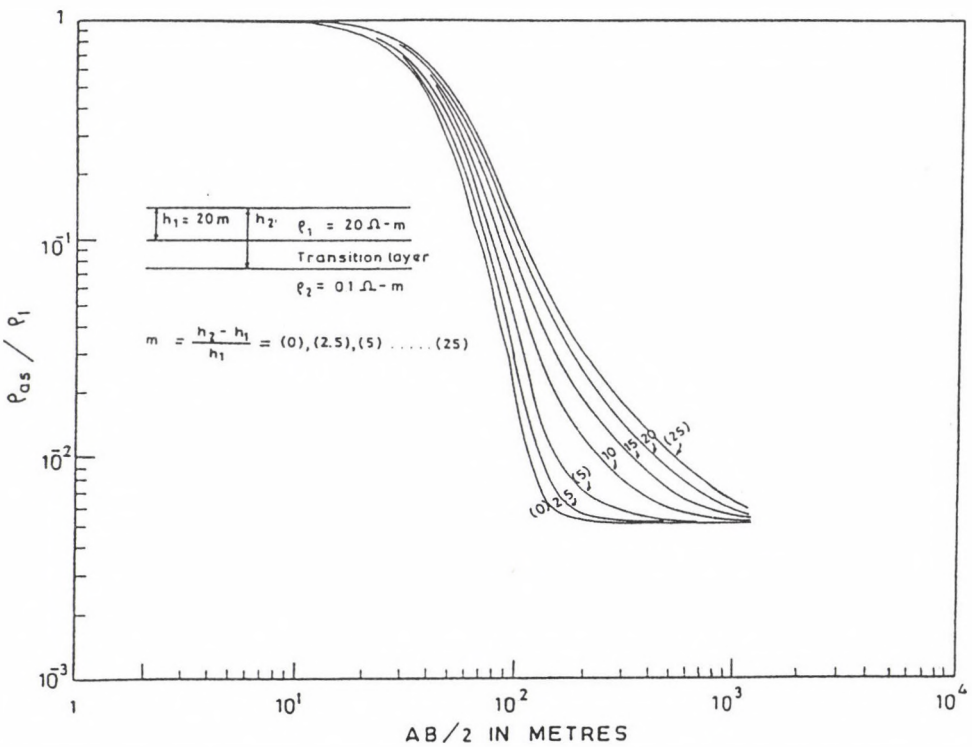


Fig. 4. The Schlumberger master curves with linear variation of conductivity of transition layer for conducting substratum ($\varrho_2 = 0.1$ ohmm)

depth to the top and bottom boundaries of the transition layer. The computer algorithm of section 2.1 (B) has been used to compute ϱ_{as}/ϱ_1 for varying thickness of transition layer and for the same values of ϱ_1 and ϱ_2. The master curves for the models having $\varrho_1 = 10$ ohmm; $\varrho_2 = 1000$ ohmm; $h_1 = 10$ m and $m = 1, 6, 10, 20, 30, 40$ and 50 have been presented in Fig. 5. The curves show that with the increase of the thickness of the transitional layer, the apparent resistivity decreases. The curves have diverging tendency as AB/2 is increased, but one expected to converge perhaps at a very large value of AB/2 (see Fig. 5). Another three layer setup having the layering parameters as $\varrho_1 = 20$ ohmm; $\varrho_2 = 1000$ ohmm; $h_1 = 20$ m and the master curves corresponding to $m = 0$, 1, 2.5, 5, 15 and 20 have been shown in Fig. 6. The characteristics of the curves is such that as the thickness of the intermediate transitional layer is increased, the Schlumberger apparent resistivity curves deplete in value and with the increase of substratum resistivity value, the set of curves shift to right thereby indicating more values for the top layer thickness, h_1 during interpretation.

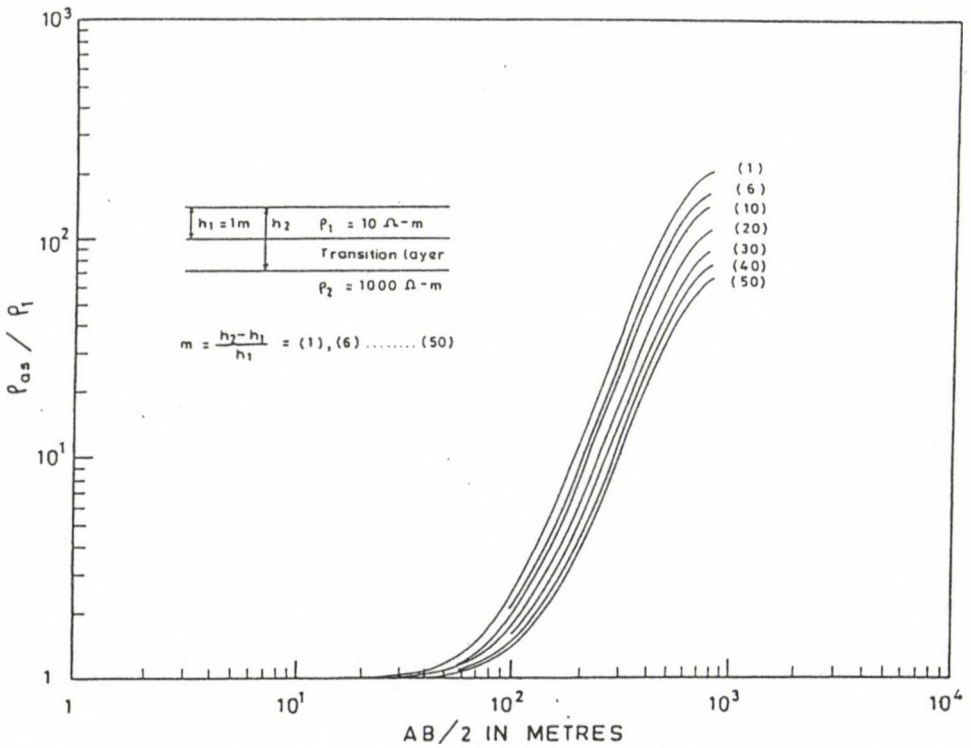


Fig. 5. The Schlumberger master curves with exponential variation of conductivity of transition layer for insulating substratum ($\varrho_2 = 1000$ ohmm)

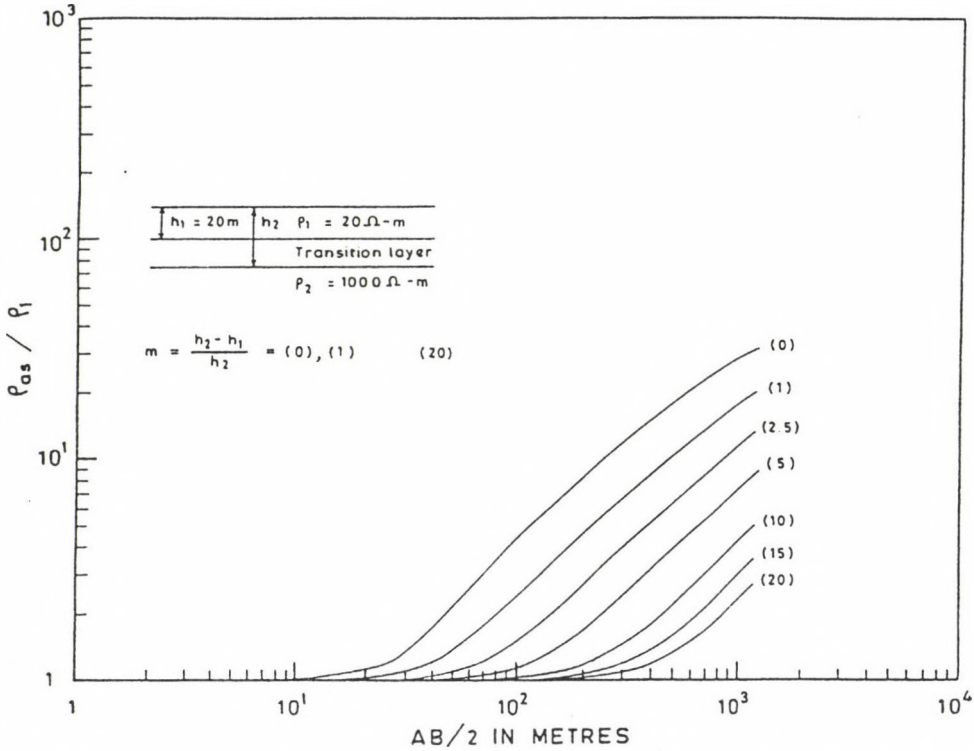


Fig. 6. Another case of the Schlumberger master curves with exponential variation of conductivity of transition layer for insulating substratum ($\varrho_2 = 1000$ ohmm)

3.4 Exponential variation with conducting substratum

As in the case of insulating model, the transitional layer conductivity variation follows the same law. The Schlumberger apparent resistivity master curves have been prepared for the model with parameters: $\varrho_1 = 20$ ohmm; $\varrho_2 = 0.1$ ohmm; and $m = 0.75$, 1.5, 2, 3, 4 and 5 which have been plotted in Fig. 7. The figure depicts that the master curves are converging at a very large value of the half-current electrode spacing, AB/2. In this case also it has been observed that the apparent resistivity values decrease with the increase of the transition layer thickness.

4. Conclusions

A computer program has been developed which successfully generates the Schlumberger master curves for the three-layered earth with an inhomogeneous interstratum having (i) a linear variation of conductivity with insulating substratum, (ii) a linear variation of conductivity with conducting substratum, (iii) an exponential variation of conductivity with insulating substratum, and (iv) an exponential variation of conductivity with conducting substratum.

A general inspection of the master curves for the above cases indicate that, (a)

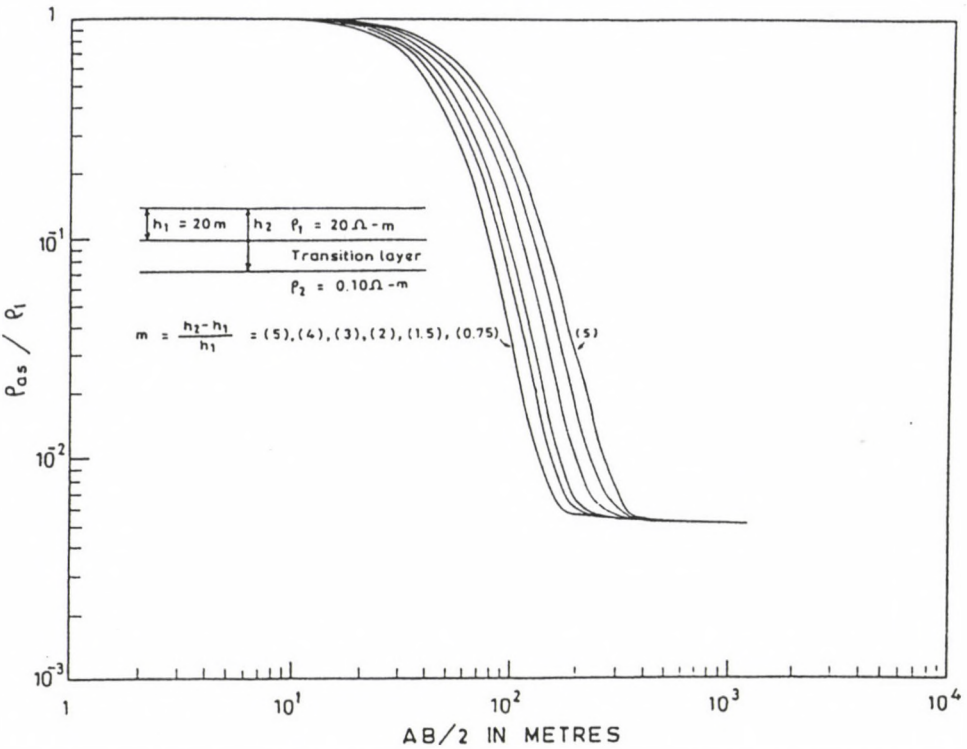


Fig. 7. The Schlumberger master curves with exponential variation of conductivity of transition layer for conducting substratum ($\varrho_2 = 0.1$ ohmm)

for $\varrho_2/\varrho_1 > 1$, an increase in the thickness of transition layer lowers the apparent resistivity values; and (b) for $\varrho_2/\varrho_1 < 1$, an increase in the thickness of transition layer enhances the apparent resistivity values. Also as the bottom layer resistivity is increased for the models with linear or exponential variation of conductivity of transition layer, the curves tend to shift towards right along abscissa axisthereby giving a higher estimate for the top layer thickness.

Acknowledgements

The financial support received from the minor research project, University Grant Commission, New Delhi is highly acknowledged. One of the authors (NPS) is grateful to the CSIR, New Delhi for providing the assistance in the form of Research Associateship.

Appendix

```
      PROGRAM RESTRAN
C     RESTRAN COMPUTES THE SCHLUMBERGER RESISTIVITY MASTER CURVES
C             OF A THREE-LAYERED EARTH SYSTEM WITH AN INTERMEDIATE
C             INHOMOGENEOUS INTERSTRATUM.
C-----------------------------------------------------------------
C     SUBROUTINE REQUIRED: SCHLMGR,QG10,BESL,FBNOR1,FBNOR2,KEREXP,
C                          AND FUNCTION ROUTINE FUNC;
C     LIBRARY ROUTINES REQUIRED: BEESJ0,BESSJ1,BESSK0 AND BESSK1.
C-----------------------------------------------------------------
C     INPUT REQUIRED:
C               H1,H2,RHO1,RHO2,RHO,RHOD,NORDER,NXX
C               XCOEFF,(RADR(1), I=1,30)
C     OUTPUT:   APPARENT RESISTIVITY VERSUS HAL-CURRENT ELECTRODA
C               SPACING (AB/2).
C-----------------------------------------------------------------
      COMMON NREAD,NPRT,NPLOT,RHO1,RHO2,RHO,RHOD,H1,H2,NCRDER,NXX
      COMMON SIGMA0,SIGMA1,SIGMA2,SIGMAOD
      COMMON/TWO/RADR1
      CHARACTER*8 FILE1,FILE2,FILE3
      DIMENSION GRND(30),RADR(30),APPR(30)
C
C   TO COMPUTE EXPONENTIAL VARIATION ASSIGN 1 TO NXX
C   OTHERWISE ASSIGN 0 TO NXX FOR LINEAR VARIATION.
C   THE PROGRAM IS WRITTEN BY DR. M. BANERJEE, D/GEOPHYSICS, B.H.U.
      WRITE(*,*) 'TYPE IN 8 CHARACTERS THE NAME OF INPUT FILE1='
      READ(*,1) FILE1
      WRITE(*,*) 'TYPE IN 8 CHARACTERS THE NAME OF OUTPUT FILE2='
      WRITE(*,'(A\)')' The Program AIMRESI developed by'
      READ(*,1) FILE2
      WRITE(*,*) 'TYPE IN 8 CHARACTERS THE NAME OF PLOT FILE3='
      READ(*,1) FILE3
    1 FORNAT(1A8)
      WRITE(*,1) FILE1,FILE2,FILE3
      OPEN(5,FILE=FILE1,STATUS='OLD')
      OPEN(6,FILE=FILE2,STATUS='NEW')
      OPEN(7,FILE=FILE3,STATUS='NEW')
C
      NREAD=5
      NPRT=6
      NPLOT=7
C
      READ(NREAD,5) H1,H2,RHO1,RHO2,RO,RHOD,NORDER,NXX
      READ(NREAD,5) XCOEFF,IMM
      READ(NREAD,*) (RADR(I), I=1,30)
C
    5 FORMAT(6F10.4,2I5)
    6 FORMAT(F10.6,I5)
```

```
      WRITE(NPRT,7) H1,H2,RHO1,RHO2,RHO,RHOD,NORDER,NXX
    7 FORMAT(5X,'H1=',F10.4,' H2=',F10.4,' RHO1=',F10.4,' RHO2=',
     *F10.4,' RHO=',F10.4,' RHOD=',F10.4,' NORDER=',I5,' NXX=',I5//)
      WRITE(NPRT,8) XCOEFF,IMM
    8 FORMAT(5X,' XCOEFF=',F10.6,' IMM=',I5//)
C
      SIGMA0D=1.0/RHOD
      SIGMA1=1.0/RHO1
      SIGMA2=1.0/RHO2
      SIGMA0=1.0/RHO
      SIGR=SIGMA2/SIGMA1
C
      IF(((H2-H1).EQ.0.0).OR.(NORDER.EQ.0)) GO TO 11
      ORDER=1.0/FLOAT(NORDER)
      TK=((SIGMA0D/SIGMA0)**ORDER-1.0)/(H2-H1)
   11 IF(((H2-H1).EQ.0.0).OR.(NORDER.EQ.0)) TK=0.0
      SIGMA0D=SIGMA0*((1.0+TK*(H2-H1))**NORDER)
      RHOD=1./SIGMA0D
      SIGR1=SIGMA0/SIGMA1
      SIGR2=SIGMA2/SIGMA0D
      SIGH1=(H2-H1)/H1
      CALL WRTM1(1,SIGR,SIGR1,SIGR2,SIGH1)
      CALL SCHLMGR(RADR,XCOEFF,IMM)
      STOP
      END
C
      SUBROUTINE SCHLMGR(RADR,XCOEFF,IMM)
C
C-----------------------------------------------------------------
C  SUBROUTINE SCHLMGR COMPUTES SCHLUMBERGER APPARENT RESISTIVITY
C  OVER A THREE LAYER EARTH MODEL WITH A TRANSITION LAYER HAVING
C  VARIABLE LAYER PARAMETERS
C-----------------------------------------------------------------
C
      COMMON NREAD,NPRT,NPLOT,RHO1,RHO2,RHO,RHOD,H1,H2,NORDER,NXX
      COMMON SIGMA0,SIGMA1,SIGMA2,SIGMA0D
      COMMON/TWO/RADR1
      DIMENSION GRND(30),RADR(30),APPR(30)
C
      DATA N/30/
C
C    XCOEFF IS GENERALLY GIVEN THE VALUE AS 0.3 OR 7.0
      XBG=0.00000001
      CALL QG10(XBG,XCOEFF,GRND,RADR,IMM)
      DO 10 K=1,N
C   APPR IS THE RATIO OF APP. RES. AND THE RHO1
      APPR(K)=1.0+2.0*RADR(K)*RADR(K)*GRND(K)
   10 CONTINUE
      WRITE(NPRT,20)
```

```
   20 FORMAT(1H1,11X,'DISTANCE R',10X,'APPR. RES.',10X,'INTEGRAND'/)
      00 30 I=1,N
      WRITE(NPRT,25) RADR(I),APPR(I),GRND(I)
      WRITE(NPLOT,26) RADR(I),APPR(I)
   25 FORMAT(10X,E10.4, 5X,E16.9, 5X,E16.8)
   26 FORMAT(2F10.4)
   30 CONTINUE
      RETURN
      END
C
      SUBROUTINE QG10(XL,XU,GRND,RADR,IMM)
C     FOR INTEGRATION
      EXTERNAL FUNC
      COMMON NREAD,NPRT,NPLOT,RHO1,RHO2,RHO,RHOD,H1,H2,NORDER,NXX
      COMMON SIGMA0,SIGMA1,SIGMA2,SIGMA0D
      COMMON/TWO/RADR1
      DIMENSION RADR(30),GRND(30),APPR(30),X(16),W(16)
      DATA N/30/
      MMB=17
      DO 20 I=1,N
      RADR1=RADR(I)
C     A=0.5*(XU+XL)
C     B=XU-XL
C     C=0.4869533*B
C     Y=0.03333567*(FUNC(A+C)+FUNC(A-C))
C     C=0.4325317*B
C     Y=Y+0.07472567*(FUNC(A+C)+FUNC(A-C))
C     C=0.3397048*B
C     Y=Y+0.1095432*(FUNC(A+C)+FUNC(A-C))
C     C=0.2166977*B
C     Y=Y+0.1346334*(FUNC(A+C)+FUNC(A-C))
C     C=0.07443717*B
C     Y=B*(Y+0.1477621*(FUNC(A+X)+FUNC(A-C)))
C     YY=Y
C    NM SHOULD EQUAL TO 1000 TO 7000
      MM=IMM
      XINC=(XU-XL)/FLOAT(MM)
      XMB=XL
      SUM=0.0
      DO 10 II=1,MM+1
      XII=FLOAT(II)
      XMB=XMB+(XII-1.0)*XINC
      SUM=SUM+FUNC(XMB)
      XMB=XL
   10 CONTINUE
      SUM=XINC*SUM
      YY=SUM
C     IF(I.EQ.3) GO TO 21
   20 GRND(I)=YY
```

```
   21 RETURN
      END
      FUNCTION FUNC(XXT)
C     TO SET THE CONDUCTIVITY VARIATION OF THE TRANSITION LAYER,
C     AND TO COMPUTE THE KERNEL FUNCTION
      COMMON NREAD,NPRT,NPLOT,RHO1,RHO2,RHO,RHOD,H1,H2,NORDER,NXX
      COMMON SIGMA0,SIGMA1,SIGMA2,SIGMA0D
      COMMON/TWO/RADR1
      DIMENSION GRND(30),RADR(30),APPR(30)
      XMB1=XXT*RADR1
      CALL BESSJ1(XMB1,VALJ1)
      XX=XXT
C     WRITE(*,*) 'XX=',XX,'RADR1,' XMB1=',XMB1,' VALJ1=',VALJ1
      IF((H2-H1).EQ.0.0) GO TO 10
      IF(NXX.EQ.1) THEN
              CALL KEREXP(BETA,XX)
              BOD1=(XX*BETA+1.0)/(XX*BETA-1.0)
      ELSE
              CALL BESL(XX,S)
              BOD1=(S+1.0)/(S-1.0)
C             WRITE(*,*) 'BOD1= ',BOD1
      ENDIF
      D1=BOD1*EXP(-2.0*XX*H1)
      FUNC=XX*VALJ1*D1/(1.0-D1)
   10 CONTINUE
C     WRITE(*,*) 'FUNC=',FUNC,' D1=',D1,' BOD1=',BOD1,' XX=',
C    *XX,' H1=',H1
      RETURN
      END
C
      SUBROUTINE BESL(XX,S)
C     TO SOLVE THE BESSEL FUNCTION FOR LINEAR VARIAITON
      COMMON NREAD,NPRT,NPLOT,RHO1,RHO2,RHO,RHOD,H1,H2,NORDER,NXX
      COMMON SIGMA0,SIGMA1,SIGMA2,SIGMA0D
      DIMENSION GRND(30),RADR(30),APPR(30)
      IF(((H2-H1).EQ.0.0).OR.(NORDER.EQ.0)) GO TO 41
      TK=((SIGMA0D/SIGMA0)**ORDER-1)/(H2-H1)
   41 IF(((H2-H1).EQ.0.0).OR.(NORDER.EQ.0)) THEN
         XM=XX
         XMMK=XM
      ELSE
         XM=ABS(XX/TK)
         XMMK=XM*(1.0+TK*(H2-H1))
      ENDIF
      XMK=ABS(XMMK)
      NORD=1+NORDER
      GO TO (10,20),NORD
   10 CALL FBNOR1(XM,XMK,S)
      RETURN
```

```
   20 CALL FBNOR2(XM,XMK,S)
      RETURN
      END
C
      SUBROUTINE FBNOR2(XM,XMK,S)
C     TO COMPUTE KERNEL FUNCTION
C
C-----------------------------------------------------------------
C  FOR (N-1)/2 IS EITHER ZERO OR INTEGER VALUE
C
      COMMON NREAD,NPRT,NPLOT,RHO1,RHO2,RHO,RHOD,H1,H2,NORDER,NXX
      COMMON SIGMA0,SIGMA1,SIGMA2,SIGMA0D
      DIMENSION GRND(30),RADR(30),APPR(30)
      CALL BESSI0(XMK,VALUE5)
      CALL BESSI1(XMK,VALUE6)
      CALL BESSK0(XMK,VALUE7,VALUE5)
      CALL BESSK1(XMK,VALUE8,VALUE6)
C
      UU=SIGMA0D/SIGMA2
      U=-(VALUE7-VALUE8*UU)/(VALUE5+VALUE6*UU)
C
      CALL BESSI0(XM,VALUE3)
      CALL BESSI1(XM,VALUE4)
      CALL BESSK0(XM,VALUE1,VALUE3)
      CALL BESSK1(XM,VALUE2,VALUE4)
C
      SS=(VALUE1+U*VALUE3)/(-VALUE2+U*VALUE4)
      S=SIGMA1*SS/SIGMA0
      END
C
      SUBROUTINE FBNOR1(XM,XMK,S)
C     TO PROVIDE KERNEL FUNCTION ARGUMENT
      COMMON NREAD,NPRT,NPLOT,RHO1,RHO2,RHO,RHOD,H1,H2,NORDER,NXX
      COMMON SIGMA0,SIGMA1,SIGMA2,SIGMA0D
      DIMENSION GRND(30),RADR(30),APPR(30)
      UU=SIGMA0D/SIGMA2
      CALL BESSI0(XMK,VAL2)
      CALL BESSI1(XMK,VAL4)
      CALL BESSK0(XMK,VAL1,VAL2)
      CALL BESSK1(XMK,VAL3,VAL4)
      U=-(VAL1-VAL3*UU)/(VAL2+VAL4*UU)
      CALL BESSI0(XM,VAL8)
      CALL BESSI1(XM,VAL6)
      CALL BESSK0(XM,VAL7,VAL8)
      CALL BESSK1(XM,VAL5,VAL6)
      SS=(VAL7+U*VAK8)/(-VAL5+U*VAL6)
      S=SIGMA1*SS/SIGMA0
      RETURN
      END
```

```
C
      SUBROUTINE KEREXP(BETA,XX)
C     TO SOLVE THE KERNEL FUNCTION FOR THE EXPONENTIAL CONDUCTIVITY
C     VARIATION OF TRANSITION LAYER
      COMMON NREAD,NPRT,NPLOT,RHO1,RHO2,RHO,RHOD,H1,H2,NORDER,NXX
      COMMON SIGMA0,SIGMA1,SIGMA2,SIGMA0D
      DIMENSION GRND(30),RADR(30),APPR(30)
      SIGMAS=SIGMA0D
      IF((H2-H1).GT.0.0) THEN
          TKE=(1.0/(H2-H1))*ALOG(SIGMAS/SIGMA0)
      ELSE
          TKE=99999.
      ENDIF
          FSQ=(TKE*TKE+4.0*XX*XX)**0.5
      P1=(-TKE+FSQ)/2.0
      P2=(-TKE-FSQ)/2.0
      F1NM=(XX+P1*SIGMAS/SIGMA2)/(XX+P2*SIGMAS/SIGMA2)
      CON10=SIGMA1/SIGMA0
      FEXP1=EXP(P1*H1)
      FEXP2=EXP(P2*H1)
      FEXP12=EXP(H2*(P1-P2))
      TNUM=FEXP1-F1NM*FEXP2*FEXP12
      DENOM=P1*FEXP1-P2*F1NM*FEXP2*FEXP12
      BETA=CON10*TNUM/DENOM
      RETURN
      END
C
      SUBROUTINE WRTM1(ISET,SIGR,SIGR1,SIGR2,SIGH1)
      COMMON NREAD,NPRT,NPLOT,RHO1,RHO2,RHO,RHOD,H1,H2,NORDER,NXX
      COMMON SIGMA0,SIGMA1,SIGMA2,SIGMA0D
      DIMENSION GRND(30),RADR(30),APPR(30)
      WRITE(NPRT,100)
  100 FORMAT(///,32X,16(1H-))
      WRITE(NPRT,101)
  101 FORMAT(32X,1H*,14X,1H*)
      WRITE(NPRT,102) ISET
C     WRITE(NPLOT,102) ISET
  102 FORMAT(32X'* < SET ',I4,' > *')
      WRITE(NPRT,101)
      WRITE(NPRT,103)
      IF(NXX.NE.1) GO TO 40
      IF(NXX.EQ.1) WRITE(NPRT,38)
   38 FORMAT(25X,'EXPONENTIAL VARIATION IN TRANSITION LAYER'/)
   40 IF(NXX.EQ.1) GO TO 44
      IF(NORDER.EQ.1) WRITE(NPRT,42)
   42 FORMAT(20X,'POWER, N=1: LINEAR VARIATION IN',
     *' TRANSITION LAYER'/)
   44 CONTINUE
      IF(SIGR.LT.1.0) WRITE(NPRT,104) SIGR
```

```
      IF(SIGR.GE.1.0) WRITE(NPRT,105) SIGR
      WRITE(NPRT,30) RHO, RHO1,RHO2,H1,H2,SIGR,SIGR1,SIGR2,RHOD,
     *SIGH1,NORDER,NXX
   30 FORMAT(10X,'RHO=',E9.2,2X,'RHO1=',E9.2,2X,'RHO2=',E9.2,
     12X,'H1=',F6.2,2X,'H2=',F6.2,2X,'SIGR=',E9.2,2X,'SIGR1=',
     2E9.2,2X,'SIGR2=',F6.2/9X,'RHOD=',F6.2,2X,'SIGH1='F6.2,2X,
     3'NORDER=',I5,2X,'NXX=',I5)
  104 FORMAT(25X,'RESISTIVE SUBSTRATUM',2X,'SIGR=',E9.2)
  105 FORMAT(25X,'CONDUCTIVE SUBSTRATUM',2X,'SIGR=',E9.2)
      WRITE(NPRT,103)
  103 FORMAT(2X,80('*'))
      RETURN
      END
C           END OF PROGRAMME "RESTRAN.F".
C-----------------------------------------------------------------
```

Rereferences

Banerjee B, Sengupta B J, Pal B P 1980a: *Geophysical Prospecting*, 28, 435–452.
Banerjee B, Sengupta B J, Pal B P 1980b: *Geophysical Prospecting*, 28, 750–758.
Jain S C 1972: *Geophysical Prospecting*, 20, 283–292.
Koefoed O 1979: *Geophysical Prospecting*, 27, 862–868.
Lal T 1970: *Pure and Applied Geophysics*, 82, 259–269.
Mallick K, Jain S C 1979: *Geophysical Prospecting*, 27, 869–875.
Mallick K, Roy A 1968: *Geophysical Prospecting*, 16, 436–446.
Patella D 1977: *Geophysical Prospecting*, 25, 699–729.
Raghuvanshi S S, Singh B 1986: *Geophysical Prospecting*, 34, 409–423.
Sri Niwas, Upadhyay S K 1974: *Geophysical Prospecting*, 22, 279–296.
Stoyer C H, Wait J R 1977: *Geoexploration*, 15, 11–18.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 121–123 (1997)

PIONEERS OF POLAR AURORAE IN HUNGARY 1768–1910 AND SPECTRAL RESEARCH AT OBSERVATORY ÓGYALLA BY KONKOLY-THEGE*

L Křivský[1]

[Manuscript received March 10, 1997]

The early history of auroral research in Hungary and especially at the observatory Ógyalla-Hurbanovo is reviewed, emphasizing Konkoly-Thege's spectroscopic research.

Keywords: aurora history; Hurbanovo Observatory; Konkoly-Thege M; Ógyalla Observatory

At the astrophysical Observatory Ógyalla, Miklós Konkoly-Thege continued with starting in 1870 the tradition of observation and research of polar aurorae of Hungarian predecessors who had achieved extraordinary success in this branch even in comparison on international level of that time.

It was Miksa Hell who headed together with János Sajnovics the expedition to Norway (1769) and later became imperial astronomer in Vienna, and Pál Makó von Kerekgede who published in German and in Latin works on polar aurorae (1775, 1781). Makó calculated distances of polar aurorae but the results were overestimated. Makó came very near to the Frenchman J D de Marian (1773) with results from the following quotation: "Northern light is an electric spark which is created by the contact of the terrestrial and solar atmospheres". By that time, this to great extent intuitive explanation of the interconnection of the Earth's atmosphere with that of the Sun corresponds with the most modern scientific knowledge on diffusion of magnetoplasma clouds from solar eruptions via corona and heliosphere to the magnetosphere and the outer atmosphere of the Earth.

Taking into consideration war and revolutionary events the interest in polar aurorae has not been renewed until after 1860. It was Guido Schenzl (1869) who, by that time, examined the variations of the geomagnetic field and who equally observed several polar aurorae; later Miklós Konkoly-Thege concentrated his attention to observations of polar aurorae with a spectroscope at the Observatory Ógyalla (1870). Further Lorand Eötvös (1871) investigated the spectrum of aurorae and defended the idea of their electric origin, too.

The observation efforts aimed at gaining unique water-coloured paintings of polar aurorae, starry sky and terrain cannot be skipped. Their author Joh Nep Fierer in Trnava was a member of Jesuit Order. He made observations in the years 1768–1778 and his pictures were published in a monography by Réthly and Berkes (1963).

*Presented at the EGS XXII. General Assembly in Vienna, April 21-25, 1997

[1]Astronomical Institute of the Czech Academy of Sciences, Ondřejov

1217-8977/97/$ 5.00

The first description of polar aurorae by Konkoly at Ógyalla was made on the 25 October 1870, he signed by error the year 1871. It was published in Beobachtungen 1879 (p. 34 and 36) as well as in Hessische Wochenschrift (1870 p. 395). By this extensive description of that phenomenon as well as by observation with spectroscope Ógyalla entered into world catalogues of polar aurorae (Fritz 1873, Réthly and Berkes 1963, Křivský and Pejml 1985, Křivský 1993–1996). One of the first spectral observations is being quoted in detail as follows: "I studied the phenomenon during its full splendour with the spectral apparatus. Two light lines were seen in the spectrum, that nearer to the more refractable end of the spectrum was extremely weak that it seemed so overradiated when the scale was illuminated that one could not measure it." This was then a very extensive polar aurora observed equally in Turkey, Greece, Italy, Portugal and Spain. In Germany observed F Zöllner (1881) the spectrum of the same polar aurora on the 25 October 1870. He also observed besides the red line the green line detected already earlier by A J Angström in the years 1867–1868 (Vogel 1872).

On the 5 November 1871 Konkoly observed polar aurora again with a spectroscope. He described this observation, as follows: "The spectroscope showed in all auroras which I could observe only the two known lines, from which that nearer to violet was extremely weak" (Konkoly 1879, Beobachtungen, p. 34).

According to notes by the event 11 February 1872, the observation of zodiac light, of the nightfall and the polar aurorae was performed by a spectroscope "Browning" with five prismas allowing simultaneous observation of the spectrum (Konkoly 1879, Beobachtungen, p. 67).

First photographs of the spectrum of polar aurorae were taken by E S King, not until April 1879 and 15 March 1898 at the Harvard College Observatory with exposures longer than 140 minutes. The most bright lines were in lengths 3892–3925 A and 4285 A (Piccering 1898).

The following polar aurorae have been optically observed in Ógyalla:

25	X	1870	31	X	1903
2, 3, 5	XI	1871	22	III	1920
11	II	1872	24	III	1940
9	IX	1898			

Descriptions of simultaneous disturbances of magnetic records have been presented usually with some polar aurorae observed in Ógyalla or other places (Réthly and Berkes 1963).

It is the merit of the founder of the Astrophysical Observatory Miklós Konkoly-Thege that auroral observations in Ógyalla performed before the 1st World War were published in scientific literature and later included in the catalogues of polar aurorae.

Konkoly devoted his work to a vast range of problems including besides astronomy also geophysics and meteorology. In the end, polar aurora is a phenomenon uniting in fact those three branches which, unfortunately, in the era of Czechoslovakia, were treated rather separately. However, during the last decades, there has

been the tendency to overcome university limits of classic branches formed by individual departments mostly thanks to the orientation of the observatory of the Geophysical Institute of the Slovak Academy of Sciences at Hurbanovo (former Ógyalla) after the year 1965. Even the mere historical data of the occurrence of polar aurora with exact timing is very important for statistical examination of the relations between the Sun and the Earth as a part of cosmic physics.

References

Eötvös L 1871: Vom Spektrum des Nordlichtes. *Természettudományi Közlöny*, III.

Fritz H 1873: Verzeichniss beobachteter Polarlichter. C Gerold's Sohn, Wien

Hell M 1777: Theoria nova lucis borealis. Ephemerides, Appendix, Vindobona

Konkoly M 1870: Bericht über ein prachtvolles Nordlicht, gesehen am 25. Oct. zu Ógyalla. *Hessische Wochenschrift*, Halle

Konkoly M 1874: Description of the Ógyalla astronomical observatory. Értekezések a Mathematikai Tudományok köréből, III, Budapest

Konkoly M 1879: Beobachtungen angestellt am Astrophysikalischen Observatorium in Ógyalla in Ungarn, I, Halle

Konkoly Thege M jr 1910: A natural phenomenon of unique beauty. *Az Időjárás*, XIV.

Křivský L 1993–1996: *Travaux Géophysiques*, Prague, 37, 1–21.

Křivský L, Pejml K 1985: *Travaux Géophysiques*, Prague, 33, 77–151.

Mairan J J D 1733: Traité physique et historique de l'Aurorae boréale. Paris

Makó P 1781: Dissertationes Physicae Quas Elucubratus est. Typis Regiae Universitatis, Budae

Piccering E C 1898: *Astr. Nachrichten*, 146, 3491, 175.

Réthly A, Berkes Z 1963: Nordlichtbeobachtungen in Ungarn (1523–1960). Akadémiai Kiadó, Budapest

Schenzl G 1984: *Zeitschr. für Meteorologie*, Vienna, 2.

Schröder W 1984: Das Phänomen des Polarlichts. Wissensch. Buchgesellsch., Darmstadt

Vogel H K 1872: *Pogg. Ann.*, 146, 569.

Zöllner F 1881: Wissenschaftliche Abhandlungen mit zwei populären Vorträgen. B. IV, Leipzig

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 125–133 (1997)

THE DEVELOPMENT OF THE SOLAR SYSTEM — RELATIONS TO PHYSICS AND COSMOLOGY

W SCHRÖDER[1] and H-J TREDER[2]

[Manuscript received June 5, 1997]

The Kantian cosmology "based on Newtonian principles" is founded on celestial mechanics, statistical mechanics and atomistic thermodynamics. However, these fundamental physical laws are working in the given cosmos.

Keywords: cosmology; Kant I; Newton I; solar history; stellar evolution

I.

Isaac Newton declared all cosmological problems to be illusory ones on the basis of his "Principia Mathematica Philosophiae Naturalis". Already in his "Scholia" to the big work he sharply rejected the Cartesian postulate that theoretical physics (in this case celestial mechanics) has to explain not only the laws governing the motions of celestial bodies, but also their existence and formation and indicated their insufficiency. In the correspondence between Clarke and Leibniz, Clarke finally formulated that it depends on the LORD's pleasure how HE created the world. Newton explained that on the basis of the principles of mechanics no cosmological initial conditions can exist. Whether the LORD created the world before one second, before six thousand years (as Newton, a Bible believing Christ had thought) or before many billion years, cannot fundamentally be decided, moreover it is physically absolutely irrelevant, too.

It was Immanuel Kant who threw discredit upon Newton's thesis when Kant developed as a devoted Newtonian (but who misunderstood here Newton's position) a "cosmogony based on Newtonian principles". The difference between Newton's and Kant's positions was for Kant at that time unclear. Newton took the reversibility of all natural processes as starting point, while Kant — presupposing the Newtonian principles which he accepted paradoxically as correct — wanted to deduce a cosmological development.

The difference of Newton's and Kant's conceptions is after all the topic of all discussions about the connection between physics, cosmology and cosmogony. A solution of this apory came later with the "Mechanical Theory of Heat".

Thermodynamics implies, in contrast to celestial mechanics, a direction of time. That is why thermodynamics leads cosmogonically, too, to an irreversible time in contrast to Newton. Wiener called this Clausius-Thomsonian time.

The thermodynamical definition of time leads in the cosmogony of the solar system to the Helmholtz-Kelvin time scale (Thomson, Lord Kelvin). A certain "age

[1]Hechelstrasse 8, D–28777 Bremen-Roennebeck, Germany
[2]Rosa-Luxemburg-Strasse 17a, D–14482 Potsdam, Germany

1217-8977/97/$ 5.00

of the Sun" and a cosmogonical time-span of twenty to thirty million years results from the classical principles of mechanics and thermodynamics together. Today the Helmholtz-Kelvin time scale is considered as a time interval in which cosmogony can be discussed without taking into account quantum mechanics, theory of relativity and nuclear physics.

II.

The connections between cosmogony of the solar system and cosmology as science about the global structure of the Universe have an aspect being for a first glance trivial (which is nevertheless perhaps not as evident), namely the fact that the solar system came into being and developed further in a given, but in time changing cosmic environment. Naturally, this cosmic environment is firstly the near vicinity of the Sun in our arm of the Milky Way, then the Galaxy in general. Our Galaxy is then again embedded (across intermediate systems) into a Meta-Galaxy and finally all physics takes place in the cosmos as it is given for us. This is, however, an evolving cosmos depending on time.

The cosmological foundation of the cosmology of the solar system is evidently of high importance and it is the cosmic indication of the initial conditions. — The Universe, more correctly the Meta-Galaxy is about ten to twenty billion years old. In order to enable the discussion with a concrete number, we put here the age of the Universe t equal to the reciprocal of Hubble's constant $t = H^{-1}$ and suppose that this is about fifteen billion years.

The age of the solar system is praeterpropter the same as the age of the Sun as a fixed star and is about four and half billion years: for simplicity it will be rounded off to five billion years. Correspondingly, the age of the solar system is one third of the age of the Universe.

The solar system is therefore on the one hand young with respect to the age of the Universe, on the other hand the age of the solar system till present is of comparable order of magnitude with that of the Universe. Significant changes in the cosmological conditions can be therefore reflected in the history of the solar system. It is, however, clear that the solar system formed in an already "old cosmos". The Sun is *de facto* a star of Disk Population I and the primary matter which can be studied in the frame of nuclear physics and from which the solar system developed, is by no means the original matter which came into being "in the first minutes of the Universe". The primary matter of the Sun (and also the primary matter of the planets) had been "cooked through" in stars of previous generations, thus its isotope composition differs significantly from that of the cosmological primary matter.

Let us imagine for sake of simplicity that stars of Disk Population I came into being from matter which have a composition being the result of development processes of stars with heavy masses from a primary Population II and that the stars of this Population had developed "shortly after the big bang" from the cosmological primary matter. The primary matter of the solar system was then due to a supernova-like end of the perhaps ten billion years long development of one or of

several primary stars. This catastrophic end of the primary stars led according to the generally accepted idea to the formation of cosmic nebulae with a new type of isotopic composition and deducible thermodynamical parameters.

The mass of the nebulae could be from some hundreds up to a few thousand solar masses and they included more or less imaginable granulations, opacities etc.

All these determine then the minimum mass M into which the nebula decomposed itself due to its gravitational instability. These minimum masses are given by Jeans' equation:

$$M = \left(\frac{40}{9}\right) \cdot \frac{3}{2} \cdot \frac{1}{\sqrt{\varrho}} \cdot \frac{(RT)^{3/2}}{f\mu} . \tag{1}$$

Here ϱ is the mass density, T the temperature and μ the average atomic weight. We remark here that two universal constants appear in Jeans' equation: the gas constant R and the gravitational constant f. Further it is to be remarked that the temperature T depends on the external radiation field.

The following history of Jeans' mass M depends strongly on its torque. It is possible to give for this an (also cosmologically interesting) estimation. Would the torque of the "primeval Sun" be great enough that the Kepler-orbits of its outermost layers would correspond to a rigid rotation of the solar body, i.e. it would be valid with the original solar radius a and with the angular velocity:

$$\textit{angular momentum} \approx \frac{2}{5} M \cdot a^2 \cdot \omega \tag{2}$$

then the contraction of the primeval Sun would result in a split up to (at least) nearly equal masses according to the theory of the equilibrium figures of rotating bodies. The mass ratios are 2:3 or 2:5 etc. That is the development known in celestial mechanics since Poincaré, Darwin and Jeans.

If the torque of the primeval Sun is very small with respect to the value in Eq. (2) then a splitting into such multiple stars is impossible. The originally existing small torque is split according to a process discussed already by Kant.

If we consider the solar system then the torque is about one thousandth part of the torque in Eq. (2) that would be necessary for a quasi-rigid rotation of the primeval Sun.

It is evident that nearly the total mass of the solar system is represented by the present solar mass, thus an observer from a greater distance (say from a distance of ten light years) could not detect the existence of the planetary companions of the Sun using methods of celestial mechanics and normal astrophysical methods. From this point of view the cosmogony of the solar system would be practically identical with the development of the Sun itself and it is our specific point of view that as observers of a planet, we are interested in peculiar effects like the existence and the development of planets.

A second fact makes this simplification dubious. The torque (be it as small as it is) is in 98 percent bound to the orbital torque of the planets (including 90 percent bound to the orbital torque of Jupiter and Saturn). That means that: for the history of the development of the Sun is the existence of the planetary system

— as bearer of the torque of the primeval Sun — highly relevant. We have here a completely uneven distribution: the main mass has practically zero torque, and the bearers of the torque have actually no mass. This fact has been since Kant and Laplace one of the decisive problems of the cosmogony and led to present research about the transport of torque not in accordance to the laws of Newtonian mechanics, but incidentally by the action of magnetic fields (Alfvén).

A Kantian cosmogony has in contrast as starting point that the torque is anyway very small and therefore the distribution of the torque can be explained as a stochastic process. This is the common point of the else in details very different cosmogonies by von Weizsäcker and Schmidt.

The development of a planetary system is in a Kantian cosmogony a necessary physical consequence. In all development histories of stars (naturally only of stars of Population I) is the original torque small in the previous sense. In all cases a planetary system similarly to our solar system has to come into being as Kant had shown it. The planetary systems may differ only in peculiarities. And these peculiarities would be then "frozen" characteristics of the primary state.

The position of a cosmogony (based on Newtonian principles) is according to Kant's idea that the development of the considered system has been practically closed since a sufficiently long past time therefore the laws of mechanics and thermodynamics are satisfactory within the system itself (with addition of nuclear physics and magnetohydrodynamics) for an explanation of the cosmogony. There is, however, a second thesis that the conservation laws of energy, torque, gravity, mass, charge etc. yield "inhibitions" for this development process: else the final state is independent of the initial state, therefore the original situation can only be reproduced after a sufficiently long time interval using "frozen-in fossil states" Kantian cosmogonies declare that the present state of the solar system inevitably developed, independently of the "fine structure" of the primeval nebula.

Here we meet a substantial criticism, the problem of relaxation times. Ideas implied in the Kantian hypothesis about the pre- and early history of the planetary system tell us that — omitting secular effects (like the Darwinian tidal friction) — the basic structures of the solar system have existed for more than four billion years as all decisive cosmological processes took place in the first hundred million years.

This is valid with high probability for the inner planets. The age of the Earth should be not very different from the age of the Sun. The outer planets which are both physically and chemically very different from the inner planets, however, are the main bearer of the torque of the solar system and their formation is correspondingly a cosmogonically important process. It has been several times remarked that already the planet Jupiter's existence influenced significantly the formation of the other planets due to its great mass, but the problem of the time interval of their formation is for the great planets much more difficult than for the small ones.

III.

All cosmological theories which declare the final state of a cosmic system as necessary and independent of all initial data, have as basis explicitly or implicitly thermodynamics and statistical mechanics in addition to celestial mechanics. The decisive cosmogonical process is the transformation of mechanical (kinetic or potential) energy into heat energy and this is consequently irreversible. That does not mean, however, that the cosmogonical system is closed, but it is exothermal. During the formation of the Sun and of its planets (and later on during the further development of the Sun) energy is transported from the system in form of radiation energy; with the main component thermal radiation.

According to Kant's and Laplace's classical cosmogonical hypotheses this heat transport is the unique connection of the solar system with the cosmic space, and this essential contact is completely described by the laws of atomistics and thermodynamics.

The contraction of the primeval Sun means from a radius $r = ra$ (i.e. from a few times the distance of Neptune) to the present solar radius $r = r_0$ (at which nuclear fusion started) according to Helmholtz' and Thomson's proof that potential energy is transformed into thermal energy. For systems with very small torque the history of the formation of the solar systems can be quantitatively characterised according to the principles of thermodynamics and statistical mechanics beginning with the separation of the primeval Sun from the cosmic nebula till the development of the present Sun using Helmholtz's and Poincaré's theorem. Poincaré's theorem connects the potential energy V of the "sun" with its internal energy U corresponding to the virial law:

$$U = \frac{1}{2} \cdot |V| = \frac{3}{10} \cdot f \cdot \frac{M^2}{r} \tag{3}$$

and this yields Helmholtz's formula for the energy difference:

$$V(r_a) - V(r_e) + U(r_a) - U(r_e) = \frac{1}{2} \cdot (V(r_a) - V(r_e)) \tag{4}$$

(r_a is the radius in the initial situation, r_e that in the final situation) between the different stages of the contraction.

The energy difference in Eq. (4), i.e. the half of the loss in potential energy is radiated as heat into the cosmos. Helmholtz has already shown that the time of the contraction of the primeval Sun from the Neptunian distance $r = r_a$ till the present radius $r = r_e$ is the characteristic time for the cosmogony of the Sun. This time is then the Helmholtz-Thomson time of about twenty to thirty million years.

During this Helmholtz-Kelvin-epoch the system losses no mass (omitting here absolutely irrelevant effects due to the special relativistic mass equivalence) and conserves its torque. This torque is taken up in Kantian cosmologies by the planetary system which possesses at present 98 percent of the torque of the solar system.

Kant's and Helmholtz's argumentation led thus to the thesis that the degrees of freedom of the torque, the orbital movements of the planets were determined during the Helmholtz-Kelvin-stage corresponding to the contraction of the primeval

Sun: practically mass-free particles take up the torque what remained during the contraction of the primeval Sun. Kant's genuine conception, namely that inelastic collisions between "planetesimals" which moved chaotically as macroscopic corpuscles and which made up the nuclei of planets on "arbitrary orbits" according to Newtonian principles fits into the Helmholtz-Kelvin-scenario and it is the central point at Schmidt. The relative velocities of the colliding corpuscles are lost at inelastic collisions and kinetic energy is transformed into thermal energy according to Carnot's formula. It is possible in this case to translate the argumentation of the statistical thermodynamics mutatis mutandis on cosmogony analogously to Boltzmann's law of equal distribution.

Kant's macroscopic "planetesimals" are, however, the main constituents between the macroscopic corpuscles of the planets' matter only within the special radiation regime in the vicinity of the Sun. At greater distances from the Sun, the "escaping" gases of the primeval nebula are the overwhelming constituents.

Moreover the process of the formation of the great outer planets is continuously a process in which macroscopic chaos is transformed into an "elementarily unordered state".

According to this hypothesis the formation of the planetary system is a transition from turbulent (macroscopically chaotic) movements to macroscopically ordered (laminar) movement due to the internal friction (kinematic viscosity) in the primeval nebula considering which the time of the formation of the planets can be much longer than one Helmholtz-Kelvin period. Additionally the state "elementary disorder" is atomistically much more probable and correspondingly the macroscopic order, too.

Von Weizsäcker brought back the Cartesian ideas into a Kantian cosmogony, when he made a supposition concerning the formation of the outer planets (this is cosmogonically much more important than that of the inner ones). At first the torque was taken up by "planetary eddies" and the next problem is the relaxation time (i.e. Reynold's number) during which the turbulent motion becomes a macroscopically ordered motion.

According to Kant's and Laplace's idea the cosmogony is running "spontaneously". The formation of the solar system is the consequence of the laws of nature on an undisturbed system. The "primeval nebula" losses energy by thermal radiation, but it does not loss mass or torque. The principles of physics — as applied to such a primeval nebula — lead finally to similar planetary systems. The problem remains, however, if this process lasts as long as the Helmholtz-Kelvin duration is, or it lasts much longer.

Alfvén developed the idea on the basis of his cosmic magnetohydrodynamics that the bulk of the initial torque was removed from the solar system by the transport of very small masses. Alfvén's hypothesis relies naturally to the relatively small total torque of the solar system, too.

The change of the torque balance has been the starting point for all "catastrophe-theories" of the cosmogony since Buffon. Basically arbitrary initial conditions are at disposal for such "open systems" (Nölke) which enable solutions for all problems; one puts simply the requested final result into the initial conditions. But in such a case the structure of the solar system is a random phenomenon — according to

the Lord's will — according to Newton and not a regular consequence according to Kant.

The notion of a "closed" cosmogonical system should naturally be accepted cum grano solis. In addition to the energetically significant thermal radiation, there are other interactions between solar system and cosmos, too. — The Sun is the source of corpuscular radiation, and the radiation pressure can expel small particles from the solar system (Schwarzschild and Arrhenius).

Further on corpuscles may leave the solar system by diffusion, too. On the other side, the solar system receives different kinds of electromagnetic radiation. It is further not impossible that masses reach the solar system sometimes from the interstellar space in form of meteorites or comets. The thesis about the "closedness" tells us only that these processes are cosmogonically irrelevant.

IV.

The problem of the "closedness" of the solar system refers to a stationary cosmos. A time dependence of the physical constants as implied by a non-stationary cosmos leads to a change in the definition of the closedness.

In spite of the fact that the age of the solar system is only one third of the age of the cosmos H^{-1}, the two times are nevertheless of comparable value. There are hypotheses in the framework of the further development of the general relativistic cosmogony which couple the value of the fundamental constants of nature with the age of the cosmos. (It is here allowed to use a linear approximation due to the short age of the solar system in comparison to the age of the Universe.)

Dirac suggested in 1937 that the gravity constant f changes with the age of the Universe and supposed for this change the formula:

$$\frac{\dot{f}\sim}{f\sim} = \alpha H\,, \tag{5}$$

where α is a constant of the order of magnitude "one" and according to Dirac it should be of negative sign. But the gas constant R should be also time dependent together with the Boltzmann constant k, which expresses the connection of the average kinetic energy of the molecules $\sim m\dot{x}^2$ with the absolute temperature T. As m is the inertial mass, the Huygens-Mach principle declares the relativity of the inertia and its general relativistic formulation in the Mach-Einstein doctrine declares the induction of the inertia to bey due to the cosmic gravitation.

Nevertheless, Dirac's hypothesis allows us two interpretations. Dirac himself prefers the second one, according to which the masses m of all particles change simultaneously with f. Then it is valid in all cases:

$$\frac{\dot{m}}{m} = \beta H\,, \tag{6}$$

where β is again a constant of the order of magnitude one and its sign depends on the concrete cosmological hypothesis.

It is to be remarked here that the constants f and R appear with very different exponents in the formulas of cosmogony which can be based either on celestial mechanics or on thermodynamics. Cosmologically grounded secular changes of these constants may be partly strengthening each other, in other effects they may have opposite consequences. As the exponents in the dependencies on f and R may be, however, partly rather big, the variations of these constants either according to Dirac's hypothesis or following the Mach-Einstein doctrine may have significant cosmogonical consequences (Treder 1990).

According to Dirac's original idea the solar temperature which is proportional to F^7, was earlier much higher than is now. Simultaneously significant geotectonic effects would appear due to the increase of the Earth's radius (Jordan, Dicke). Moreover, the distance of the planets from the Sun would also change according to Kepler's III. Law

$$\frac{\dot{r}}{r} = \frac{-2\dot{f}}{f} \tag{7}$$

thus according to these cosmological hypotheses, the present distances of the planets are random values of a time-dependent function.

Dirac's later interpretation of his hypothesis and a fortiori the Mach-Einstein doctrine also imply the possibility of a secular variation of the angular momentum J of the planetary system (Treder 1972) according to

$$\frac{\dot{J}}{J} = H\,. \tag{8}$$

This change corresponds to a tide-like coupling between the solar system and the Universe. In the case of such a coupling between cosmos and solar system the cosmogonically so important law of the conservation of the torque would be hurt; we had to start from the idea that the torque of the solar system could significantly have differed from its present value at the time of the formation of the planetary system.

It may be of interest to list again the physical and cosmological time-scales which are of significance for the cosmogony of the solar system: classical thermodynamics and gas theory lead to the Kelvin-Helmholtz period of a few times 10^7 years. The nuclear physically understandable age of the Sun is a few times 10^9 years (Bethe-Weizsäcker era). The cosmological age of the Universe is $\sim H \sim 10^{10}$ years. Finally the duration during which the Sun's mass can be supposed to remain unchanged is about $\sim 10^{12}$ years (Eddington era), but the Sun would explode much earlier as a nova.

Acknowledgement

The authors thank the late Professor Hannes Alfvén for helpful and fruitful discussions.

References

Alfvén H 1943: On the cosmogony of the solar system. Stock. Obs. Ann., 14, No. 5.
Alfvén H 1966: Worlds — Antiworlds. Freeman, San Francisco
Brush S G 1992: *Plasma Sci.*, 20, 577–589.
Kant I 1954: Kant's gesammelte Schriften. Vol. I., Edition 1910, Reimer, Berlin, 183–191.
Kant I 1954: Ob die Erde veralte? Kant's gesammelte Schriften. Vol. I., Edition 1910, Reimer, Berlin, 193–213.
Laplace P S de 1796: Exposition du systeme du monde. Du Cercle Social, Paris
Treder H-J 1972: Die Relativität der Trägheit. Akademie-Verlag, Berlin
Treder H-J 1990: Kosmologie und Kosmogonie. Akademie-Verlag, Berlin

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 135–150 (1997)

THE MID-GERMAN CONDUCTIVITY ANOMALY AND ITS CONTRIBUTION TO THE DETERMINATION OF THE VARISCAN DEEP STRUCTURE

G PORSTENDORFER[1]

[Manuscript received April 7, 1997]

It is possible to construct a precise isoline map of the scalar magnitude of the induction arrows for Germany in the period range 10 to 20 min based on the results of different study groups. In the central part of Germany, a minimum zone indicates high crustal conductivity to the North of the renewed southward increase of these scalar magnitudes and it corresponds to the location of the Mid-German Crystalline Zone in the border area between Saxothuringicum and Rhenoherzynicum. Using magnetotelluric, refraction seismic, reflection seismic, geothermal, magnetic and gravimetric results and the results of geological and geochemical studies, the author tries to chose geotectonic models of the Variscan which can be harmonised with the present knowledge about the causes of high electric conductivity in the central part of the Earth's crust.

Keywords: electromagnetic induction; induction arrows; magnetotellurics; Mid-German conductivity anomaly; Variscan structure

Geomagnetic deep sounding is the most advantageous method to trace areal distribution of areal conductivity anomalies. Maps of the "Wiese-arrows" and "complex induction arrows", respectively were compiled several times by Berktold for the western, and by Ritter for the eastern parts of Germany, then they were summarised in a common map for the whole of Germany (Berktold and Ritter 1992).

The groups in München (Hofer 1990), Göttingen (Peter 1994), Frankfurt (ERCEUGT-Group 1990), Münster (Jödicke 1990), Berlin (Tauber 1993), Potsdam (Eisel 1995), Niemegk (Ritter 1991), Leipzig (Schwarz, current activities) and Freiberg (Börner 1995) added in the nineties new data to the existing ones. Therefore it became desirable to compile a new map of the scalar magnitudes of the arrows for Germany, taking into account the results of the Prague group, too, in the border area to the Czech Republic (Červ et al. 1994, Klause et al. 1994, Pečova and Praus 1996) (Fig. 1).

This map is valid in its present form for the period range 10 to 20 min in which a maximum of the real part yields the transfer function Z/H with small or evanescent imaginary parts. It is based on more than 250 measuring sites with a regrettable inhomogeneous distribution, thus this map is partly of a schematic character. Areas with a significant change in the direction of the arrows implying a zero crossing along a detectable strike direction are indicated by dotted isolines. Such areas occur in the area of the "North German-Polish Conductivity Anomaly", indicated by vertical shading, at the "Göttingen D-Anomaly", indicated by oblique shading and in the

[1]D–09599 Freiberg, Mendelejew Str. 38, Germany

1217-8977/97/$ 5.00

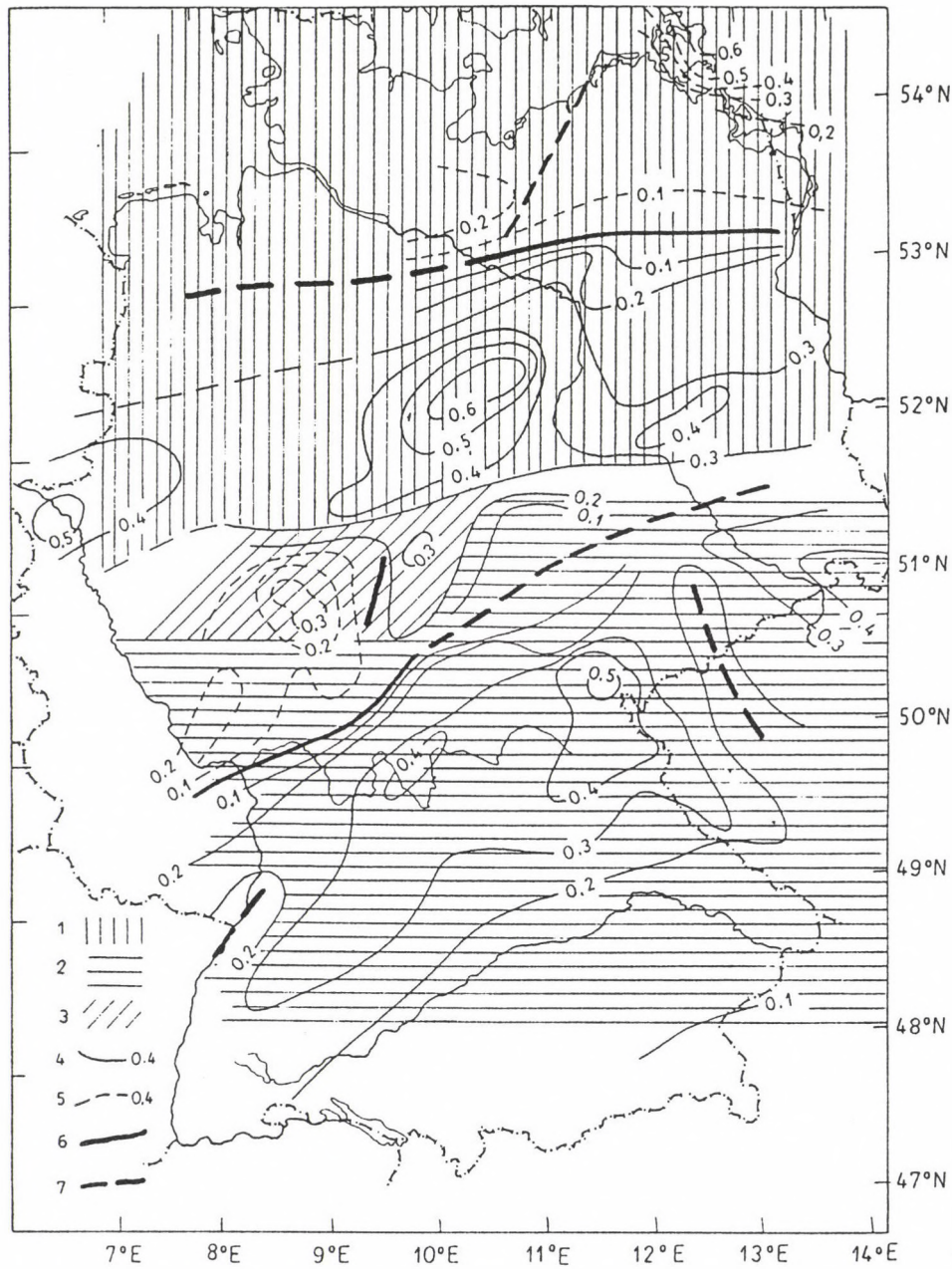


Fig. 1. Map of the absolute values of the transfer function Z/H for the period range 10 to 20 min (Porstendorfer 1995). 1 – Area of the "Northern German – Polish Conductivity Anomaly", 2 – Area of the "Mid-German Conductivity Anomaly", 3 – Area of the "Göttingen D-Anomaly", 4 – Z/H isoline (positive), 5 – Z/H isoline (negative), 6 – Z/H minimum zone (sure), 7 – Z/H minimum zone (unsure)

area of the "Mid-German Conductivity Anomaly", indicated by horizontal shading. This paper deals mainly with the latter anomaly. Within this anomaly, there are two further minimum zones in the area of the Upper Rhine Valley and in the border region between Eastern and Western Erzgebirge.

According to the location and strike of this "Mid-German Conductivity Anomaly" is a connection of this anomaly with the zones of Moldanubicum, Saxothuringicum and Rhenoherzynicum of the Variscan Orogenesis very likely. These zones were detected by Kossmat based on the investigation of sedimentary, volcanological, tectonic and metamorphic rocks (Fig. 2). This idea has been expressed several times by different authors within the ERCEUGT-Group in Western Germany and within the ZENTROSEIS-Group in Eastern Germany. A comprehensive concentration gave Berktold (1995).

The northern part of this anomaly in Eastern Germany is determined by the last traces of the Northern German-Polish Conductivity Anomaly, the renewed increase of the absolute value of Z/H is to be emphasised here toward south followed by a slow decrease toward the Alps. According to experiences with models of Geomagnetic Deep Sounding, the centre of the corresponding zone of high conductivity is to be looked for in the area of the plotted minimum zone as far as the maximum of the gradients of the renewed increase of the Z/H values. This area lies in Saxothuringicum approximately in the Mid-German Crystalline Zone.

A great number of magnetotelluric soundings, then corresponding modellings and inversions made by the mentioned working groups led to a more precise determination of the distribution of the resistivity. Such results were reported by the members of these groups. It can be stated in general that especially the phase curves of the E-polarization with respect of the regional Variscan strike direction indicate with a maximum up to 80° in a period range of 30 to 100 s a conducting crustal zone at relatively short periods (the corresponding minimum of ϱ_s follows partly only in the period range > 1000 s), and a depth range of 10 to 20 km is to be attributed to these values where a continuous increase of the conductivity from south toward north and a pulling-up of the conductive formations in the area of the Mid-German Crystalline Zone were established. The lowest values of the specific resistivity lie at 0.1 to 0.3 Ωm and the conductance of the conducting crustal layer amounts to about 10 000 Siemens. This conductivity is thus higher than that given for the North German-Polish Conductivity Anomaly for the Pre-Zechstein formations. Figure 2 shows the position of two geotraverses from the Alps to the Baltic Sea. It contains information about the characteristic distribution of the transfer function Z/H together with hints from magnetotelluric measurements to conductive zones as presented in Fig. 3.

Figure 4 is a further step forward as it includes results from refraction seismics and means a petrographic-geotectonic *a priori* interpretation.

Both Franke et al. (1990) and the ZENTROSEIS-Group (Bormann et al. 1989) prefer here an overthrusting tectonics directed from south to north. The question of the connection of the conductivity anomaly with other geophysical data leads first of all to a coincidence of a clear negative velocity anomaly in a depth of 10 km, whereas in the east profile relative low velocities around 5.9 km/s are indicated

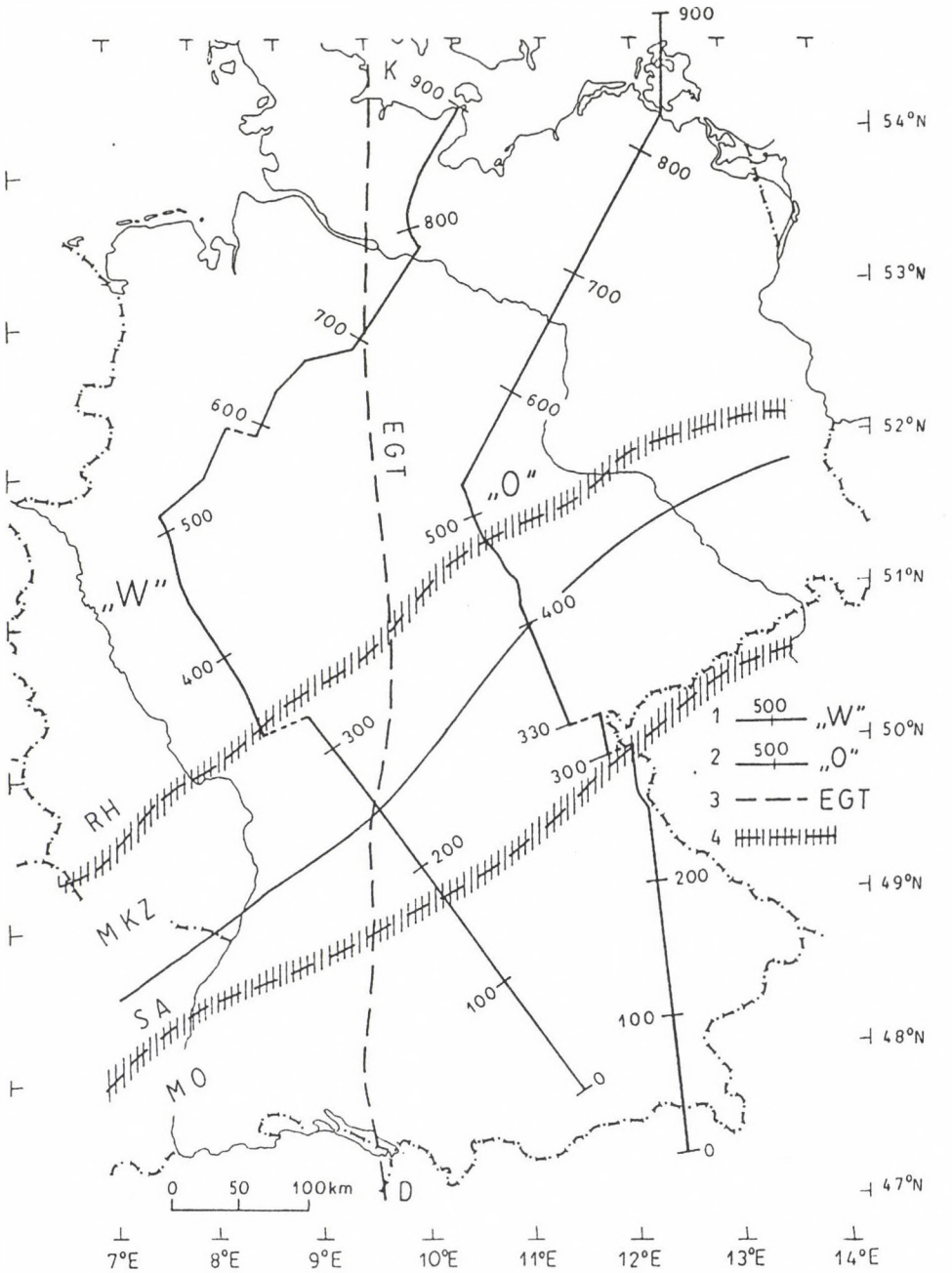


Fig. 2. Variscan zones after Kossmat. MO – Moldanubicum, SA – Saxothuringicum, RH – Rhenoherzynicum, MKZ – Mid-German Crystalline Zone, 1 – Western profile "W" of the ERCEUGT-Group (Bahr et al. 1990) with km-values, 2 – Eastern profile "O" by Porstendorfer (1995) with km-values, 3 – European Geotraverse, 4 – Border zone of the Variscan Kossmat-zones

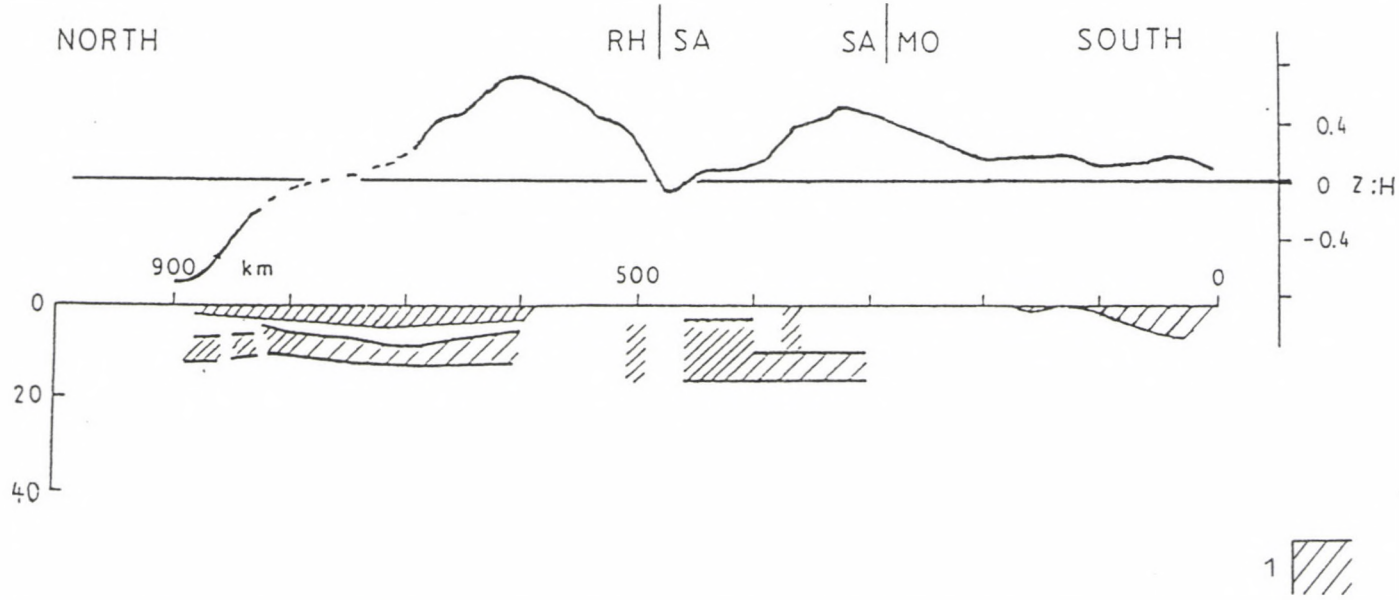


Fig. 3a. Absolute values of the transfer function Z/H and hints at conducting zones based on magnetotelluric results along geotraverses in Fig. 2. Eastern profile (after Porstendorfer 1995), 1 – conducting zones, MO – Moldanubicum, SA – Saxothuringicum, RH – Rhenoherzynicum

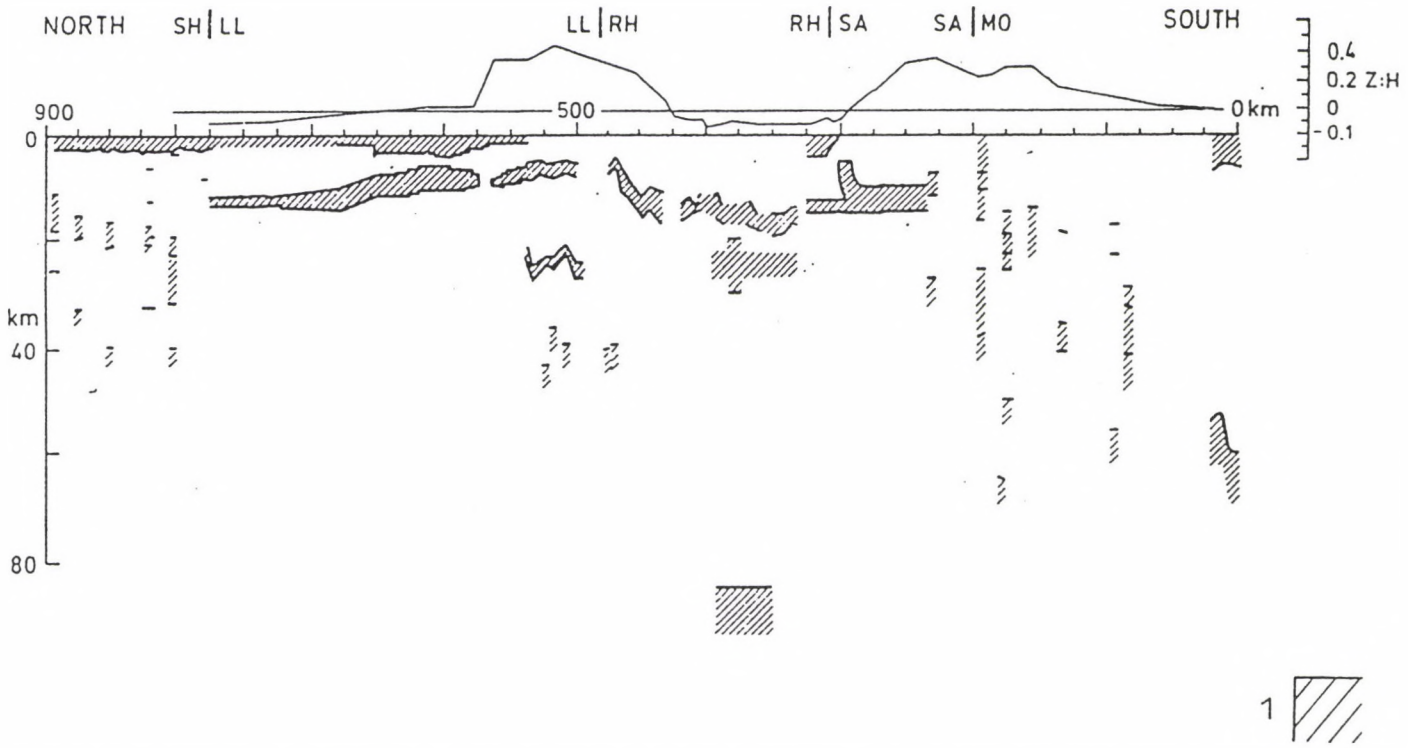


Fig. 3b. Absolute values of the transfer function Z/H and hints at conducting zones based on magnetotelluric results along geotraverses in Fig. 2. Western profile compiled by the ERCEUGT-group (Bahr et al. 1990), 1 – conducting zones, MO – Moldanubicum, SA – Saxothuringicum, RH – Rhenoherzynicum

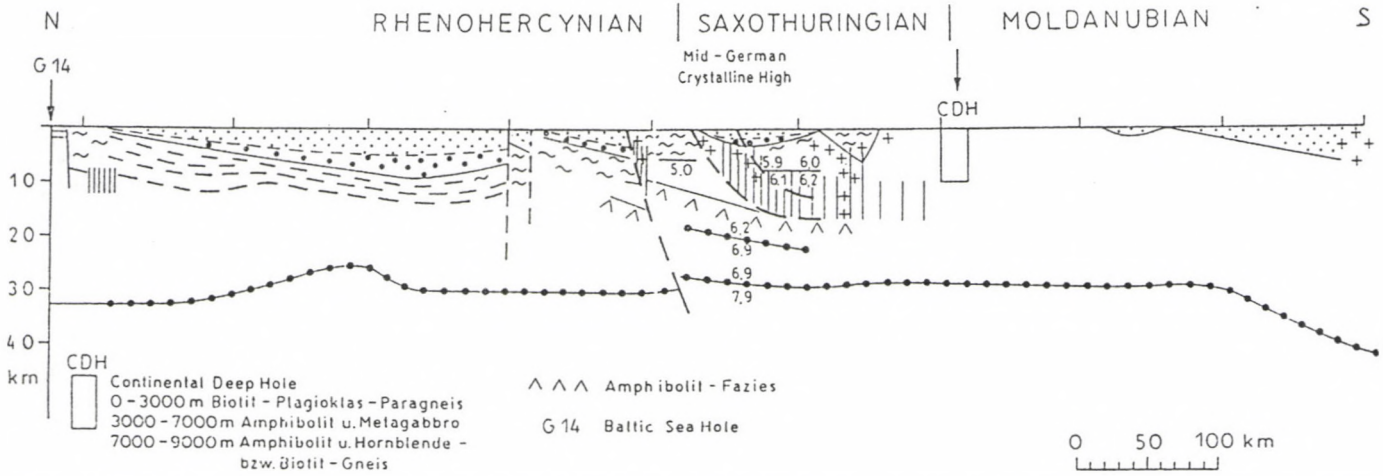


Fig. 4a. Combination of magnetotelluric results with results of reflection seismics and an *a priori* petrographical-geotectonic interpretation. East profile (Porstendorfer 1995). The eastern profile includes the results of the ZENTROSEIS-group along the THUMARK-SOUTH profile.

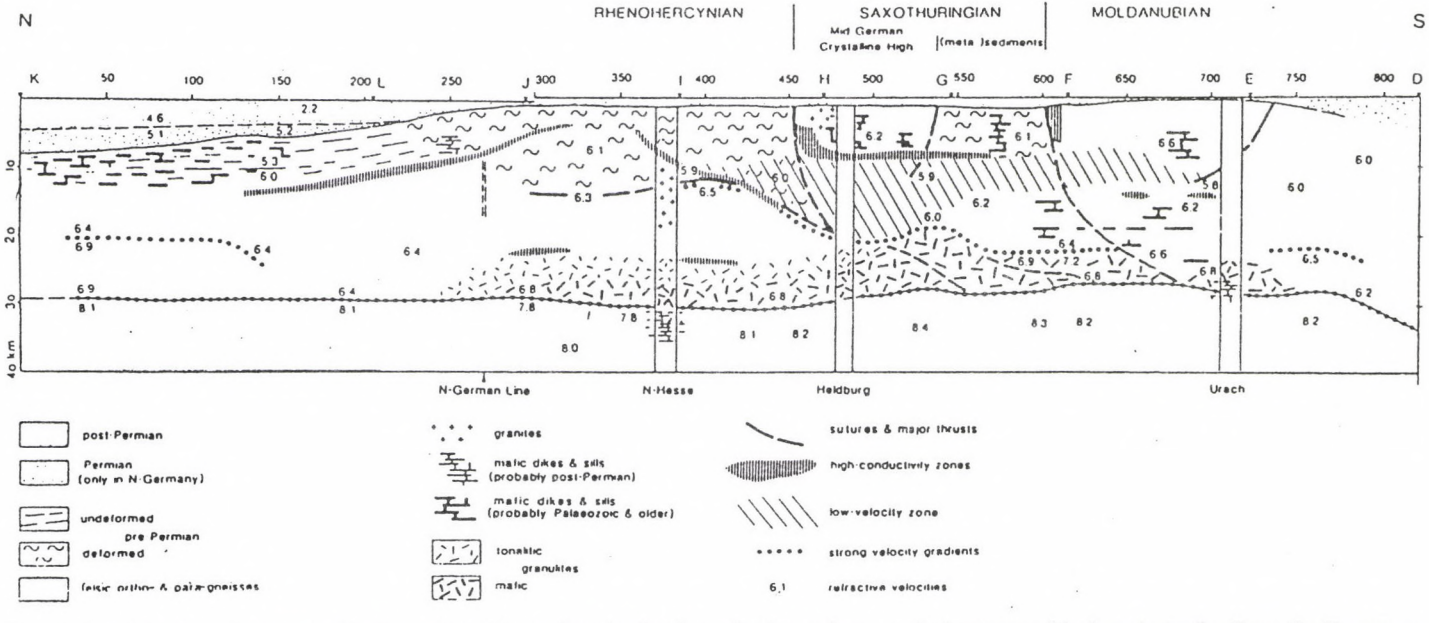


Fig. 4b. Combination of magnetotelluric results with results of reflection seismics and an *a priori* petrographical-geotectonic ollows the European Geotraverse. West profile = EGT (ERCEUGT-group, Bahr et al. 1990). The western profile (after Franke et al. 1990) follows the European Geotraverse.

in a depth of 10 km. Knothe (1972) found such low-velocity-zones especially along International Deep Seismic Profile VI, which crosses the east profile in the area of the Mid-German Crystalline Zone. Similarly the residuals of submerging waves obtained from quarry explosions with respect to normal traveltime curves hint at this zone as far as the area north of Vogtland (Porstendorfer 1980).

The similarity between the conductivity anomaly and anomalous zones obtained by geomagnetic and gravimetric surveys is illustrated in Fig. 5. A range of positive magnetic anomalies at about the southern border of the Mid-German Crystalline is accompanied by a range of low gravimetric values to the north of it. For comparison, the minimum zone of Z/H is also plotted in Fig. 5. The gravimetric minima are supposed to be due to acid intrusions in the area of the Mid-German Crystalline Zone during the maximum of the collision of the Variscan Orogen about 340 to 330 million years B.P. (Early Carboniferous), the geomagnetic maxima hint at basic bodies, supposedly, however, of an Early Palaeozoic age.

Deep Reflection Seismic crosses with the profiles DEKORP 2 S and MVE-9 (West) the anomalous zones roughly parallel to the strike. The interpretation by Heinrichs et al (1994) is based at both profiles on an element with high reflection intensity and occurrence frequency with a characteristic reflection dip. This element is called Tauber-Vesser zone and it corresponds again with positive zones of geomagnetic anomalies at the southern border of the Mid-German Crystalline Zone (Fig. 6).

The map of the Moho-depth indicates a slight upwelling in the area of the Mid-German Crystalline Zone up to values of about 28 km. This can be connected to a movement toward isostatic balance and hints at previously much thicker crust in this area.

Finally reference is to be made at the map of the thermal flux (Čermak 1990) which shows no anomalous values in the area of the Mid-German Crystalline Zone and which leads — accepting suppositions about the radiogenous heat production and about the heat conductivity — to a depth > 25 km of the 400°C isotherm. In order to reach some idea about the possible causes of the high electrical conductivity in the middle crust the processes are to be considered which lead to the development of high conductivity in the crust starting from sedimentary formations.

As initial facies is a sandy-clayey one supposed in which organic material is embedded partly in a disperse form, partly — as e.g. in Carboniferous age sediments — in form of carbon layers. Moreover stratiform disperse ore deposits are also allowed, e.g. as sea bottom ore mud in rift zones.

As long as the sedimentary material lies near to the surface, the conductivity is determined by the contiguous pore volume of mineral containing water, its salinity and by the surface conductivity of clay.

If the material sinks and diagenesis and regional metamorphosis take place under ever increasing pressure and temperature conditions, new minerals develop, the pores close and the remaining water is built into solid mineral bonds. This so-called "green schist facies" is formed at temperatures above 300°C. In the transition to the so-called "amphibolite facies" at temperatures above 400°C crystalline water becomes again free water and a highly mineralised fluid phase develops which covers

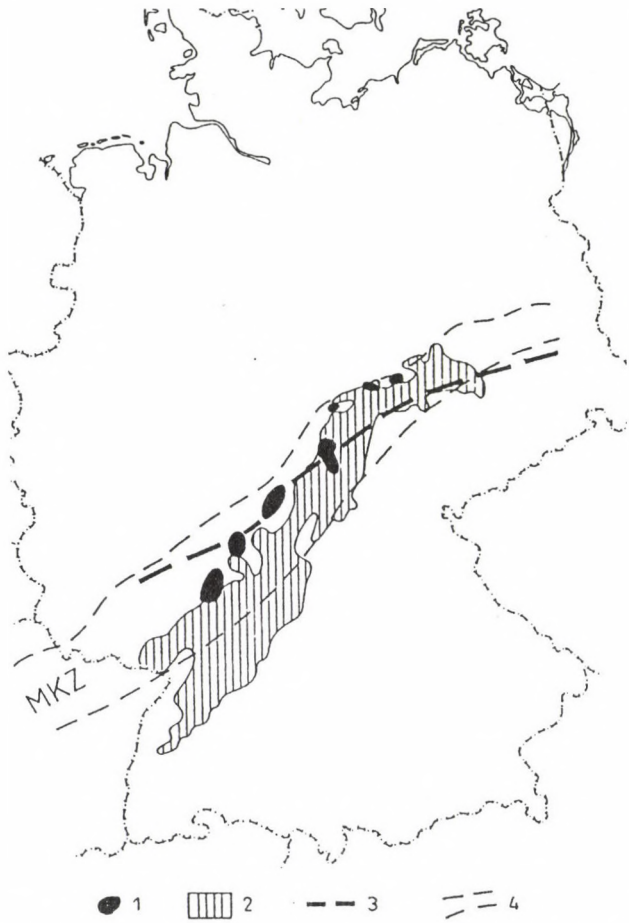


Fig. 5. Anomaly ranges of positive magnetic and negative gravimetric extremum values in the area of the Mid-German Crystalline Zone (Heinrichs et al. 1994, Bahr et al. 1994) compared to the axes Z/H of the Mid-German Conductivity Anomaly 1 – Area of negative gravimetric extrema, 2 – Area of positive magnetic extrema, 3 – Minimum values Z/H of the Mid-German Conductivity Anomaly, 4 – Area of the Mid-German Crystalline Zone, MKZ

the rock matrix in a skin-like form. This cover eventually enables then initial melting due to the high water content. The continuing metamorphosis leads then finally to a complete dehydration of the rocks in the so-called "granulite facies". Figure 7 shows the conductivity from laboratory experiments under high pressure (after Glover and Vine 1995) and at high temperature in conditions of the amphibolite facies.

The petrophysical conditions of the amphibolite facies have further importance with respect to seimics. The low rock permeability in greater depths hampers a quick migration of the fluid phase, therefore a pore pressure builds up which adds to the matrix pressure. In other words the pressure of closed systems is increasingly taken up by the water volume. Now as the seismic velocity is determined

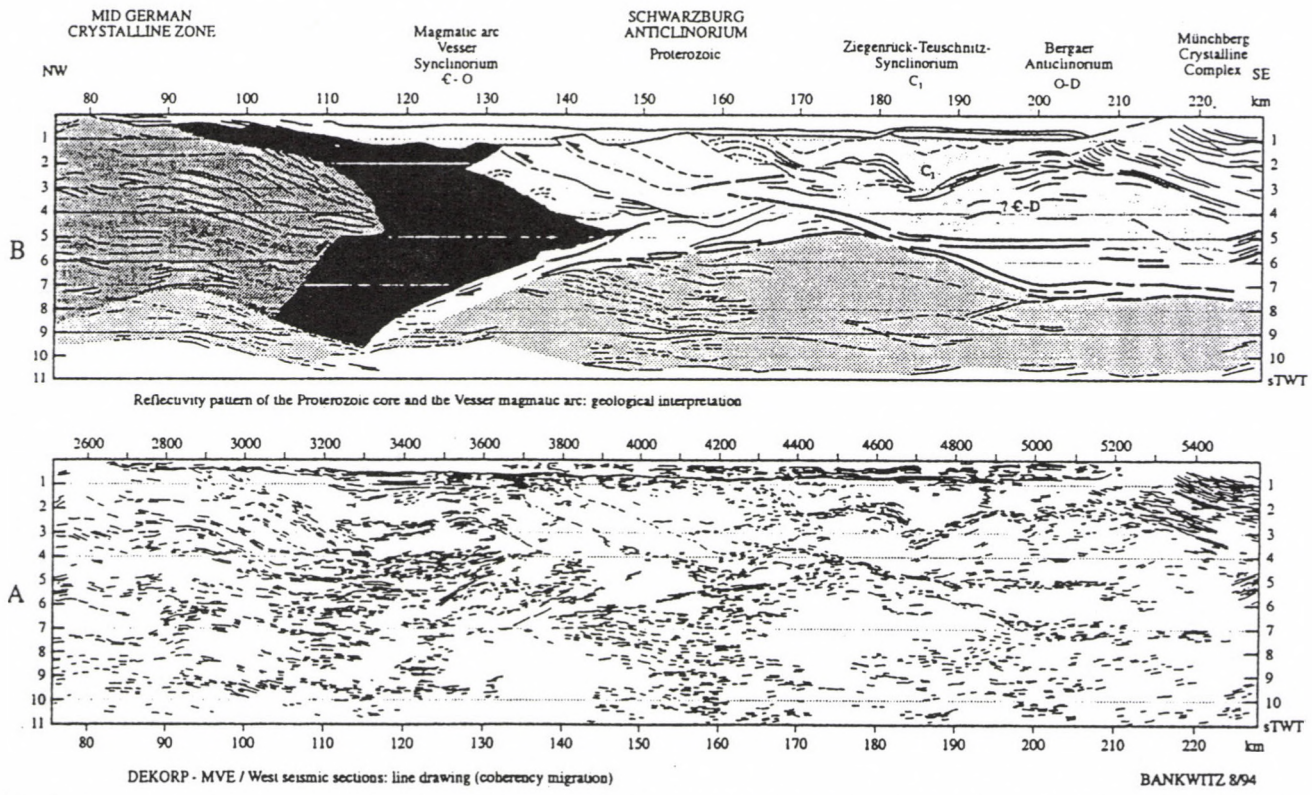


Fig. 6. Geotectonic interpretation of the DEKORP-profile MVE/West crossing the Mid-German Crystalline Zone after Bankwitz (Heinrichs et al. 1994)

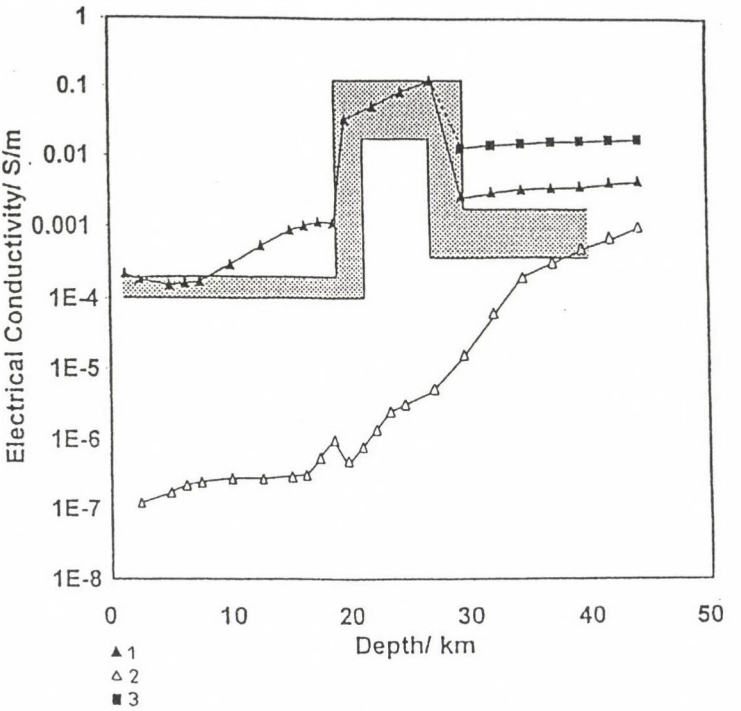


Fig. 7. Depth curves of the specific resistivity in continental areas (Glover and Vine 1995) for a geothermal gradient of 20.4 C°/km. 1 – 0.5 m NaCl saturated rock without graphite, 2 – dry rock, 3 – 0.5 m NaCl saturated rock with graphite

by the pressure on the rock matrix, a low velocity zone develops as frequently observed in connection with high crustal electrical conductivities (Schön, personal communication in Porstendorfer 1973). If the water migrates then the boundary zone between brittle and ductile deformation is a possible accumulation and migration zone with a relatively high secondary porosity at about 300°C. The remaining organic material is subjected diagenesis and metamorphism in ranges of higher pressure and temperature, with increasing carbonisation which can even reach the graphitic stage ($T \cong 300$°C). According to experience gained at the continental deep borehole a migration to release zones, i.e. to disturbance zones and overthrust shearing paths occurs and contiguous films of extremely high conductivity develop (ELEKTB-Group 1994).

Let us now try to find starting points how to include these facts and considerations about electric conductivity into existing geotectonic ideas about the Variscan deep structure.

A basic variant of such an idea is represented in Fig. 8 which was presented at the Berg- und Hüttenmännischer Tag 1975 without knowledge about the Mid-German Conductivity Anomaly (Porstendorfer et al. 1976). The initial situation is a Central European ocean with possible rift zones (opening areas). Some hints exist about this

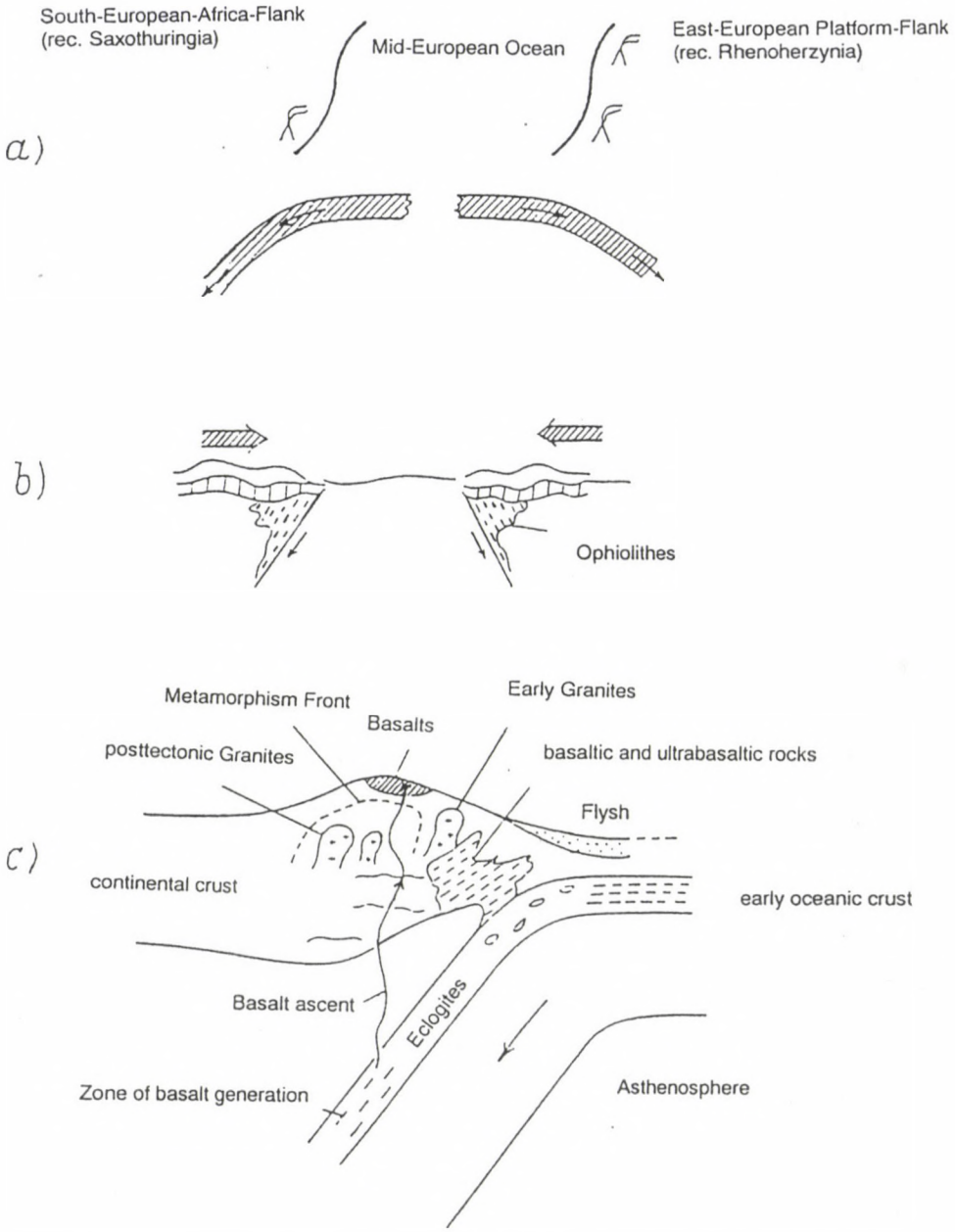


Fig. 8. Plate tectonic idea on the development of the Variscan after Porstendorfer et al. (1976). a) Opening process, b–c) Different stages of the closing process

ocean in form of MORB (mid/oceanic rift basalts, Werner 1993). The oceanic area closed due to the collision of Gondwana and Laurussia till the end of Carboniferous and this closure might lead to corresponding subductions and crustal thickenings. Hints at a so-called "subduction signature" exist from geochemical investigations, too (Wenzel 1995) nevertheless, here caution is advisable as products may be contaminated in the crustal zone. This scenario already contains the necessary sinking zones which are necessary to explain the high electrical conductivity.

A more detailed description of this scenario was given by Franke (1989). His scheme shows that a continental area (terrane) is supposed within this Central European Ocean in form of the Central European Plate which meant the precursor of the later Mid-German Crystalline Zone from the Ordovician on. It is characteristic for the supposed closing process that it happened earlier in the area between Saxothuringicum and Moldanubicum than between Rhenoherzynicum and Saxothuringicum. The crustal thickening in connection with this closure led therefore in the southern part that deep rocks passed through the amphibolite facies and later in consequence of overthrusting they even reached the surface, while the condition in the northern part (i.e. in the border zone between Rhenoherzynicum and Saxothuringicum) might led to the release of fluids.

The extraordinarily clear disturbance tectonics in the area of the Mid-German Crystalline Zone (especially in the Vesser/Tauber element) could result in the frequent occurrence of graphitic zones, thus the second explanation of the conductivity anomalies can also contribute to the anomaly.

A decision in favour of this or that supposition is only possible based on experiments to distinguish electronic and electrolytic conductivity as it has been already started in Pretzschner's work (1993) on surface-near rocks of the Carboniferous of the Ziegenrücker Mulde in Thuringia within the region of the Mid-German Conductivity Anomaly.

References

Behr H I, Dürbaum H I, Bankwitz P eds 1994: *Zeitschr. f. Geol. Wiss.*, 22, 6, 649–764.

Berdichevsky M N, Vanyan L L, Feldmann J S, Porstendorfer G 1972: *Gerl. Beitr. Geophys*, 81, 187–196.

Berktold A 1969: Tiefensondierung mit Hilfe der Variationen des erdmagnetischen und erdelektrischen Feldes längs eines Profils vom Oberpfälzer Wald durch das nördliche Alpenvorland bis zu den Kitzbühler Alpen. Diss. an der Naturwiss. Fak. der Universität Müchen

Berktold A 1995: Regionale Verteilung der elektrischen Leitfähigkeit der Erdkruste in Mitteleuropa. Geophysikalisches Barbara-Kolloquium, Institut für Geophysik der TU Bergakademie Freiberg

Berktold A, Ritter E 1992: The European Geotraverse Atlas Map 12, Low-Resistivity-Anomalies (Central Europe)

Börner R U 1995: Die Leitfähigkeitsverteilung der Erdkruste zwischen Frankenwald und Harz — abgeleitet aus der Inversion magneto-tellurischer Messungen. Diss. Math.-Naturwiss. Fak., TU Bergakademie Freiberg

Bormann P, Bankwitz P, Schulze A 1989: Geophysikalische Ergebnisse und geologische

Konsequenzen tiefenseismischer Untersuchungen in der DDR. Freiberger Forschungshefte, C 440.

Čermák V et al. 1990: In: The European Geotraverse, Integrated Studies, European Science Foundation

Červ V, Pek I, Praus O 1994: Magnetotelluric and magnetovariational measurements in southwest Bohemia, data processing and modelling experiments. Koll. Erdmagnet. Tiefenforschung, Höchst

Eisel M 1995: Interpretation magnetotellurischer Messungen im Umfeld der kontinentalen Tiefbohrung unter besonderer Berücksichtigung lateraler anisotroper Leitfähigkeitsstrukturen. Geoforschungszentrum Potsdam, Scientific Technical Report, 95/13.

ELEKTB-Group (Bahr K et al.) 1994: *Mittelungen der Deutschen Geophys. Gesellschaft*, 4, 2–40.

ERCEUGT-Group (Bahr K et al.) 1990: *Tectonophysics*, 207, 123–139.

Franke W 1989: *Tectonophysics*, 169, 221–228.

Franke W, Giese P, Grosse S, Haak V, Kern H, Mengel K, Oncken O 1990: Geophysical imagery of geological structures along the central segment of the EGT. In: 5. Earth Science Study Centre, Rauischholzhausen

Glover P, Vine F 1995: *Surveys in Geophysics*, 16, 5–36.

Grosse S, Behr H J, Edel J B, Heinrichs T 1992: *Tectonophysics*, 207, 97–121.

Hofer S 1990: Abteilung der Leitfähigkeitsverteilung in der Erdkruste im Bereich der großtektonischen Grenze Rhenoherzynikum-Saxothuringikum östlich des Rheingrabens, aus einer gemeinsamen Interpretation flächenhafter Messungen der Magnetotellurik und der erdmagnetischen Tiefensondierung. Diss. an der Fak. für Geowiss. der Universität München

Hoffmann N, Fluche B, Müller W 1994: Erste Ergebnisse neuer magnetotellurischer Messungen in Nordostdeutschland. Ein Statusbericht. Koll. Erdmagnetische Tiefensondierung, Höchst

Heinrichs T, Giese P, Bankwitz P, Bankwitz E 1994: *Zeitschr. f. Geol. Wiss.*, 22, 771–801.

Jödicke H 1990: Zonen hoher elektrischer Krustenleitfähigkeit im Rhenoherzynikum und seinem nördlichen Vorland. Hochschulschriften, Bd. 24, Münster

Johannes W 1989: *Veröffentl. Niedersächs. Akad. Wiss.*, No. 1, 49–55.

Klause B, Murasch K, Schwarz G 1994: Breitbandige erdmagnetische und magnetotellurische Sondierungen in Westböhmen. Koll. Höchst

Knothe Chr 1972: *Geod. Geoph. Veröff.*, Reihe III, No. 27, 59–68.

Pečová, J, Praus O 1996: *Studia geoph. et geod.*, 40, 50–76.

Peter Ch 1994: Elektrische Leitfähigkeitsmodelle der Erdkurste in Südhannover, Nordhessen und im westlichen Thüringen, abgeleitet aus der elektromagnetischen Induktion durch natürliche Felder. Diss. an der Math.-Naturwiss. Fak. der Universität Göttingen

Porstendorfer G 1973: *Geod. Geoph. Veröff.*, Reihe III, No. 28, 33–40.

Porstendorfer G 1980: *Freib. Forsch. Heft.*, C 356, 29–41.

Porstendorfer G 1995: In: Koll. "Observation and Interpretation of Geomagneatic Variation", U Schmucker ed., Reinhausen/Göttingen

Porstendorfer G, Baumann L, Olszak G 1976: *Zeitschr. Geol. Wiss.*, 4, 1033–1047.

Praus O, Pečová J, Petr V, Babuska V, Plomerova 1990: *Phys. Earth Planet. Int.*, 60, 212–228.

Pretzschner C 1993: Quantitative Modellierung des Einflusses der Polarisierbarkeit auf elektromagnetische Nah- und Fernfeldmethoden. Diss. Fakultät für Math. und. Naturwiss., TU Bergakademie Freiberg

Ritter E 1991: Zusammenstellung von Wiesepfeilen im Rahmen des DFG-Projektes "Die elektrische Leitfähigkeit von Kruste und Mantel im Süden Ostdeutschlands"

Tauber S 1993: Die Leitfähigkeitsverteilung in den nördlichen Varisziden untersucht mit den Methoden der Magnetotellurik und der geomagnetischen Tiefensondierung auf einem Profil vom Oberpfälzer Wald ins Vogtland. Institut für Geologie, Geophysik und Geoinformatik der FU Berlin, Diplomarbeit

Walter R 1992: Geologie von Mitteleuropa. E Schweizenbartsche Verlagsbuchhandlung, Stuttgart

Wenzel Th 1995: Persönliche Mitteilung

Werner C D 1993: Basische Magmatite im Unter- und Mittelharz. Zbl. Geol. Paläont. I, No. 9/10, 1257–1283.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 151–168 (1997)

GEOELECTRICAL STRUCTURE OF THE EARTH'S MANTLE IN PANNONIAN BASIN

V Yu Semenov[1], A Ádám[2], M Hvozdara[3], V Wesztergom[2]

[Manuscript received April 10, 1997]

Data from four geomagnetic observatories: Nagycenk, Tihany, Hurbanovo and Wien-Kobenzl together with magnetotelluric data obtained at 4 points situated not far from these observatories were analyzed to estimate the geoelectrical mantle structure in the Pannonian Basin. Response functions were estimated in the period range from seconds to several hours by the MT method and from 3 days to 1 year by using Dst variations. The published responses obtained from Sq variations were added to remove gaps between MT and MV curves. These deep soundings have confirmed that the upper mantle conductance below the Pannonian Basin is at least twice higher than below the Canadian and Ukrainian shields. The asthenosphere was detected with a conductance of about 4 kS at 80–280 km depths. A conductive zone (20 kS) was required by inversions at the depth of about 500 km. The mid mantle conductive zone was estimated to 550 kS at 850 km depth.

Keywords: deep electromagnetic soundings; mantle conductivity; Pannonian Basin

Introduction

Recently, a local magnetovariational method was developed and by it, reliable results were obtained for the mantle conductivity by Roberts (1984), Schultz and Larsen (1987) and Semenov (1989). The results obtained by this method may be linked to deep magnetotelluric soundings at a regional level. Using this possibility, the geoelectrical structure of the mantle beneath the observatory Tucson in North America (Egbert and Booker 1992), in the stable at Canadian craton (Schultz et al. 1993), in German territory (Bahr et al. 1993) and in an active subduction zone at Sakhalin Island (Semenov and Rodkin 1996) was determined. We present here a new investigation of the mantle conductivity structure in the Pannonian Basin which is considered as a "hot region" in Central Europe (Ádám et al. 1989). Complete deep soundings were carried out in the Hungarian and Slovakian geomagnetic observatories: Tihany, Nagycenk and Hurbanovo (Fig. 1).

[1]Institute of Geophysics, Polish Academy of Sciences, Ks. Janusza 64, PL–01–452 Warsaw, Poland

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

[3]Institute of Geophysics, Dubravska 9, 842 28 Bratislava, Slovak Republic

1217-8977/97/$ 5.00

Fig. 1. Locations of geomagnetic observatories and the MT measurements sites in the Pannonian Basin

Methods and assumptions

Response functions sensitive to the mantle conductivity structure can be obtained by using Dst geomagnetic variations with known field source structure. In the case of a purely first source term ($P_1^0(\cos\vartheta) \sim Z$) of the spherical harmonic expansion, the ratio of vertical (Z) to horizontal (H) field components can be con-

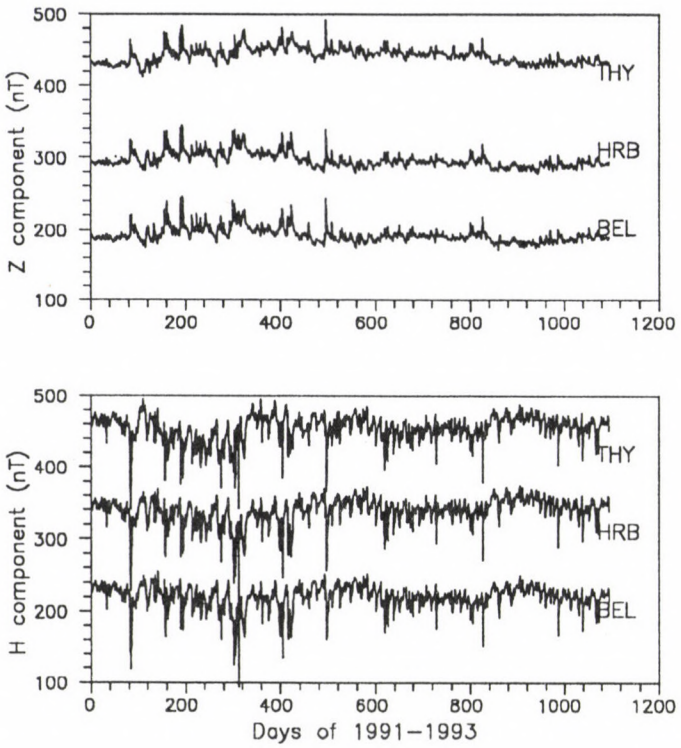


Fig. 2. The daily mean data of observatories: Tihany, Belsk and Hurbanovo in the years 1991–1993

verted into the complex apparent resistivity (Rokityansky 1982):

$$\varrho(\omega) = \frac{i\omega\mu}{4} \cdot R^2 \cdot \tan^2\Theta \cdot \left|\frac{Z(\omega)}{H(\omega)}\right|^2 .$$

The Earth's radius (R) and the geomagnetic co-latitude of an observatory (Θ) must be known for this procedure.

Response functions sensitive to the crust and upper mantle conductivity structures are obtained by assuming a plane wave source model of geomagnetic variations in a wide band of periods shorter than one day. By means of the magnetotelluric method, the components of the impedance tensor can also be transformed into the equivalent tensor of the apparent resistivity (Semenov 1992):

$$\begin{aligned} \varrho_{xx} &= (Z_{xy} \cdot Z_{xy} - Z_{xx} \cdot Z_{yy})/i\omega\mu , & \varrho_{xy} &= Z_{xx} \cdot (Z_{yx} - Z_{xy})/i\omega\mu , \\ \varrho_{yy} &= (Z_{yx} \cdot Z_{yx} - Z_{xx} \cdot Z_{yy})/i\omega\mu , & \varrho_{yx} &= Z_{yy} \cdot (Z_{xy} - Z_{yx})/i\omega\mu . \end{aligned}$$

The principal directions of that tensor can be found by several ways (e.g. as the minimum ratio of the additional (ϱ_{ij}) to main (ϱ_{ii}) resistivities) which are expected to be less disturbed by subsurface inhomogeneities.

To enable a combination of both MT and MV results we have to assume:

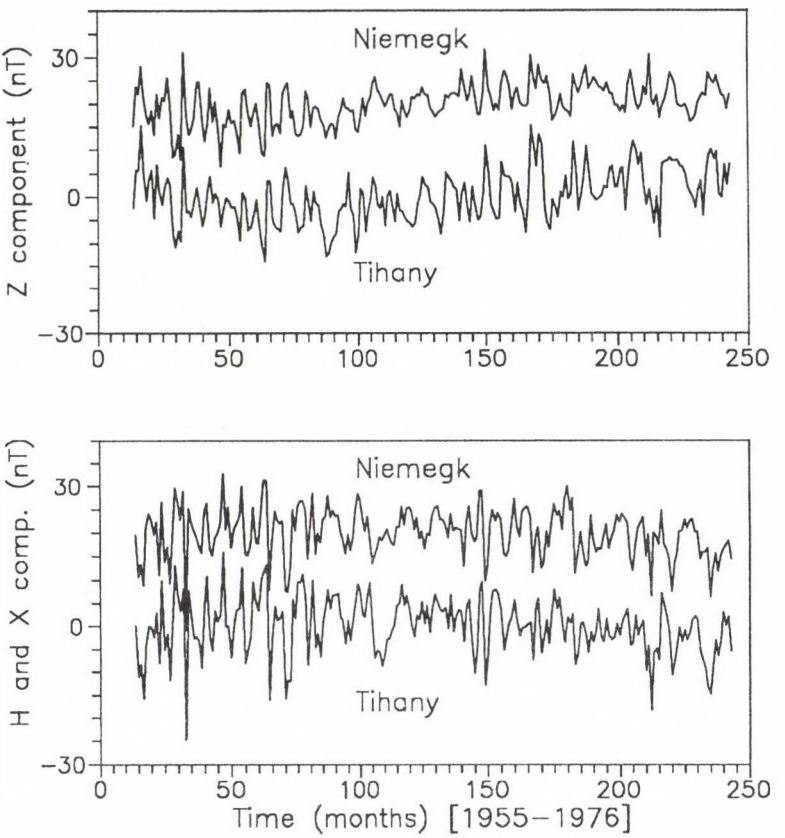


Fig. 3. Example of the monthly mean data at the Tihany and Niemegk observatories in the years 1955–1976

— that the deep mantle medium is isotropic therefore it is possible to combine the scalar MV and tensor MT curves;

— that the region is situated rather far from sharp contrasts in subsurface conductivity (Takeda 1993);

— that the MT results in the chosen direction are only disturbed by the static shift effect at the longest periods.

Then using the coincidence of the phase curves due to apparent resistivity of MT and MV methods, the long period part of MT curves can be shifted until the level required by the regional MV sounding result is reached. As a criterion of the true level for the shift, the minimum difference between the 1D inversion results obtained by different methods (D+ and Occam, for example) can be considered.

Table I. Deep MV sounding results in the Pannonian Basin

Wien-Kobenzl (WIK)*

T day	ϱ Ohmm	Error Ohmm	φ degr.	Err. degr.
73.0	2.40	0.36	−39	9
40.0	2.71	0.27	−55	6
27.0	4.09	0.22	−56	3
18.2	4.55	0.23	−62	3
13.7	5.08	0.36	−61	4
11.0	6.06	0.31	−57	3
9.0	6.54	0.42	−68	4
7.8	8.41	0.73	−60	5
6.9	7.97	1.42	−40	10
4.9	8.28	0.91	−73	6
1.0+	35.70	7.20	−48	15
0.5+	43.40	8.70	−47	15

Hurbanovo (HRB)

T day	ϱ Ohmm	Error Ohmm	φ degr.	Err. degr.
474.1	0.55	0.20	−80	16
158.0	2.20	1.16	−60	22
51.0	4.50	1.35	−58	20
21.0	5.68	0.68	−62	16
9.9	8.76	1.50	−58	12
4.9	12.24	2.70	−68	14
2.0	23.30	8.16	−72	14
1.0+	35.70	7.20	−48	15
0.5+	46.90	9.40	−49	15

Tihany (THY)

T day	ϱ Ohmm	Error Ohmm	φ degr.	Err. degr.
262.2	1.27	0.36	−38	21
154.2	2.28	0.62	−57	13
90.7	3.78	1.20	−54	19
53.5	2.77	0.59	−52	12
31.5	4.07	1.05	−59	11
18.4	5.23	0.95	−76	10
10.9	7.02	0.74	−68	6
6.3	13.01	1.26	−77	5
4.0	14.63	2.55	−74	9
1.0+	35.70	7.20	−48	15
0.5+	46.80	9.40	−48	15

Hurbanovo (HRB)*

T day	ϱ Ohmm	Error Ohmm	φ degr.	Err. degr.
73.0	2.08	0.36	−31	10
40.0	2.90	0.35	−46	7
27.0	3.66	0.24	−47	4
18.2	3.92	0.28	−54	4
13.7	3.98	0.41	−52	6
11.0	4.98	0.39	−52	4
9.0	5.66	0.42	−53	4
7.8	8.42	0.85	−51	6
6.9	6.27	1.09	−47	10

+Olsen (1994)
*Roberts (1984)

Data processing

The daily mean values of geomagnetic field at the observatories Tihany (THY) and Hurbanovo (HRB) were used for soundings together with the data of the reference Polish observatory Belsk as shown in Fig. 2. Besides, the 20 years long time series of monthly mean values for these observatories were compared with the Niemegk (NGK) data. Examples of the simultaneous monthly averages are shown in Fig. 3.

The first step in the analysis was a rotation of the reference systems for the observatories to determine into the direction of the maximum Dst field (Schultz and Larsen 1983). The following mean values were found: −19° for Hurbanovo, −22° for Belsk, −19° for Niemegk and −9° for Tihany. The agreement between the horizontal Dst fields and the first spherical harmonic model of the source was

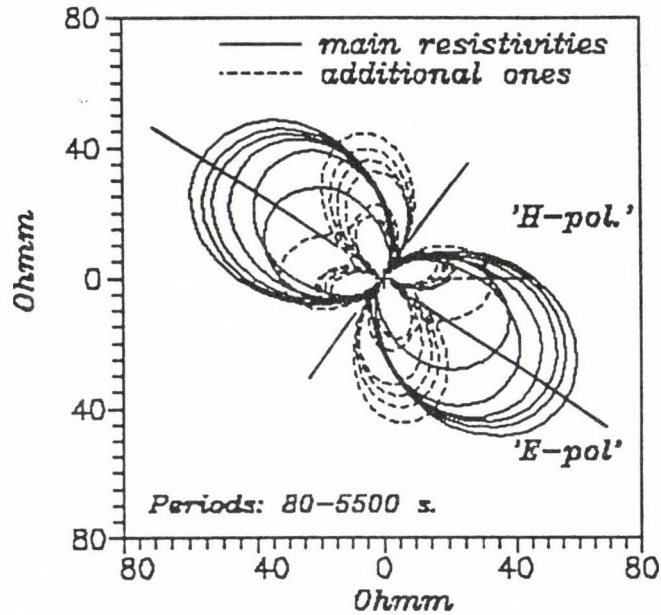


Fig. 4. Polar diagrams of the apparent resistivities and the chosen directions for interpretation (Litke)

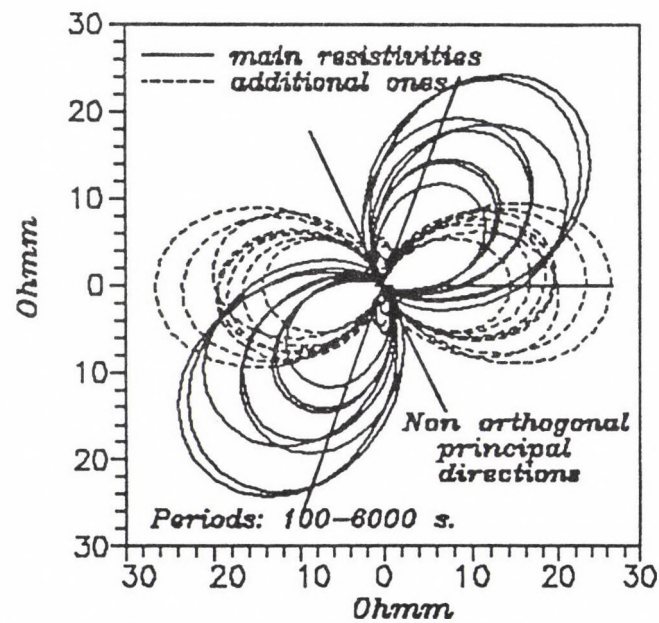


Fig. 5. Polar diagrams of the apparent resistivities and the chosen directions for interpretation (Srobarova)

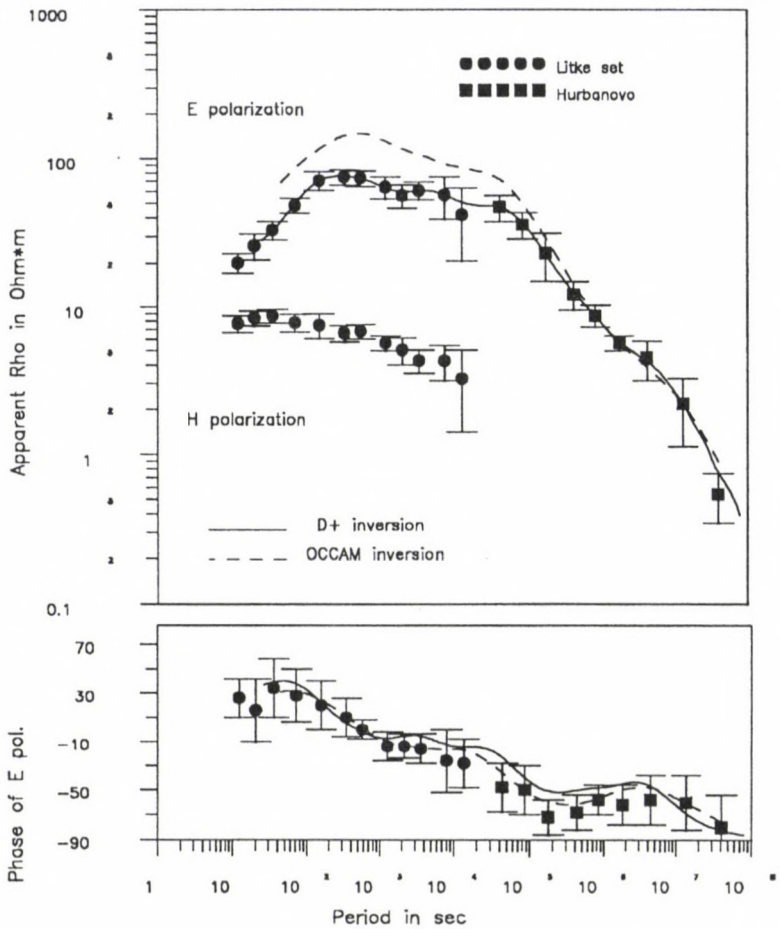


Fig. 6. Comparison of original experimental responses and inversion results for Litke with the Hurbanovo observatory

tested (if the coherence were more than 0.9). The period ranges which were in contradiction with the source model were excluded from the further analysis.

The transfer functions between the vertical (Z) and horizontal (H) components were computed only if the coherence values were greater than 0.7 between them. The amplitude of these transfer functions were corrected by the improved remote reference method (Semenov and Zakharova 1989). The yearly oscillations were removed during the data processing of the monthly mean values. The transfer functions satisfying all these criteria were converted into apparent resistivity. Finally, the robust technique was applied for estimations at each observatory and the general results are presented in Table I together with converted results obtained by Roberts (1984) and Olsen (1994).

The long period magnetotelluric data were partly obtained during the interna-

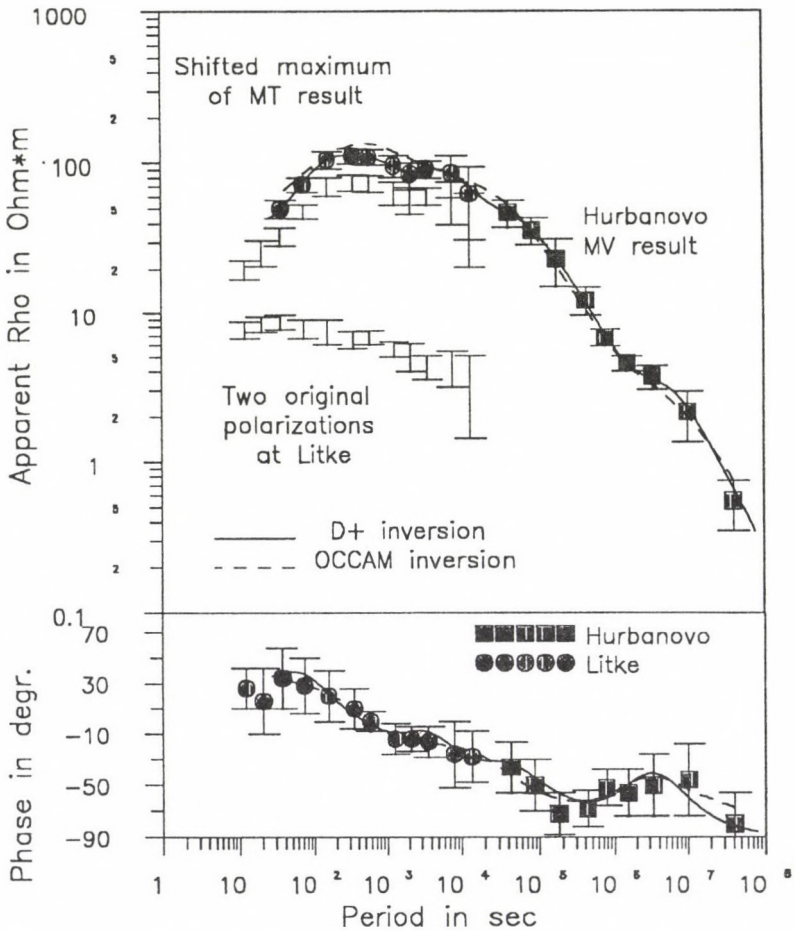


Fig. 7. The comparison of shifted MT responses and inversion results for Litke with the Hurbanovo observatory

tional investigations across the Tatra Mountains in 1994 (Ernst et al. 1997). We have used here two sets of the MT measurements from Litke in Hungary and from Srobarova in Slovakia (see Fig. 1). The polar resistivity diagrams are presented in Figs 4, 5 together with the directions of the minimum additional resistivities chosen for interpretation. Besides, published MT sounding results of Tihany (Ádám 1966) and of Nagycenk (Ádám 1993) were used for the joint observatory data interpretation.

Analysis and inversions

The MT results obtained at Litke were combined with the previously averaged MV results from the nearest observatory Hurbanovo (see Table I). The maximum curve from that MT sounding was chosen for the joint inversion due to the agreement between phases and between levels of the apparent resistivity curves (see Fig. 6).

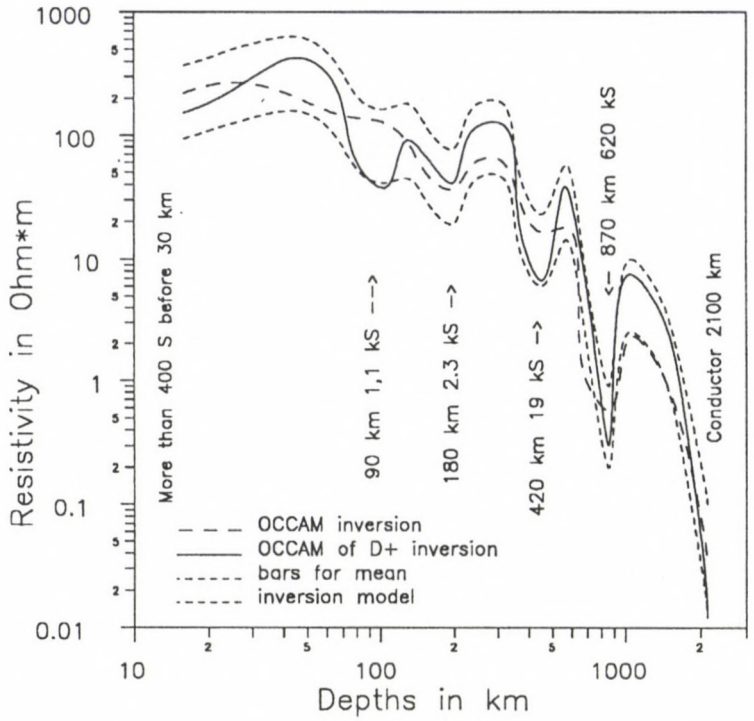


Fig. 8. The mean mantle geoelectrical structure at Litke

As a first step Parker's inversion (D+ algorithm) (Parker and Whaler 1981) was used for the estimation of the deep conductivity structure. It was established that a one dimensional model can be constructed for all periods simultaneously (Fig. 6).

However, as the D+ model gives the sharpest distribution, the Occam inversion method (Constable et al. 1987) was used to construct the "minimum structure" of the mantle conductivity. This inversion, in contrast with the D+ one, is more sensitive to the phase of the complex apparent resistivity. This algorithm has shown a discrepancy between the phase curve and the resistivity level of the MT result (see Fig. 6). Then, the resistivity modulus was shifted by about 1.5 times, as required by the Occam inversion. The new comparison of the shifted result and its inversions are shown in Fig. 7.

Both kinds of inversion results are presented for the station Litke in Fig. 8. The two geoelectrical conductivity distributions give us possibility to construct a mean model and indirectly estimate its error bars. The bar boundaries were assumed for these data twice less and twice more than the mean model at every depth as shown in Fig. 8. The real D+ inversion results are written at this figure. It is seen there that all obtained inversions are in agreement within the error bars of the mean model. The accuracy could be improved either by the higher quality of the initial data or by comparison of several independent results in a region.

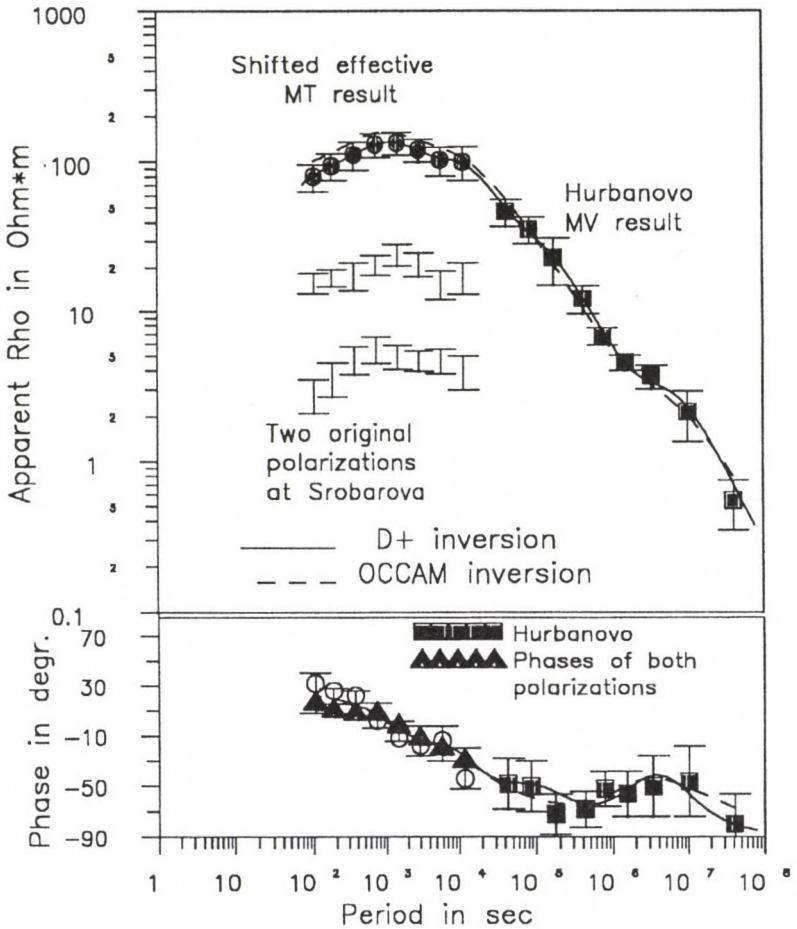


Fig. 9. The comparison of shifted MT responses and inversion results for Srobarova with the Hurbanovo observatory

The MT sounding results at Srobarova (near the Hurbanovo observatory in the Slovak Republic) are presented in Fig. 9. Using the coincidence of the phases for both chosen directions the effective curve was obtained (by averaging the complex apparent resistivities). In accordance with a coincidence between the phase curves obtained by MT and MV methods, the effective MT module values were shifted by 12 times until the level required by the MV result at Hurbanovo. Using both kinds of inversions for the joint result the geoelectrical structure was estimated in the same way as it was described for Litke station. The misfits for these data are shown in Fig. 9 and the mean mantle structure is presented in Fig. 10.

The same principles were applied in the case of the MT sounding at Nagycenk, published by Ádám (1993). The MT result was combined with the MV result obtained by Roberts (1984) for the Wien-Kobenzl observatory as shown in Fig.

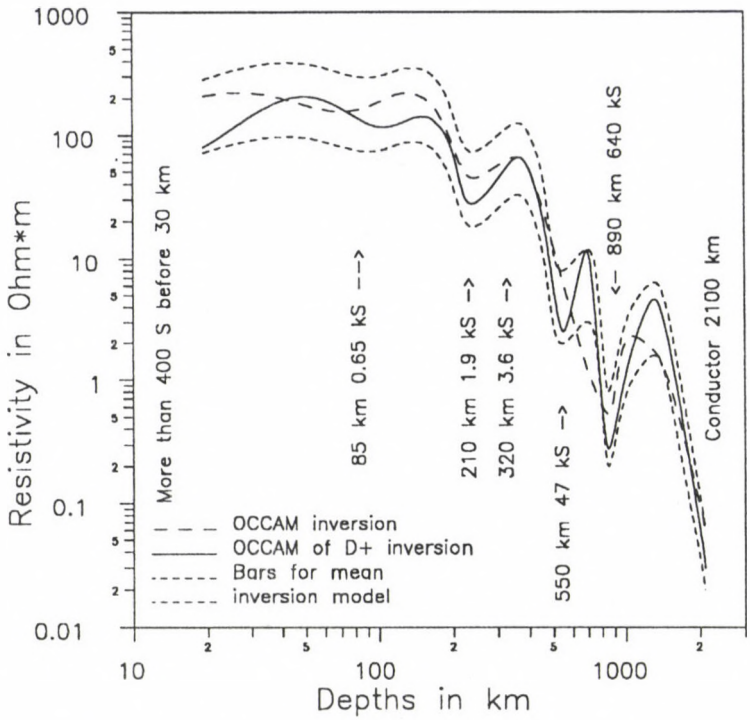


Fig. 10. The mean mantle geoelectrical structure at Srobarova

11. After shifting the effective MT results by about 3.2 times the joint data were inverted and the misfits are also shown in Fig. 11. The mean model is presented in Fig. 12, where the bar boundaries were reduced by 1.5 times in comparison with the previous result.

Finally, the MT result at Tihany observatory published by Ádám (1966) was combined with the MV result from this observatory data (see Table I). Using the D+ inversion the apparent resistivity phases were estimated from the experimental module values. The effective MT curve was combined with the MV sounding without any shifting at this observatory as it shown in Fig. 13 together with the inversion responses. The mean model is presented in Fig. 14 where the bar boundaries were assumed as for MT station Litke.

The most critical part of the joint inversions of the deep MT and MV sounding curves is the fitting of these curves to each other. As it is known, the MT apparent resistivity curves could be strongly distorted by near-surface inhomogeneities partly by electric charges (static shift), partly by induction effects. A very crucial example for this problem is the distortion of the sounding curves in Tihany observatory lying near to a great fracture (so-called Balaton line) along which rocks of different resistivity meet and therefore the sounding curves of E and B polarizations strongly differ. The B polarized resistivity values strongly increase or decrease over

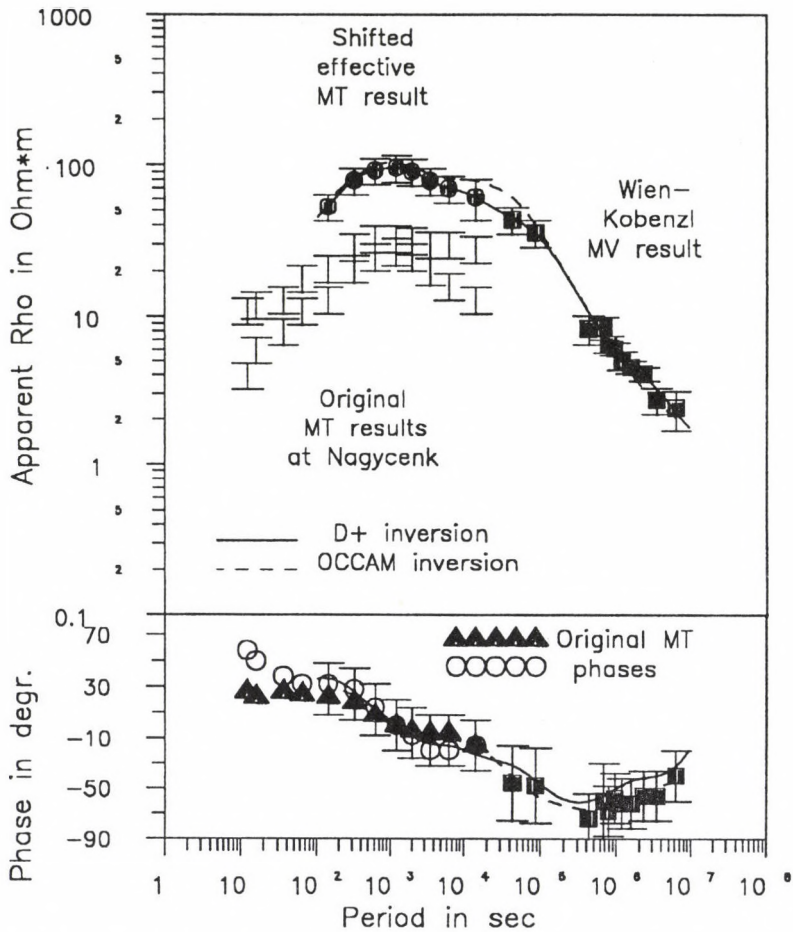


Fig. 11. The comparison of shifted MT responses and inversion results for Nagycenk with the Wien-Kobenzl observatory

or under the real resistivity value depending on which part of the tectonic line the measurement has been made. The effect of this tectonics has been confirmed by measurement in a nearby sounding site (Aszófő). Due to this distortion the upper mantle resistivity structure obtained by inversions also strongly differs in Tihany from the other ones (see Fig. 15).

The shift of the MT sounding curves — carried out to get a better fit with the MV curves — effects the surface S-value which is then different from the locally determined one and causes a change in the parameters of the deeper structures, too (e.g. in the depth in the asthenosphere and in its conductance). As in the case of the studied sounding curves, the shift was always upwards (towards greater MV apparent resistivity) therefore, the depth values are greater and conductances are less than those determined from the original MT curves. Thus, our inversion model

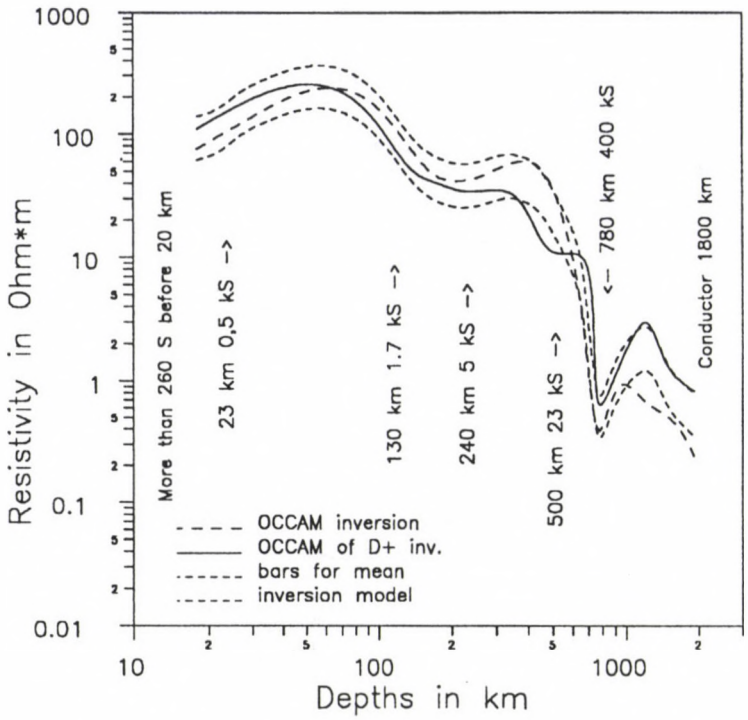


Fig. 12. The mean mantle geoelectrical structure at Nagycenk

can be called as a *minimum conductivity structure model* for the Pannonian Basin.

The distortion effect of the MT curves in the deeper part of the geoelectrical structure — e.g. in the conductivity increase at a depth of about 500–800 km — disappears.

As Olson's MV values attached to the MT ones are regional responses (Olsen 1994), the fitting could be improved if these values would be derived from the local Sq values measured at the observatories in question.

Interpretation

All types of inversions for the four considered deep sounding have confirmed the existence of a conductive layer in the mid mantle (depth about 800–900 km) with a conductance of about 550 kS. If we suggest that the thickness of this layer is about 150 km, then its conductivity is like that of the ocean water. Maybe a more striking fact is the relatively high resistivity in the lower mantle just beneath this layer with the maximum at about 1200 km depth. Deeper, the lower mantle conductivity is increasing and the resolving power is not enough to distinguish any features of its structure.

The upper mantle structure was defined more precisely by MT and MV methods. The comparison of all obtained models has shown that except the Tihany model

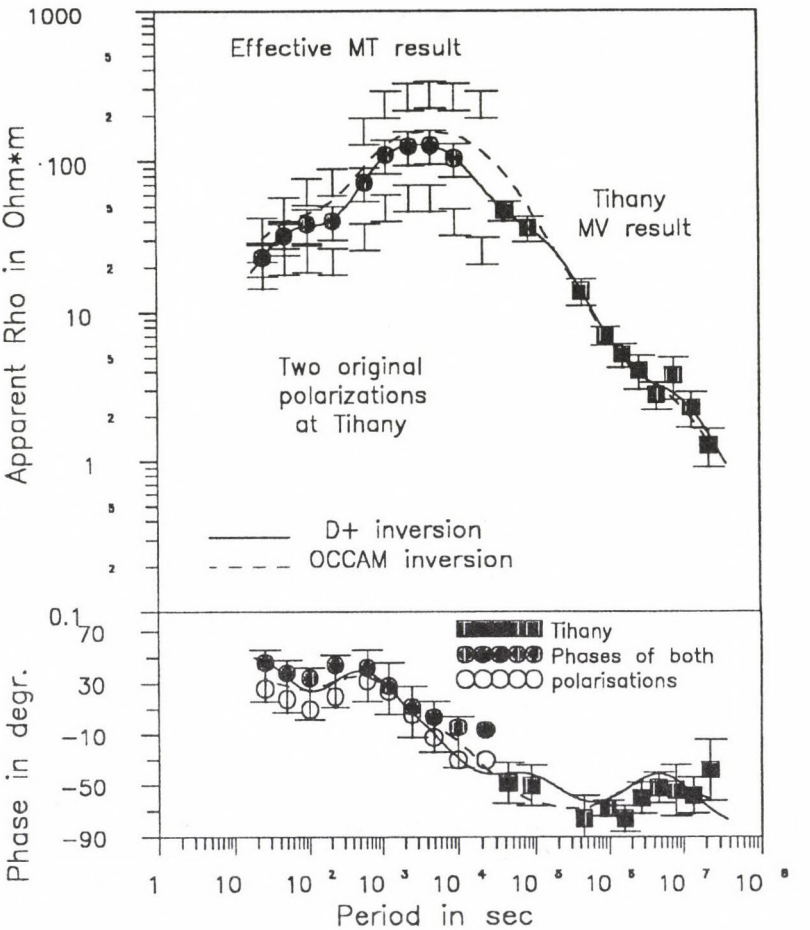


Fig. 13. The comparison of effective MT responses and inversion results for the Tihany observatory data

these results are similar. That is why we averaged the models for Litke, Srobarovo (HRB) and Nagycenk (NCK). The averaged model is presented in Fig. 16. This model shows the existence of a rather high subsurface conductance, which could be one of the sources of the regional shift in the Pannonian Basin (surrounded by resistive mountains).

The main feature of this region is the rather highly conductive upper mantle (about 25 kS) which is twice higher than that for the stable Canadian craton (about 10 kS) (Schultz et al. 1993) and three times higher than that for the Ukrainian shield (about 8 kS) (Semenov et al. 1997). Within the upper mantle in the Pannonian Basin three conductive zones are observed: at the depths 80–120 km (about 1 kS), at the depths 200–280 km (about 3 kS) and about 450-500 km depths (20 kS). The first two zones are associated with the asthenosphere being here much shallower

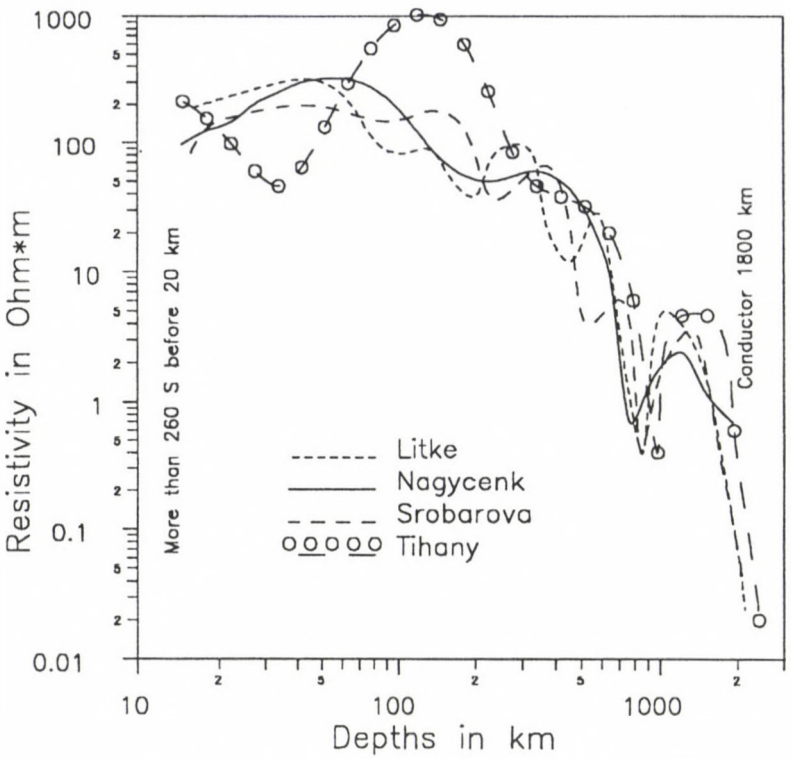


Fig. 14. The mean mantle geoelectrical structure at Tihany

than below the East European plate (180 km depth). This fact can be interpreted as follows: the bottom of the East European lithospheric plate is situated at depths of about 150 km while the same boundary in the Pannonian Basin is at about 70–80 km. This conclusion is confirmed by the heat flow data, too (Ádám 1993).

Finally the mean geoelectrical structure for the Pannonian Basin can be compared with the other published results including the deep sounding at the Tucson observatory situated not so far from a rift zone (Egbert and Booker 1992). After converting the model to conductivity values, this comparison is presented in Fig. 17. The 920 km seismic discontinuity was recently detected (Kawakatsu and Miu 1994).

Discussion

The comparison of three results for different geotectonic situations (Fig. 17) shows that the increase of the mantle conductivity is not monotonic with depth for all regions. However, the highly resistive lithosphere is followed by the conductive asthenosphere everywhere. There are differences in the depths and conductivities of these zones. The next significant increase of the conductivity is between about

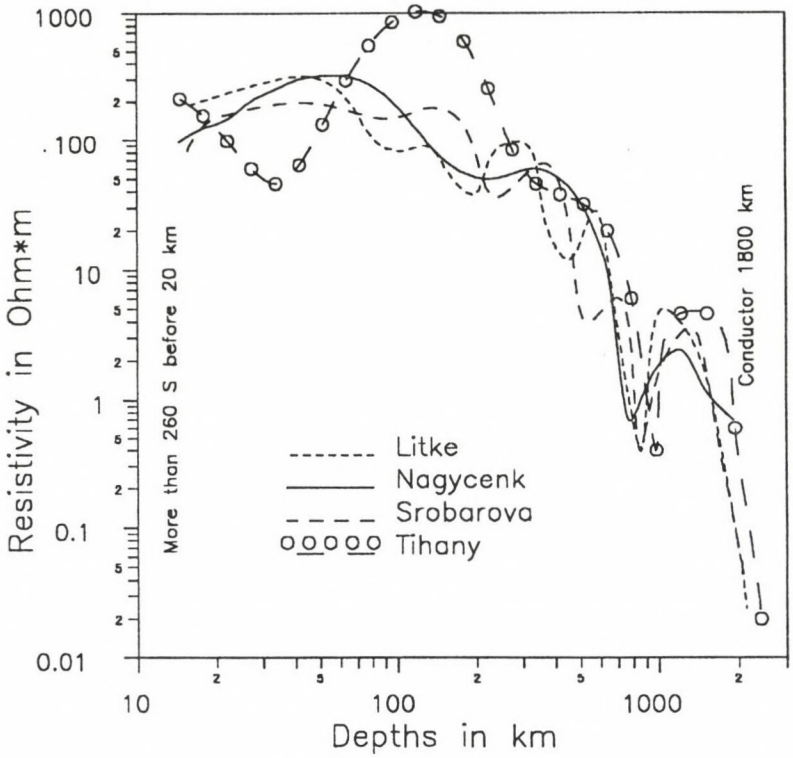


Fig. 15. The comparison of all estimated mean mantle structures in the Pannonian Basin

400 km (Ádám 1980) and about 800 km, i.e. the upper-lower mantle boundary. The main feature of the mid mantle is the conductive layer situated beneath the transition zone, which was predicted as a global phenomenon (Zharkov 1983) due to phase transition of the olivine minerals (Mg, Fe) SiO_4 to the spinel structure. Perhaps, the middle mantle geoelectrical structures are slightly different for plates or shields from those in rift zones.

If we compare geoelectrical structures with known seismic discontinuities (Zharkov 1983) the first increase of conductivity (see Fig. 17) may be associated with the bottom of lithosphere (from about 80 km for rift zones through 140 km for platform to 180 km for stable cratons). The slight plateau would be in accordance with a seismic asthenosphere (from the bottom of lithosphere to the upper edge of the mantle transition zone), which may be layered according to seismic as well as geoelectrical results. The next significant increase of the mantle conductivity would be in accordance with the mantle transition zone (420–670 km depth). In this case, the conductive layer at the mid mantle may be associated with a zone between the lower — upper mantle boundary and the recently detected seismic boundary at a depth of 920 km (Kawakatsu and Niu 1994). The suggested analogy can be considered as an initial step for future investigations of the Earth's mantle.

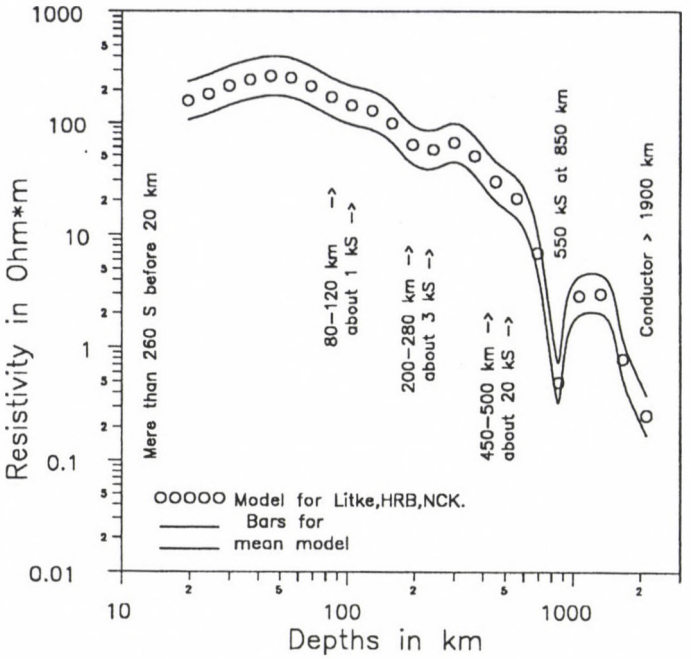


Fig. 16. The mean geoelectrical structure of the mantle in the Pannonian Basin according to all results with the exception of Tihany

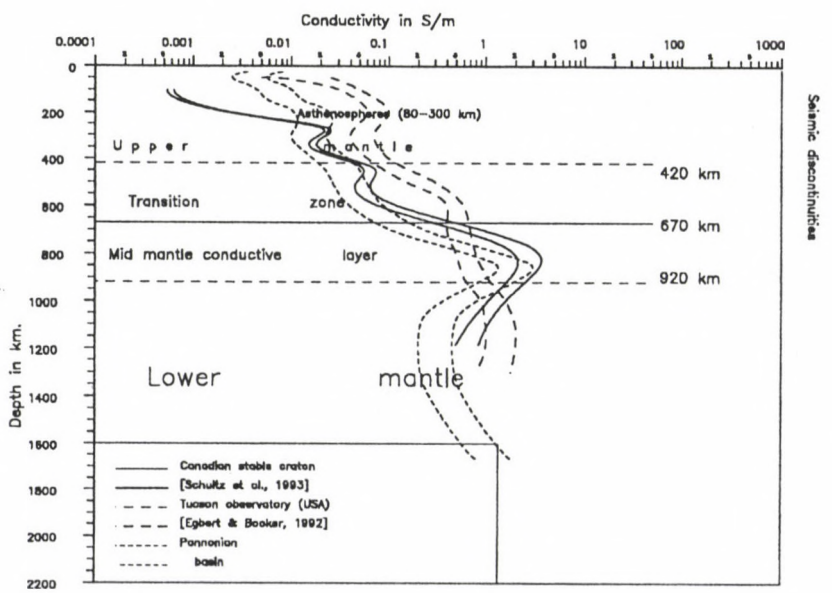


Fig. 17. Comparison of geoelectrical mantle structures at Pannonian Basin, Canadian craton and Tucson observatory

Acknowledgements

This research was supported by the grant No 6 P04D 03512 of the Polish Committee of Scientific Investigation and by the grant of the Hungarian Research Found (OTKA) No T014882. The authors thank to Mr A Körmendi for supplying the geomagnetic data of the Tihany observatory.

Rererences

Ádám A 1966: *Geofizikai Közlemények*, 17, 39–50.
Ádám A 1980: *J. Geomag. Geoel.*, 32, 115–124.
Ádám A 1993: *Acta Geod. Geoph. Mont. Hung.*, 28, 151–195.
Ádám A, Landy K, Nagy Z 1989: *Tectonophysics*, 164, 361–368.
Bahr K, Olsen N, Shankland T J 1993: *Geophysical Res. Letters*, 20, 2937–2940.
Constable S C, Parker R L, Constable C G 1987: *Geophysics*, 52, 289–300.
Egbert G D, Booker J R 1992: *J. Geophys. Res.*, 97, 15099–15115.
Ernst T, Jankowski J, Semenov V Yu, Ádám A, Hvozdara M, Jozwiak W, Lefeld J, Pawlishin J, Szarka L, Wesztergom V 1997: Electromagnetic soundings across the Tatra Mountains. *Acta Geoph. Polon.* 45 (in press)
Kawakatsu H, Miu F 1994: *Nature*, 371.
Kharin E P, Semenov V Yu 1989: *Il Nuovo Cimento.*, 12, 547–553.
Olsen N 1994: *DGG*, 15, 66–73.
Parker R L, Whaler K A 1981: *J. Geophys. Res.*, 86, 9574–9584.
Roberts R G 1984: *Geophys. J. Roy. Astr. Soc.*, 78, 547–572.
Rokityansky I I 1982: Geoelectromagnetic investigation of the Earth's crust and mantle. Springer-Verlag, Berlin, New-York
Schultz A, Larsen J C 1983: *J. Geomag. Geoel.*, 35, 835–846.
Schultz A, Larsen J C 1987: *Geoph. J. Roy. Astr. Soc.*, 88, 733–761.
Schultz A, Kurtz R D, Chave A D, Jones A G 1993: *Geophysical Res. Letters*, 20, 2941–2944.
Semenov V Yu 1989: *Fizika Zemli*, 25, 221–226.
Semenov V Yu 1992: In: Proceed. of the Symposium of 10 years Finnish-Soviet co-work in geoelectrics, University of Oulu, Geoph. Dept., Report 18, 96–103.
Semenov V Yu, Rodkin M V 1996: *J. Geodynamics*, 21, 355–364.
Semenov V Yu, Zakharova O K 1989: In: Report 16 of the Oulu University, Geoph. Dept., Finland, 85–89.
Semenov V Yu, Kulik S N, Logvinov I M 1997: Estimations of the geoelectrical structure of the mantle in Ukraine. Trans. of Ukrainian Academy of Science (in press).
Takeda M 1993: *Geophys. J. Int.*, 114, 289–292.
Weidelt P 1972: *Geophysik*, 38, 257–289.
Zharkov V N 1983: Internal structure of the Earth and Planets (in Russian). Nauka, Moscow

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 169–191 (1997)

ROLE OF SOLAR WIND WAVE EVENTS IN THE GENERATION OF GEOMAGNETIC PULSATIONS

T A PLYASOVA-BAKOUNINA[1]

[Manuscript received April 23, 1997]

In contrast with the popular model of two sources of Pc pulsations, "external" and "internal", a new concept assuming a common extra-magnetosphere origin is suggested, the idea being that the two distinct types of upstream waves in the solar wind lead to two types of Pc pulsations on the ground.

The transverse monochromatic $F - B$ dependent upstream waves are presumed responsible for generation of the solar wind-controlled Pc_{sw} pulsations, while the multichromatic highly compressional upstream waves are presumed responsible for generation of the magnetosphere-controlled Pc_{mg} pulsations. "High latitude entry" for Pc_{sw} and "direct entry" through the sub-solar magnetopause into the magnetosphere for Pc_{mg} is suggested and discussed. Experimental evidence supporting these hypotheses is presented.

Keywords: geomagnetic pulsations; interplanetary magnetic field; magnetospheric resonator; solar wind

Introduction

Geomagnetic pulsations play an important role in the physics of near-Earth cosmic space for understanding of the solar-terrestrial relationship and studying of the pathways of the penetration of the wave energy of the solar wind into the magnetosphere.

Geomagnetic pulsations are the oscillations of the geomagnetic field in the milliherz frequency band. The main part of it is the regular Pc2-4 pulsations with the range of periods from 5 to 200 seconds. They are most of the time observed on the sun-lit side of the Earth as a continuous regular wave form signal lasting several hours, as a result of which they have received the name "pulsations continuous".

In Table I, all types of mechanisms proposed for the origin of Pc2-4 are classified according to the location of the sources:
— at the solar wind
— on the magnetopause, and
— inside the magnetosphere. In the earliest hypothesis (Dungey 1954), pulsations are considered as oscillations of a magnetospheric resonator. The proposed source of excitation of the resonator is a Kelvin-Helmholz (K-H) instability on the magnetopause (Dungey 1962). This mechanism is developed in a number of papers (Southwood 1968, 1974, Chen and Hasegawa 1974, Kivelson and Pu 1984, Kivelson and Southwood 1986, Southwood and Kivelson 1990, Fedorov 1992), where the

[1]Institute of Physics of the Earth, Moscow, Russia

1217-8977/97/$ 5.00

Table I. Location of the sources of geomagnetic pulsations: in the solar wind, on the magnetopause and inside the magnetosphere

Solar wind	Magnetopause	Inside magetosphere	Authors
		generation eigen oscillation magnetosph. resonator Pc2–5	Dungey 1954
	generation (K-H instability)		Dungey 1962
	generation	wave transformation into Alfvén and compressional Pc2–5	Southwood 1968, 1974, Chen and Hasegawa 1974 Kivelson and Pu 1984, Southwood and Kivelson 1990 Fedorov 1992
		generation Pc2–5 (drift, waves)	Hasegawa 1979, Pochotelov and Pelipenko 1976
		generation on plasmapause	Ong and Roderick 1972, Buldurev et al. 1974 Kovner et al. 1976b, Maltzev 1978
generation of monochrom. waves Pc2–4_{sw}		Pc2–4_{sw}	Troitskaya et al. 1971, Plyasova-Bakounina 1972 Greenstadt 1972, Gul'elmi 1974 Russell and Fleming 1976, Verő 1979 Varga 1980, Odera 1986, Yumoto 1988 Engebretson et al. 1990, 1991
generation of monochrom. waves Pc2–4_{sw}	amplification of surface waves driven by K-H instability	Pc2–4_{sw}	Plyasova-Bakounina 1992 Engebretson 1995a Anderson and Engebretson 1995
generation of multichromatic highly compres. wawes		amplification on eigen frequencies of magn. resonator Pc2–5_{mg}	

mechanism of transformation of magnetosonic waves into Alfven and cavity modes are considered.

Hasegawa and Chen (1974), Pochotelov and Pelipenko (1976) place the source of the generated waves deep inside the magnetosphere, where drift instabilities of different types are developed.

Maltsev (1971), Ong and Roderick (1972), Buldurev et al. (1973) and Kovner et al. (1976b) place the source of the waves at the plasmapause, where a K-H instability can be developed.

All of these theories explain very well the resonance features of pulsations, but cannot explain the relationships between Pc and parameters of the solar wind and IMF.

Troitskaya, Plyasova-Bakounina and Gul'elmi (1971) were the first to place the source of the Pc pulsations in the solar wind near the bow shock, where the monochromatic upstream waves of 5–100 s period are observed (Fairfield 1969, Russell and Hoppe 1983). Later this idea of extra-magnetospheric origin was developed in a number of papers (Plyasova-Bakounina 1972a, 1972b, 1992, Greenstadt 1973, Gul'elmi 1974, Russell and Fleming 1976, Verő 1979, Varga 1980, Odera 1986, Yumoto 1988, Engebretson et al. 1990, 1991, Verő et al. 1994).

I. Relationship of micropulsations Pc2-4 to parameters of the solar wind and IMF

The idea of extra-magnetospheric origin of Pc arises from the relationship between the period of the pulsations and the magnitude of IMF B.

Figure 1a shows the dependence of T on B, which is approximated by the formula

$$T = 160/B\,,$$

where T is period of Pc 2-4 in seconds and B is the magnitude of IMF in nT (Troitskaya et al. 1971). A similar form of dependence exists between the gyroperiod of protons in the magnetic field, T_B, and B, the magnitude of the magnetic field: $T_B = 2\pi \cdot m_p c/eB$, where m_p is mass of a proton, e is charge of an electron and c is the velocity of light. Inserting these values in the equation gives $T_B = 65/B$. This relationship combined with our experimental result suggests the relationship $T = 2.5T_B$. Kovner et al. (1976a) consider the generation of Pc waves as due to two-beam instability (solar wind and beam of upstream accelerated protons). Via routine procedures, we arrive at the implied dispersion equation. Solution of this equation gives the dependence of the growth rate of the excited waves on frequency ω. The dependence has a maximum at $\omega = 0.4\omega_B$, where ω_B is the proton gyrofrequency, or $T = 2.5T_B$. This relationship is shown on Fig. 2, together with observed T data from the ground-based geomagnetic station Borok, versus T_B. The T_B values were calculated using B data which were recorded simultaneously with ground-based pulsation data at Explorer-34 the figure shows excellent agreement between theory and observation. A similar dependence, namely $T = 3T_B$, was observed by Fairfield (1969) for the Alfven upstream waves. These waves were

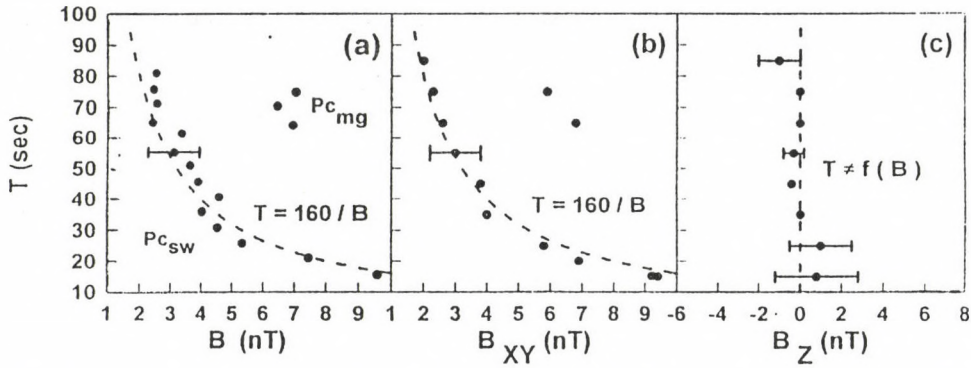


Fig. 1. Dependence of periods (T) of Pc2-4 pulsations at Borok station on a) B-magnitude of IMF, b) on B_{XY} (projection of B on xy-plane), and c) on B_Z (vertical component of B). T is in sec, B is in nT. IMF data were collected by IMP-3 between September-November 1965 and IMP-4 between September-December 1966. Horizontal lines are scattering bars with values shown being typical of non-plotted values. The dashed line is function $T = 160/B$. Pc_{sw} and Pc_{mg} are explained in the text

recorded in the solar wind near the bow shock and are supposed to arise due to the beam instability of the accelerated protons reflected from the bow shock.

Figures 1b and 1c show the dependence of T on B_{XY} (the projection of B on the ecliptic plane) and T on B_Z (component of B perpendicular to the ecliptic plane). It is seen that a close relationship exists between T and B_{XY} (actually, the relationship between T and B_X, the component of B towards the sun, is markedly closer than between T and B_Y), and that no relationship exists between T and B_Z (Plyasova-Bakounina 1992).

Other parameters of the solar wind (density of protons, velocity, temperature, dynamic pressure, etc.) are of no significance for periods of Pc2-4 pulsations (Fig. 1 in Gul'elmi et al. 1973).

Amplitudes of Pc2-4 on the ground are strongly controlled by orientation of IMF (azimuthal angle φ), exactly as the amplitudes of the upstream waves in space are controlled (Fig. 1 in Plyasova-Bakounina 1972a, also Bolshakova and Troitskaya 1968, Webb and Orr 1976, Wolf et al. 1980, and Wolf et al. 1987, Verő et al. 1995).

Amplitudes of Pc2-4 also are controlled by the speed of the solar wind (Saito 1964, Vinogradov and Parchomov 1974, Kovner et al. 1976a, Wolf et al. 1980, Wolf et al. 1987) but the time scale of influence is measured in days in contrast with the influence of orientation of IMF on amplitude, which is measured in minutes.

II. Distribution of the intensity of Pc2-4 along the ground surface

Based on Fig. 1a, the whole class of Pc 2-4 pulsations can be divided into 2 types:

a) *solar wind-controlled* pulsations Pc_{sw} (index sw means solar wind) — pulsations whose period strongly depends on B (points near the line $T = 160/B$), and

b) *magnetosphere-controlled* pulsations Pc_{mg} (index mg means magnetosphere) — pulsations the periods of which are not related to B (points far away from the line $T = 160/B$). The unambiguous indication of two types of pulsations also was made by Miletits et al. (1988).

The location of the sources of both types of Pc can be determined by investigation of the distribution of the intensity of Pc 2-4 along the ground surface. Plyasova-Bakounina et al. (1984, 1986) investigated the meridional distribution of the intensity of Pc 2-4 separately for Pc_{sw} and Pc_{mg}, mainly for stations at middle and high latitudes (53°–77° N), the location of the network of stations being shown on Fig. 3. The analytical procedures were described in detail in Plyasova-Bakounina et al. (1986).

Figure 4 shows examples of the distribution of intensity for Pc_{sw}, the peculiarity of which is the "polar" maximum of distribution of intensity, which is located at latitudes higher than 74° in the region of approximate projection of the day-time polar cusp. This result later received support from Bolshakova and Troitskaya (1984), Morris and Cole (1987), and Engebretson et al. (1989). We suggested that tangential flowing hot solar wind plasma can penetrate directly into the cusp

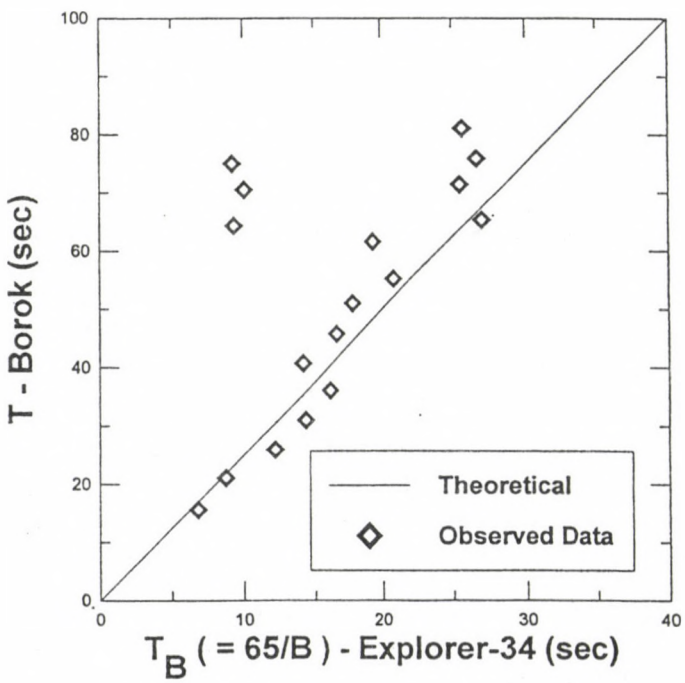


Fig. 2. Dependence of periods of Pc2-4 pulsations recorded on the ground at Borok on the gyroperiod of protons in the Interplanetary Magnetic Field (data from Explorer-34 collected between September-November 1967). Line shows the theoretical dependence of the periods of the excited waves on the gyroperiod of the protons, $T = 2.5T_B$, a relationship developed by consideration of the two-beam instability of the solar wind. See also text

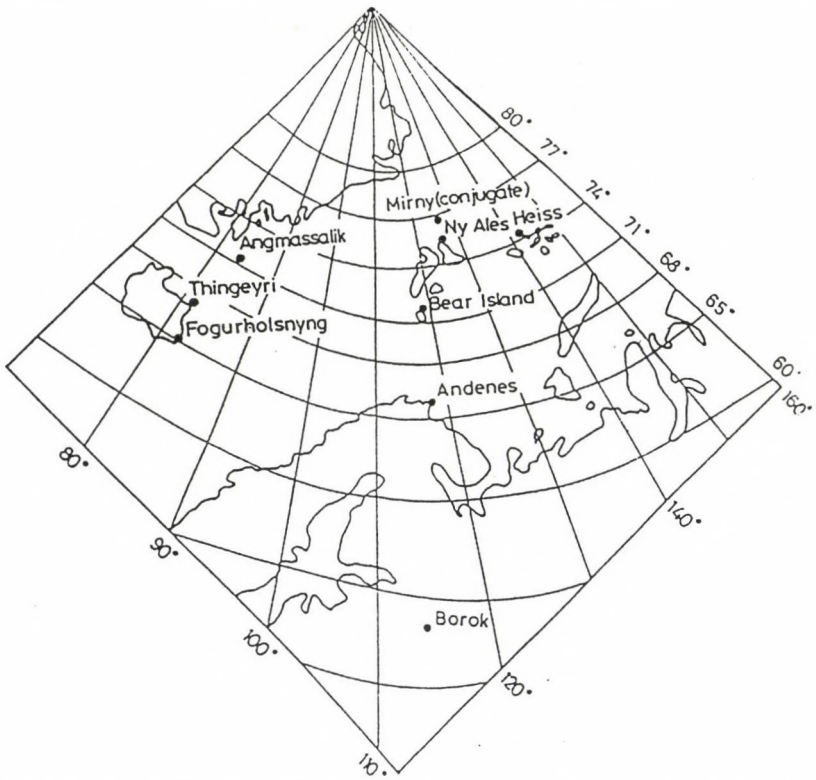


Fig. 3. Location of 9 geomagnetic stations: Ny Alesund, Bear Island, Andenes, Angmagssalik, Thingeyri and Fogurholsnyng (Siegen U., Germany); Borok, Heiss and Mirny (Inst. Phys. Earth, Moscow, Russia)

region, carrying any periodic oscillations deep within the magnetosphere. We follow Lanzerotti et al. (1972) in assuming that, from cusp to low and middle latitudes, Pc_{sw} energy is transferred via an ionospheric current system, as pulsations are observed in the auroral luminosity at high latitudes (cusp-cleft) (Chernous et al. 1986, Engebretson et al. 1990). This fact indicates that the conductivity of the ionosphere is modulated by Pc pulsations. Engebretson et al. (1991) and Anderson and Engebretson (1995) developed and extended the model of "high latitude entry" using the mechanism developed by Southwood and Kivelson (1991) and Kivelson and Southwood (1991).

Figure 5 shows the distribution of the intensity along the meridian for Pc_{mg}, a distribution which is sharply different from that for Pc_{sw}. The Pc_{mg} maximum is located at latitudes lower than 65°, just in the region which corresponds to the projection of those L-shells on which the Alfven waves' resonance amplification of geomagnetic pulsations exists for the frequency band of the considered Pc pulsations.

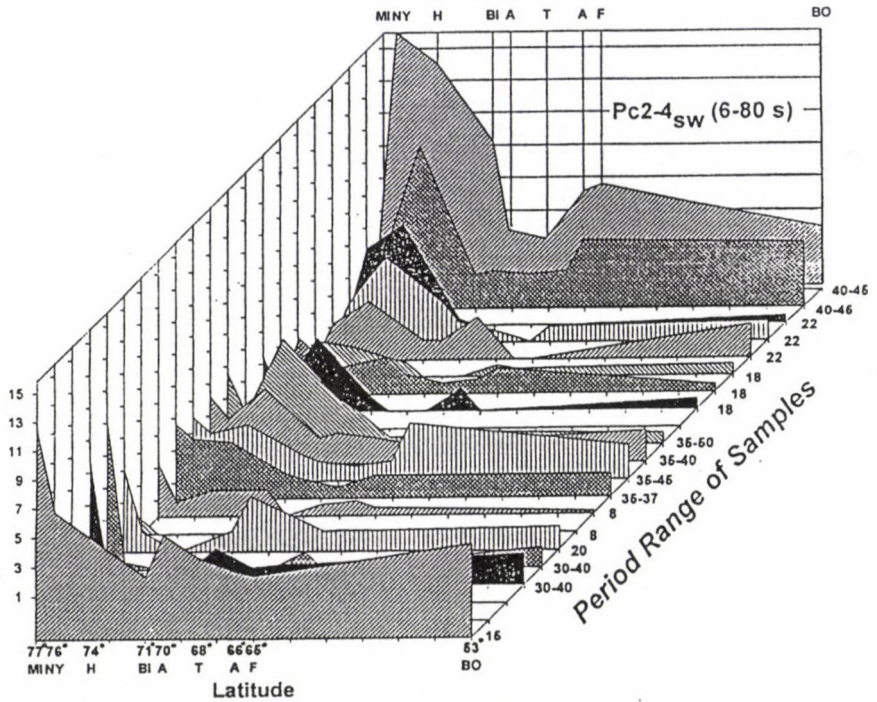


Fig. 4. Dependence of amplitude A_{max} of the solar wind-controlled Pc2-4_{sw} pulsations on latitude. The X-axis is latitude, the latitudes of the 9 Russian and German stations being indicated. These stations operated during August-October 1978. The Y-axis is the maximum amplitude of pulsations A_{max} in nT. Each curve illustrates separate events recorded at the network of stations. The period ranges of individual samples are shown (Plyasova-Bakounina et al. 1986)

Analysis of the data from this network of stations shows that the two types of pulsations, Pc_{sw} and Pc_{mg}, have different patterns of diurnal variation (see Fig. 6). For Pc_{sw} the maximum corresponds to the local pre-noon hours, while for Pc_{mg} the maximum corresponds to local noon (Plyasova-Bakounina 1992). The pre-noon maximum of Pc_{sw} pulsations seems to be related to the character of the source of the Pc_{sw} pulsations. As the maximum growth rate of excited $F-B$ dependent upstream waves corresponds to the orientation of IMF to the morning side (Kovner et al. 1976a), it can lead to the appearance of Pc_{sw} in the morning hours. The pre-noon maximum of Pc_{sw} receives support also from the experimental results presented below (Fig. 8). The noon maximum of Pc_{mg} pulsations supports the noon entry of the energy of compressional waves into the magnetosphere. Such "direct entry" of compressional waves through the subsolar point of the magnetosphere is discussed by McKenzie (1970), Verzariu (1973), Yumoto and Saito (1983) and Yumoto et al. (1985).

The influence of the sign of the IMF sector upon the appearance of Pc_{sw} pulsations on the ground is illustrated in Fig. 7, which is derived from analysis of Pc2-4 events as recorded at the Borok station, Russia (Plyasova-Bakounina 1992). It is

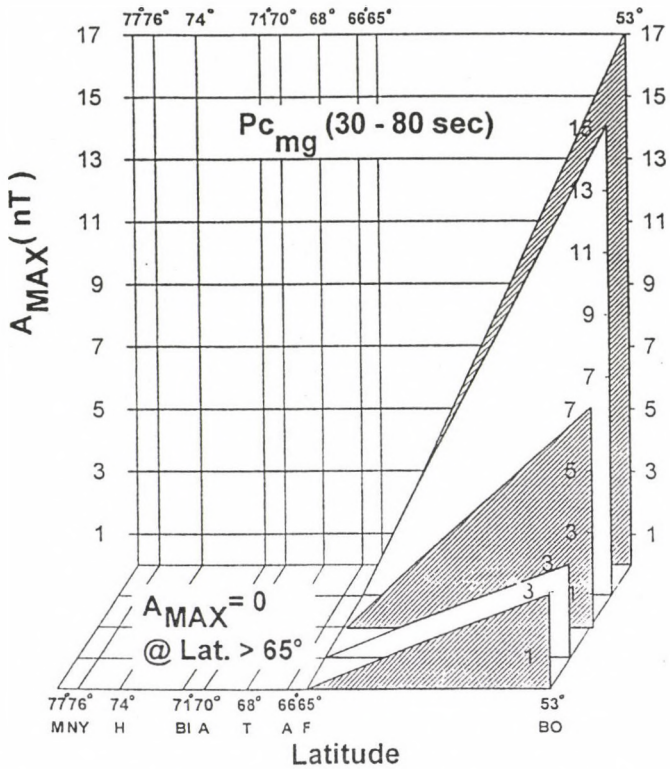


Fig. 5. Dependence of amplitude A_{max} of the magnetosphere-controlled Pc2-4_{mg} pulsations on latitude. The X-axis is latitude with 9 Russian and German stations are marked, which operated during August-October 1978. The Y-axis is the amplitude in nT. Each curve illustrates separate pulsation events recorded at the network of stations (Plyasova-Bakounina et al. 1986)

shown that the correlation $T(B)$ is strong in (+) sector of IMF and is approximated by function $T = 160/B$ (solid line in A) in contrast to the (−) sector of IMF, where the correlation is considerably less. The close correlation between T and B in (+) sector indicates that pulsations are composed mainly of Pc_{sw}. Pulsations in (−) sector are composed mainly of Pc_{mg} as there no functional correlation is observed. We interpret this result that in (+) sector there exists the condition for penetration of upstream waves into magnetosphere, due to reconnection of the Bx-component of IMF with geomagnetic field lines (model of Kwok and Lee 1984).

For (+) sector of IMF the frequency of occurrence of Pc_{sw} and Pc_{mg} versus the orientation of IMF is demonstrated in Fig. 8, which is derived from analysis of Pc 2-4 events as recorded at the Borok station (Plyasova-Bakounina 1992). The preferred IMF orientation for Pc_{sw} appearance is 140° (field lines being inclined to the morning side), while for Pc_{mg} the preferred IMF orientation is 180° (field lines in the direction towards the sun). The difference between the position of two maxima of the two distributions is real, which was verified by Fisher's criterion. Also, it can be seen that both types of pulsations are strongly attenuated for $\varphi = 90°$. This

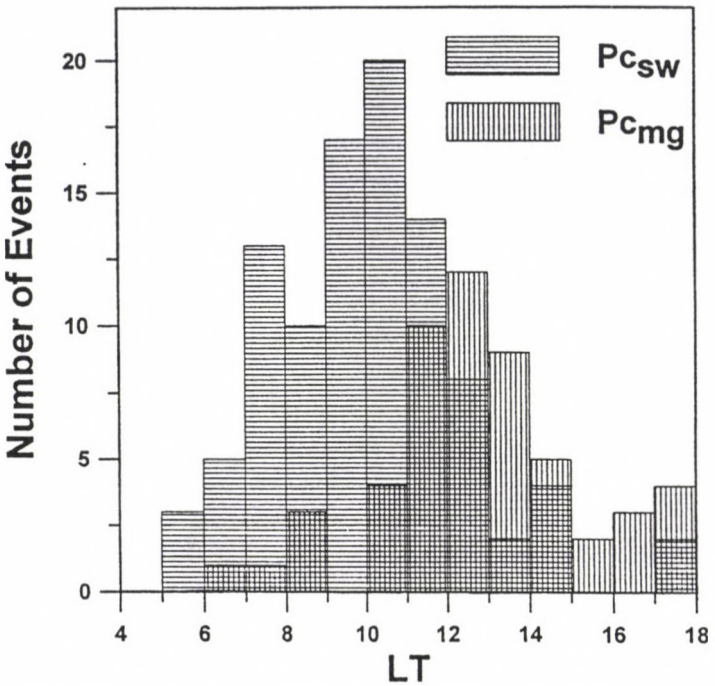


Fig. 6. The diurnal variation of the frequency of occurrence of two types of pulsations Pc_{sw} and Pc_{mg} recorded at Borok station during August-October 1978 (Plyasova-Bakounina 1992)

dependence is not valid for the (–) sector of IMF. The 140° orientation for Pc_{sw} is supported by the theoretical conclusion of Kovner et al. (1976a), where it was demonstrated that the maximum growth rate of upstream waves corresponds to a 150° orientation of IMF.

III. Source of Pc_{sw} pulsations

The direct comparision of the wave activity in the solar wind before the bow shock and on the ground is the decisive element in the solution of the problem of the source of the Pc_{sw} pulsations. Choosing only monochromatic upstream waves in the solar wind (Explorer-34, Heos-1) and simultaneously observed pulsations on the ground (Borok, Russia) (Plyasova-Bakounina 1992), we compare their periods which display marked coincidence (Fig. 9).

The dynamic spectra of monochromatic wave events simultaneously recorded by the Heos satellite near the bow shock in the solar wind, and on the ground at Borok, show that:

a) isolated wave events are identified very well;

b) their periods are the same at both sites;

c) given the abrupt onset of the wave events at the satellite, it is possible to define the time delay (τ) between the onset of pulsations at Heos and at Borok.

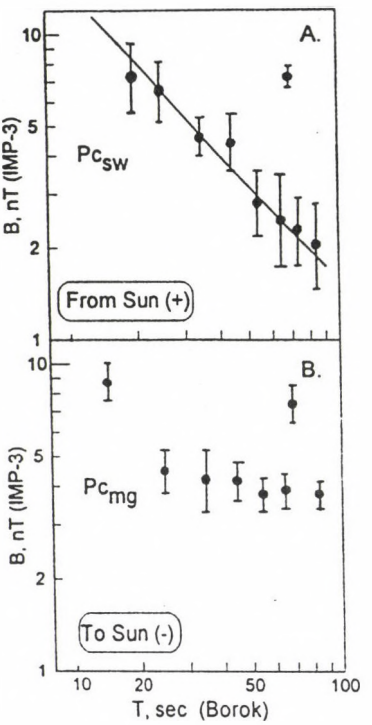


Fig. 7. A plot of the relation between the Pc2-4 pulsation period at Borok (T) and the magnitude of IMF (B) measured by the IMP-3,4 satellites during periods of sunward (A) and anti-sunward (B) directed field. The IMF data were collected by satellites between September-October 1965 and September-December 1966. The number of data points used in sections A and B are 128 and 109, respectively. The solid points are average values of B in 10 second windows of period T, error bars for one standard deviation being shown. The solid line in section A is the function $T = 160/B$, the least squares fit to the 128 data points. Pc_{sw} and Pc_{mg} are explained in the text

Figure 10 shows the time delay between the onset of pulsations at Heos and at Borok, plotted as a function of the upstream distance (D) between the satellite and the bow shock (Plyasova-Bakounina et al. 1978). One can see from the figure that the time delay is greater when the satellite is closer to the bow shock. We can interpret $t2$ as the time of propagation of the waves from the bow shock to the ground, and $t1$ as the time for the beam of reflected protons to reach the satellite, as upstream waves will not appear at a satellite until the protons reach that position. An estimate of the energy of reflected protons using time-delay gives the values 1.6–2.8 kev, close to satellite measurements.

In some cases, there is no pulsation activity observed on the ground during monochromatic pulsation activity in the solar wind. This situation is illustrated on Fig. 11. Figure 11 is a scatter-plot of the 41 monochromatic pulsation events of Heos-1, showing the solar wind velocity V_{sw} as a function of the sunward component of the Alfven velocity in the solar wind C_{A1x} (Golikov et al. 1980). The symbols indicate the presence or absence of Pc_{sw} pulsation activity at Borok during

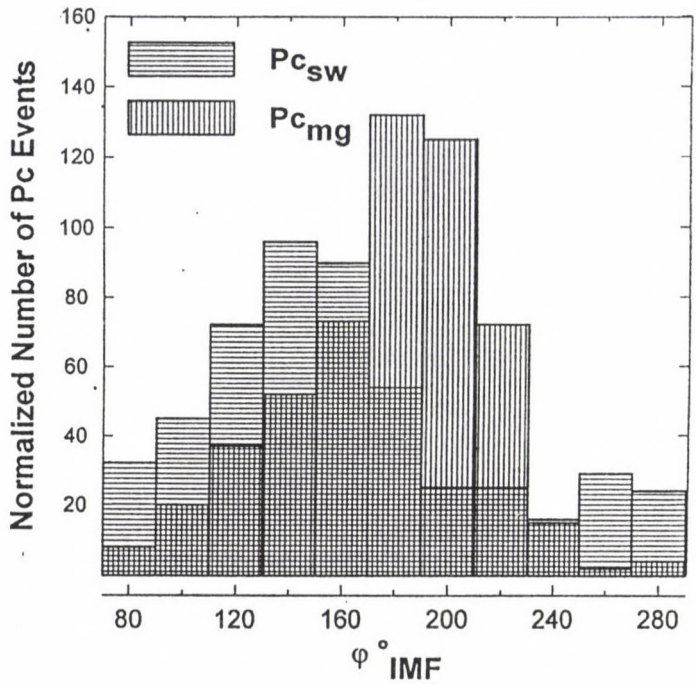


Fig. 8. The histogram distribution of the number of Pc events (Pc_{SW} and Pc_{mg}) versus the simultaneous orientation of IMF (azimuthal angle φ) normalized against the probability of occurrence of φ. Pc pulsations measured at Borok, φ angle data from IMP-3,4 satellites collected between September-October 1965 and September-December 1966. The dependence is shown for anti-sunward direction of IMF ((+) sector)

monochromatic pulsation activity at Heos-1. One can see that the coincidence between the signal in the solar wind and on the ground exists only when V_{sw} is greater than some critical velocity V_{cr}, where $V_{cr} = C_{A1x}$ + constant. We interpret the condition of Pc_{sw} presence on the ground as a condition for the development of a K-H instability on the boundary of the magnetosphere ($V_{sw} > V_{cr}$). We conjecture that the surface waves, which are developed on the flanks of the magnetopause as a result of this instability (Kivelson and Pu 1984), are amplified by a non-linear mechanism of the exchange of energy with the monochromatic transverse upstream waves which are propagating with the stream lines flowing from the nose of the magnetosphere to high latitudes. As a result of amplification of surface waves in the frequency band of monochromatic upstream waves, we observe them on the ground into the cusp projection region as Pc_{sw}.

IV. Pc_{sw} inside the magnetosphere

The important element in understanding penetration of upstream waves to the ground is observation of Pc_{sw} inside the magnetosphere. Comparison of data from the geostationary satellite ATS-1 with ground data simultaneously recorded at

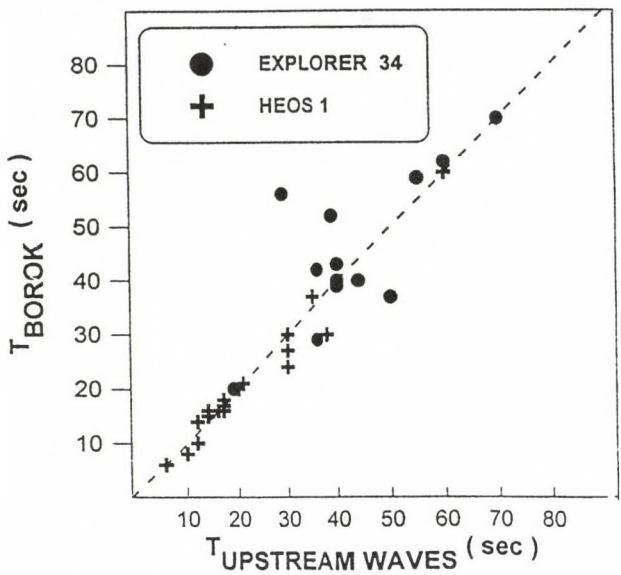


Fig. 9. Periods of simultaneously observed ground pulsations (Borok) and upstream monochromatic waves measured by Explorer-34 between September-November 1967 and measured by Heos-1 between December 1968 – April 1969 (Plyasova-Bakounina 1992)

Petropavlovsk-Kamchatsky, Russia, indicates that two groups of events are present (Kovner et al. 1976b, Plyasova-Bakounina et al. 1982):.

1. one for which the periods of events at both sites are identical (5% of all daytime events), and

2. the other for which periods do not coincide (95% of all daytime events).

The upper part of Fig. 12 demonstrates the period of Pc events versus the magnitude of IMF B for the first group, the lower part of the figure giving similar data for the second group. It is seen that for the first group there is a correlation with B, indicating that these are Pc_{sw} events. The second group shows no correlation of the events with B, indicating the group to be composed of Pc_{mg} events.

Arthur and McPherron (1977), Greenstadt and Olson (1977), Takahashi et al. (1984), Yumoto et al. (1985), Odera (1986), Engebretson et al. (1987), and Anderson et al. (1991) have shown that the amplitude of Pc3-4 pulsations within the magnetosphere strongly depends upon orientation of the IMF azimuthal angle, φ. As the majority of events observed within the magnetosphere are Pc_{mg}, it means that they are controlled by orientation of IMF but their spectra are not controlled by B.

In disagreement with Yumoto et al. (1984), Tahakashi et al. (1984) did not find any correlation of the frequency of pulsations within the magnetosphere with magnitude B, which verifies our result (in a forthcoming paper, we will suggest an interpretation of Yumoto's data consistent with our and Takahashi's results). If we

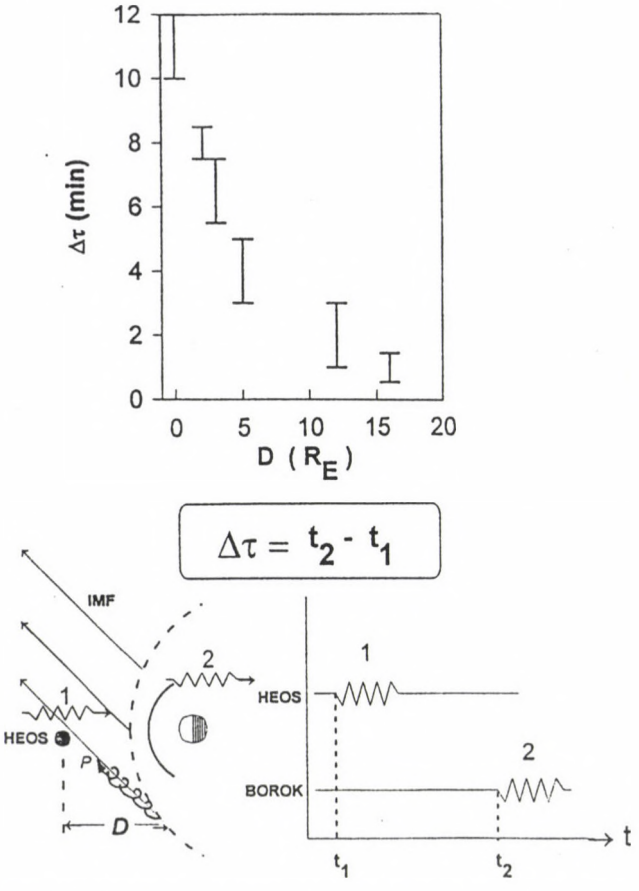


Fig. 10. Time delay ($\Delta\tau$) between the onset of pulsation activity at Heos and at Borok plotted as function of the upstream distance between the satellite and the bow shock (D)

combine Figs 12 (A) and (B) no correlation of T with B will be observed. Engebretson et al. (1991) demonstrated simultaneous exitation of similar frequency bands both within the magnetosphere and on the ground. Addition, they demonstrated that the calculated frequency of generated monochromatic upstream wave, using the formula given by Troitskaya et al. (1971) and Gul'elmi (1974), occurs in the observed frequency band of ground observations. This is as expected because as shown above, the frequency bands of excitation of Pc_{mg} and Pc_{sw} are the same. This characteristic feature is the main difficulty in separation of the two types of pulsations on the ground, when we try to solve the problem of ground-based monitoring of IMF.

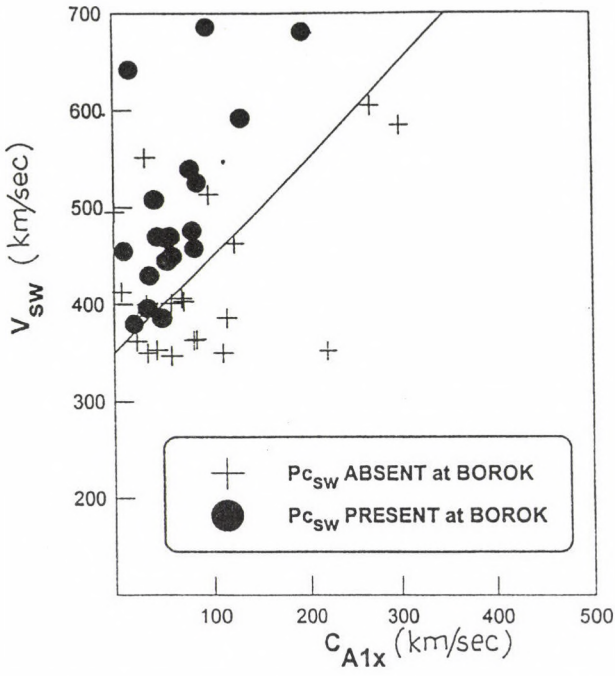


Fig. 11. Scatter plot of the 41 pulsation events recorded at Heos-1, showing the solar wind velocity V_{sw} as a function of the sunward component of the Alfven velocity in the solar wind, C_{A1x}. The symbols indicate the presence or absence of Pc_{sw} pulsation activity at Borok during pulsation activity at Heos-1

V. Ground-based monitoring of IMF using micropulsation data

The relationship between T and B raises the possibility of measuring T on the ground as a basis for defining B in the solar wind. In an attempt to exploit this idea, Gul'elmi and Bolshakova (1973) measured the average period of all ground-recorded pulsation activity, using this period to calculate B (B-index). They achieved very poor correspondence between calculated and observed B values for the solar wind (Russell and Fleming 1976). Their lack of success derived from their failure to limit their measurements to use of Pc_{sw}, which are the only pulsations containing information about B. How to select Pc_{sw} from all other ground-recorded pulsations of the frequency band of Pc2-4 has been described by Plyasova-Bakounina et al. (1982). In this paper, the fact has been established that a characteristic feature of Pc_{sw} pulsations is their global distribution, thus giving general world-wide utility to this proposed procedure.

Figure 13 shows the experimental results of a comparison of the pulsations recorded at two stations located at similar latitudes but 32° different in longitude, Hartland (England) and Borok (Russia) (Plyasova-Bakounina et al. 1982). The X axis shows the period, calculated by the formula $T = 160/B$, where B is the magnitude of IMF. The Y axis shows the period of pulsations actually observed

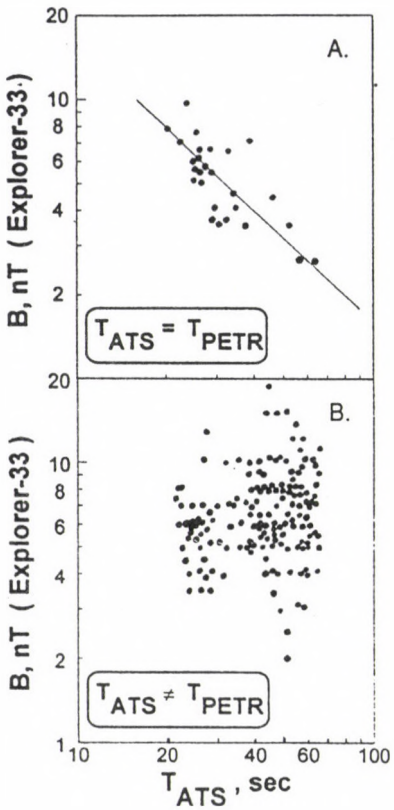


Fig. 12. Period of Pc events (10-100 sec) observed within the magnetosphere at the geostationary satellite ATS-1 (T_{ATS}) versus the magnitude of IMF (B) observed by Explorer-33 for two groups of events: (A) when $T_{ATS} = T_{PETR}$; and (B) when $T_{ATS} \neq T_{PETR}$. T_{PETR} is the period of Pc events measured simultaneously at the ground station Petropavlovsk-Kamchatsky located close to the footstep of the field line connecting ATS-1 and the ground surface. The data were collected between January-May 1968. The solid line in section A is the function $T = 160/B$, the least squares fit to the data points. Pc_{sw} and Pc_{mg} are explained in the text

at Hartland and Borok. One can see that each station records a broad frequency band which corresponds poorly to the calculated period $T = 160/B$. However, the lowest portion of the figure illustrates that the common part of the frequency bands recorded by both stations simultaneously has very high correlation with $T = 160/B$. This correlation indicates these pulsations are Pc_{sw}. A similar result is illustrated on Fig. 13 for two stations at the same meridian but different latitudes (30° difference in latitude), Lovozero and Borok, both in Russia.

The method of ground-based monitoring of the magnitude of IMF is developed in Plyasova-Bakounina and Münch (1991). The calculated coherence functions (CF) of the signals recorded at Lindau, Germany, and Sverdlovsk, Russia (common latitude, longitudes different by 38°) have their maxima for both components of the magnetic field H and D at the frequency (F_C) of the signal from the solar wind. Table II

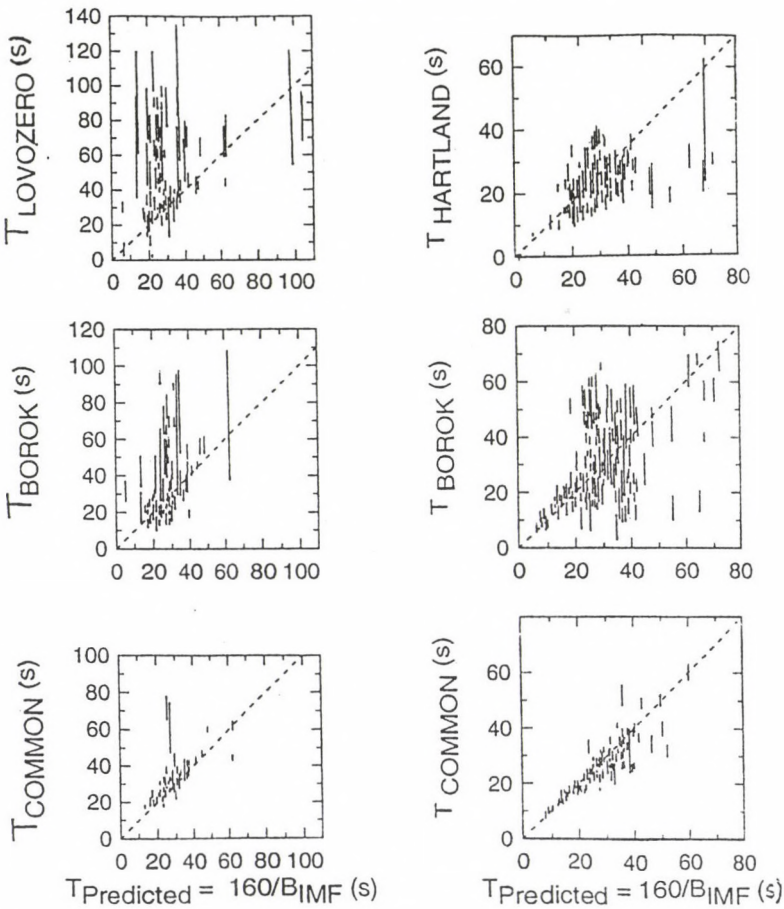


Fig. 13. Period range of pulsations simultaneously recorded by two pairs of stations (Borok - Lovozero, and Borok - Hartland) versus the predicted period of upstream waves in the solar wind. Data for each station as well as data for periods common to the data of the two stations are shown. $T_{\mathrm{PRED}} = 160/B$. See text for further discussion

demonstrates that the values of B calculated according to this frequency F_C coincide with the real values of magnitude as determined from satellite measurements.

Discussion and conclusion

From the set of described results, the following broad concept of extra-magnetospheric origin of the Pc pulsations is suggested and illustrated by Fig. 14.

Broadly speaking, there are two principal types of upstream waves near the bow shock (Russell and Hoppe 1983, Yumoto 1988). One type is mainly transverse monochromatic waves driven by reflected protons with an "intermediate" function of velocity distribution with frequencies obeying the relation $\omega = 0.4\omega B$ (Russell

Table II. This demonstrates agreement between the magnitude of IMF B received from satellites (B_{real}) and magnitude of IMF calculated using ground-based method (B_{calc}). ΔF_H, ΔF_D – frequencies correspond to the maximum of the Coherence Function calculated betwen Pc2–4 pulsations recorded at Lindau and Sverdlovsk, H and D referring to these components of the magnetic field of the pulsations. ΔF_C – common part of frequency band observed at both stations simultaneously.

$$\Delta B_{calc} = 6.25 \Delta F_C$$

No.	Date d.mo.yr.	Time UT	ΔF_H (mHz)	ΔF_D (mHz)	ΔF_C (mHz)	ΔB_{calc}	B_{real}
1.0	5.09.74	06.00–06.30	40–37	37	37	5.9	5.9 ± 4
2.0		06.30–07.00	38–33	36	36	5.8	5.9 ± 4
3.0		07.00–07.30	34	100			–
4.0		07.30–08.00	71	55			–
5.0	6.09.74	06.00–06.30	30	43			–
6.0		06.30–07.00	27.8	28–26	28	4.5	4.4 ± 3
7.0		07.00–07.30	28–25	37–28	28	4.5	4.5 ± 3
8.0		07.30–08.00	25–13	31–24	25	4.0	4.5 ± 3
9.0	7.09.74	06.00–06.30	25	20–14			–
10.0		06.30–07.00	72	59			–
11.0		07.00–07.30	20	25;14			–
12.0		07.30–08.00	20	20;14	20	3.2	3.6 ± 3
13.0		08.00–08.30	40–31	33	33	5.3	5.3 ± 4
14.0		08.30–09.00	25–20	29			–
15.0	8.09.74	06.00–06.30	34–27	27	27	4.3	4.3 ± 3
16.0		06.30–07.00	32–27	27	27	4.3	4.3 ± 3
17.0		07.00–07.30	29	46–29	29	4.7	4.5 ± 3
18.0		07.30–08.00	27–26	27	27	4.3	4.5 ± 3
19.0		08.00–08.30	27	14			–
20.0		08.30–09.00	22–18	25			–

and Hoppe 1981). They are responsible for the creation of Pc_{sw} pulsations as demonstrated in Plyasova-Bakounina et al. (1978), Greenstadt and Olson (1979), Wolf et al. (1980), Russell and Hoppe (1981), Verő (1986) and Plyasova-Bakounina (1992).

The second type — highly-compressional waves of irregular complex form (multichromatic waves) — are driven by "diffuse" protons. Such protons occur as a result of scattering of the beam of energetic reflected protons. As pulsations inside the magnetosphere a) display the dependence of the amplitude of such pulsations on the orientation of IMF and b) do not display $F - B$ dependence, the character of the source of such pulsations is required to have the following characteristics:

1. their source is located in the solar wind (as their dependence on IMF orientation requires)
2. their source has no $F - B$ dependence
3. their source is multichromatic (broadband) in nature because inside the magnetosphere the resonant pulsations are driven at several L-shells simultaneously.

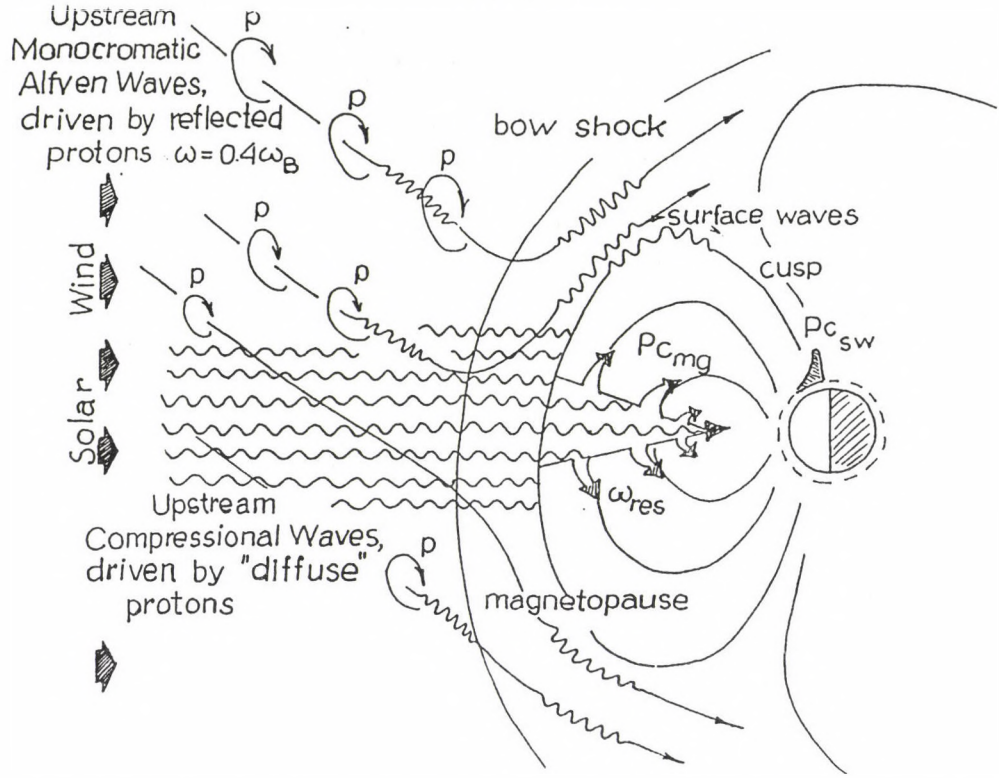


Fig. 14. Diagram illustrating how the two types of upstream waves (Alfven and compressional waves) in the solar wind are sources of the two types of pulsations observed at the ground (Pc_{sw} and Pc_{mg})

We believe that a source with the characteristics described above is the highly compressional upstream waves of irregular wave form which are generated by "diffuse" protons.

Given a source within the solar wind, we require an explanation of how the wave energy transfers from the solar wind into the magnetosphere. This issue, which is huge and complicated, cannot be detailed here but some of the mechanisms supporting our main hypothesis will be outlined.

The compressional waves responsible for Pc_{mg} can penetrate into magnetosphere due to provocation of the motions of the boundary of the magnetosphere. The penetration of normally incident compressional waves into the subsolar magnetopause (tangential discontinuity) has been considered both theoretically (Mc Kenzie 1970, Verzariu 1973, Yumoto et al. 1984, Yumoto 1988, Southwood and Kivelson 1990) and experimentally (Wolf and Kaufman 1975, Greenstadt et al. 1983, Engebretson et al. 1991, Song et al. 1993). In contrast with Verzariu (1973) and Greenstadt et al. (1983), who reported a transition coefficient of incident power of compressional waves 1–2%, Song et al. (1993) obtained an observed coefficient of 18%. The cou-

pling of the compressional waves with the local resonant oscillations are considered by Chen and Hasegawa (1974), Southwood (1968, 1974), Southwood and Kivelson (1990), Fedorov (1992).

Considering the magnetopause as a rotational discontinuity and under the assumption that the reconnection of field lines takes place, Kwok and Lee (1984) considered the propagation of ULF waves of all polarisations across the magnetopause. They demonstrate that in this case (in contrast with Verzariu's model) the wave power even can be amplified.

In contrast with compressional waves, transverse waves cannot directly penetrate across the sub-solar magnetopause; being guided by magnetic field lines, they move with the stream lines of the solar wind, flowing over the magnetopause to the high latitudes. A possible mechanism of penetration of the energy of the monochromatic transverse waves is exchange of energy with surface waves on the magnetopause, those surface waves being driven by K-H instability on the flanks of the magnetosphere in the high-latitude regions (Kivelson and Pu 1984). We believe that the corresponding frequencies in the wide spectrum of surface waves, amplified by interaction with monochromatic upstream waves ($\omega = 0.4\omega B$) by some as yet unknown mechanism, can be observed at the ground surface as Pc_{sw}. This our interpretation is confirmed by the location of the maximum of Pc_{sw} intensity in the cusp region, to where the magnetopause is projected. The role of the K-H instability in generation of Pc3-4 pulsations is noted also by Singer et al. (1977), Greenstadt et al. (1979) and Verő (1986). Another possibility for penetration of the energy of transverse Alfven waves can be realized under the condition of reconnection (Kwok and Lee 1984). The fact that Pc_{sw} pulsations are recorded on the ground when the radial component of IMF (B_X) is directed away from the sun implies that the reconnection of B_X with magnetospheric field lines takes place in the point which is projected into the cusp/cleft region. The transverse waves are guided along cusp field lines to the ground surface. Southwood and Kivelson (1991) and Kivelson and Southwood (1991) developed the mechanism of the non-linear transport into the cusp latitudes of the magnetosheath pressure impulses impinging on the magnetopause. We believe that this model qualitatively fits better for explanation of Pi1 and Pi2 pulsations connected with substorms than for monochromatic Pc_{sw} pulsations. Finally, we note that Russell and Elphic (1978) proposed a much different model of penetration of the energy of transverse waves by creating on the magnetopause mini-flux transfer events (FTEs). Though their mechanism may quite possibly be correct for upstream wave penetration into the magnetosphere, presently available data cannot determine the relationships between "FTE events" and "Pc_{sw} pulsations".

Table III compares the characteristics of monochromatic transverse upstream waves in the solar wind ($\omega = 0.4\omega B$) and Pc_{sw} on the ground, while Table IV compares the characteristics of the multi-chromatic compressional upstream waves in the solar wind and Pc_{mg} pulsations.

It is seen that characteristics of the ground pulsations of Pc_{sw} and Pc_{mg} and of the satellite-observed MHD-upstream waves are very similar.

We propose that the MHD transverse monochromatic waves are the source of Pc_{sw} for the following reasons: first, because their period T obeys the some rela-

Table III. Characteristics of upstream monochromatic Alfven waves and the solar wind-controlled Pc_{sw} pulsations on the ground

Upstream monochromatic Alfven waves	Pc_{sw}
T fixed	T fixed
$T = 160/B$	$T = 160/B$
Cannot propagate directly across magnetopause.	Pc_{sw} nearly absent inside magnetosphere.
Propagate through day cusp.	"Polar" maximum of intensity.
Maximum of appearance at $\varphi = 150°$	Maximum of appearance at $\varphi = 150°$. Pre-noon maximum of diurnal appearance.
At $\varphi = 90°$, no Alfven waves	At $\varphi = 90°$, no Pc_{sw} activity at the ground.

Table IV. Characteristics of upstream highly-compressional waves of the complex irregular wave forms and the magnetosphere-controlled Pc_{mg} pulsations

Upstream compressional waves	Pc_{mg}
T not fixed	T not fixed
$T \neq 160/B$	$T \neq 160/B$
Can penetrate directly across subsolar magnetopause.	Pc_{mg} present in magnetosphere. Noon maximum of occurrence. Maximum occurrence at $\varphi = 180°$
Irregular wave form. Broad bands waves	several L-shells are excited simultaneously
Transformed into Alfven resonance oscillations	Resonance characteristics. Maximum of distribution of intensity at $\varphi < 65°$.
At $\varphi = 90°$, no compressional waves	At $\varphi = 90°$, no Pc_{mg} activity either in magnetosphere or at the ground.

tionship ($T = 160/B$) as does the period of Pc_{sw} observed on the ground; second, because their occurrence probability is controlled by the IMF azimuthal angle φ in the same manner as are ground-observed pulsations Pc_{sw}. The maximum occurrence of pulsations both in the solar wind and on the ground is at φ values of $140° - 150°$. For monochromatic upstream waves, this position of the maximum was demonstrated theoretically by Kovner et al. (1976a) — for Pc_{sw}, this relationship is illustrated by Fig. 6 of this paper; third, the morning diurnal maximum of Pc_{sw} can be explained by the orientation of IMF. The generation of MHD monochromatic waves requires an inclination of IMF field lines into the morning sector, which results in the appearance of Pc_{sw} in the morning hours; fourth, as the majority of monochromatic waves are of transverse type, they cannot penetrate directly through the tangential discontinuity, i.e., the magnetopause. They propagate with the stream lines flowing up the magnetosphere. Penetration of their energy into the magnetosphere and to the ground surface can be expected at the flanks

of the magnetopause at high latitudes where the reconnection processes play an important role and where the surface waves can be developed. Pc_{sw} display their maximum of intensity in the region of the cusp and Pc_{sw} appears on the ground when — (a) the condition of reconnection exists, the B_X component directs toward the sun, and — (b) the condition of the development of the surface waves on the boundary of the magnetosphere is fulfilled. As the reconnection inevitably is accompanied by surface wave generation (Semenov et al. 1994), the appearance of Pc_{sw} in connection with these two phenomena is probably not accidental.

We propose the broad-band compressional upstream waves as the source of Pc_{mg} for the following reasons: first, because these highly-compressional waves can directly propagate (even if with large attenuation) through the boundary of the magnetosphere, coupling into resonance oscillations of the field lines; second, as they are irregular wave form multichromatic signals, these compressional waves can drive Alfven resonances in several L-shells throughout the magnetosphere; third, as they are multichromatic waves, they display an absence of F-B dependence; fourth, since they can penetrate across the subsolar point of the magnetopause, the near-noon maximum of occurrence of Pc_{mg} is easily explained; fifth, being connected with the appearance of the beam of reflected protons which are directly dependent on the orientation of IMF, the compressional waves may well be controlled by the orientation of IMF, thus explaining the observed strong control of Pc_{mg} pulsations by the orientation of IMF.

The suggested model of formation of Pc is distinguished from earlier concepts of an extra-magnetospheric origin in that it considers the entire class of Pc (Pc_{sw} + Pc_{mg}) as a class of pulsations having a common energetic source — MHD-upstream waves in the solar wind. Futher observations and analysis may require modification of our model.

References

Anderson B J, Engebretson M J 1995: *J. Geophys. Res.*, 100, 9591.

Anderson B J, Potemra T A, Zanetti L Z, Engebretson M J 1991: In: Physics of Earth Plasmas (1990) SPI Conference Proceedings and Reprint Series, 10, 419, Scientific Publishers Inc., Cambridge, Mass.

Arthur C W, McPherron R L 1977: *J. Geophys. Res.*, 82, 5738.

Bolshakova O V, Troitskaya V A 1968: *Dokl. Akad. Nauk USSR*, 180, 343.

Bolshakova O V, Troitskaya V A 1984: *Geomagn. Aeron.*, 24, 633.

Buldurev V S, Kovtun A A, Fam Van Chi 1973: *Geomagn. Aeron.*, 13, 136.

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

Chernous S A, Vorobjev V A, Tagirov V R, Rotwell R R, Crivs C, Lanchest B 1986: Proceeding of International Symposium "Polar Geomagnetic Phenomena", Souzdal, USSR, 23.

Dungey J W 1954: Penn. State Sci. Rep. No. 69

Dungey J W 1962: In: Geophysics, De Witt C et al. ed., Gordon and Breach, New York, 503.

Engebretson M J, Zanetti L J, Potemra T A, Baumjohaun W, Luehr H, Acuna M H 1987: *J. Geophys. Res.*, 92, 10053.

Engebretson M J, Anderson B J, Cahill L J, Arnoldy R L Jr, Newell P T, Meng C I, Zanetti L J, Potemra T A 1989: *J. Geophys. Res.*, 94, 17143.

Engebretson M J, Anderson B J, Cahill L , Arnoldy R L Jr, Rosenberg T J, Carpenter D L, Gail W B, Eather R H 1990: *J. Geophys. Res.*, 95, 2447.

Engebretson M J, Cahill L J, Arnoldy R L Jr, Anderson, B J, Rosenberg T J, Carpenter D L, Inan U S, Eather R H 1991: *J. Geophys. Res.*, 96, A2, 1527.

Fairfield D H 1969: *J. Geophys. Res.*, 74, 14, 3541.

Fedorov E N 1992: Doctoral Thesis, Inst. Phys. Earth, Moscow (in Russian)

Golikov Yu V, Plyasova-Bakounina T A, Troitskaya V A, Chernikov A A, Pustovalov V V, Hedgecock P C 1980: *Planet. Space Sci.*, 28, 535.

Greenstadt E W 1973: Proceeding of the Conference on Solar Terrestrial Relations, D V Venkatosan ed., Univ. Calgary, Alta.

Greenstadt E W, Olson J V 1977: *J. Geophys. Res.*, 82, 4991.

Greenstadt E W, Olson J V 1979: *J. Geophys. Res.*, 84, 1493.

Greenstadt E W, Singer H J, Russell C T, Olson J V 1979: *J. Geophys. Res.*, 84, 527.

Greenstadt E W, Mellott M M, McPherron R L, Russell C T, Singer H J, Knecht D J 1983: *Geophys. Res. Lett.*, 10, 659.

Gul'elmi A V 1974: *Space Sci. Rev.*, 16, 331.

Gul'elmi A V, Bolshakova O V 1973: *Geomagn. Aeron.*, 13, 459.

Gul'elmi A V, Plyasova-Bakounina T A, Chepetnov R V 1973: *Geomagn. Aeron.*, 13, 331.

Hasegawa A, Chen L 1974: *Space Sci. Rev.*, 16, 347.

Kivelson M G, Pu Z Y 1984: *Planet. Sp. Sci.*, 32, 1335.

Kivelson M G, Southwood D J 1986: *J. Geophys. Res.*, 91, 4345.

Kivelson M G, Southwood D J, 1991: *Geophys. Res.*, 96, A2, 1661.

Kovner M S, Lebedev V V, Plyasova-Bakounina T A, Troitskaya V A 1976a: *Planet. Sp. Sci.*, 24, 267.

Kovner M S, Kuznetzova V A, Lebedev V V, Plyasova-Bakounina T A, Troitskaya V A 1976b: *Annales Geophys.*, 32, 3, 189.

Kwok Y C, Lee L C 1984: *J. Geophys. Res.*, 89, 10697.

Lanzerotti L J, Lie H P, Tartaglia N A 1972: *Science*, 178, 499.

Maltsev Yu P 1971: *Geomagn. Aeron.*, 11, 925.

McKenzie J F 1970: *Planet. Space Sci.*, 18, 1.

Miletits J, Verő J, Stuart W 1988: *J. Atm. Terr. Phys.*, 50, 6449.

Morris R J, Cole K D 1987: *Planet. Space Sci.*, 35, 1437.

Odera T J 1986: *Rev. Geophys.*, 24, 1, 55.

Ong R S, Rederick N 1972: *Planet. Sp. Sci.*, 20, 1.

Plyasova-Bakounina T A 1972a: *Geomagn. Aeron.*, 11, 4.

Plyasova-Bakounina T A 1972b: Candidat Degree Thesis, Inst. Phys. Earth, Moscow (in Russian)

Plyasova-Bakounina T A 1992: Doctoral Thesis, Inst. Phys. Earth, Moscow (in Russian)

Plyasova-Bakounina T A, Golikov Yu V, Troitskaya V A, Hedgecock P C 1978: *Planet. Sp. Sci.*, 26, 547.

Plyasova-Bakounina T A, Stuart W F, Charchenko I P, Troitskaya V A 1982: Issledov. po Geom. Aeron. i physiki Solntza, Irkutsk (in Russian) , 62, 103

Plyasova-Bakounina T A, Münch J W, Gauler H, Troitskaya V A 1984: In: Program and Theses of "All-Union Seminar, on Geomagnetic Pulsations, Irkutsk, June 1984", Proc. SibIZMIR of Acad. Sci. of USSR, Irkutsk (in Russian)

Plyasova-Bakounina T A, Münch J W, Gauler H, Troitskaya V A 1986: *Acta Geod. Geoph. Mont. Hung.*, 21, 143.

Plyasova-Bakounina T A, Münch J W 1991: *Acta Geod. Geoph. Mont. Hung.*, 26, 263.
Pochotelov O A, Pelipenko V A 1976: *Geomagn. Aeron.*, 16, 296.
Russell C T, Elphic R C 1978: *Space Sci. Rev.*, 22, 681.
Russell C T, Fleming B K 1976: *J. Geophys. Res.*, 81, 5882.
Russell C T, Hoppe M M 1981: *Geophys. Res., Lett.*, 8, 615.
Russell C T, Hoppe M M 1983: *Space Sci. Rev.*, 34, 155.
Saito T 1964: *Rept. Ion. Space Res.*, 18, 260.
Semenov V S, Farrugia C J, Biernat H K, Lebedeva V V, Rijnbeek R P 1994: *Geophys. Res. Lett.*, 21, 22, 243.
Singer H J, Russell C T, Kivelson M G, Greenstadt E W, Olson J V 1977: *Geophys. Res. Lett.*, 4, 337.
Song P, Russell C T, Strangeway R J, Wygant J R, Cattell C A, Fitzenreiter R J, Anderson R R 1993: *J. Geophys. Res.*, 98, A1, 187.
Southwood D J 1968: *Planet. Sp. Sci.*, 16, 587.
Southwood D J 1974: *Planet. Sp. Sci.*, 22, 483.
Southwood D J, Kivelson M G 1990: *J. Geophys. Res.*, 95, 2301.
Southwood D J, Kivelson M G 1991: *J. Geophys. Res.*, 96, A1, 67.
Takahashi K, McPherron R L, Greenstadt E W, Neeley C A 1981: *J. Geophys. Res.*, 86, A7, 5472.
Takahashi K, McPherron R L, Terasawa T 1984: *J. Geophys. Res.*, 89, A5, 2770.
Troitskaya V A, Plyasova-Bakounina T A, Gul'elmi A V 1971: *Dokl. Akad. Nauk USSR*, 197, 1312.
Varga M J 1980: *Atmos. Terr. Phys.*, 42, 365.
Verő J 1979: Proceed. Inter. Workshop Sel. Top; Magn. Phys., 177, Jap. IMS Comm., Tokyo
Verő J 1986: *J. Geophys. Res.*, 60, 106.
Verő J, Zieger B, Lühr H 1994: Solar Wind Sources of Magnetospheric ULF Waves, Geophys. Monograph 81, AGU
Verő J, Best I, Vellante M, Lühr H, de Lauretis M, Holló L, Märcz F, Strestik J 1995: *Ann. Geophys.*, 13, 689.
Verzariu P 1973: *Planet. Space Sci.*, 21, 2213.
Vinogradov P A, Parchomov V A 1974: *Geomagn. Aeron.*, 15, 109.
Webb D, Orr D 1976: *J. Geophys. Res.*, 81, 5941.
Wolf A, Kaufman R L 1975: *J. Geophys. Res.*, 80, 1764.
Wolf A, Lanzerotti L J, Maclennan C G 1980: *J. Geophys. Res.*, 85, 114.
Wolf A, Kamen E, Lanzerotti L J, Maclennan C G, Bamber J F, Venkatesan D 1987: *J. Geophys. Res.*, 92, A1, 168.
Yumoto K 1988: *Geomag. Geoelectr.*, 40, 293.
Yumoto K, Saito T 1983: *J. Gephys. Res.*, 88, 10041.
Yumoto K, Saito T, Tsurtani B T, Smith E J, Akasofu S I 1984: *J. Geophys. Res.*, 89, 9731.
Yumoto K, Saito T, Akasofu S I, Tsurtani B T, Smith E J 1985: *J. Geophys. Res.*, 90, 6439.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 193–202 (1997)

ARE PULSATIONS WITHIN THE MAGNETOSPHERE AFFECTED BY THE ORIENTATION AND MAGNITUDE OF IMF?

T A PLYASOVA-BAKOUNINA[1]

[Manuscript received April 23, 1997]

The paper discusses observational and theoretical characteristics of geomagnetic pulsation having a source in the solar wind (Pc_{sw}) and in the magnetosphere (Pc_{mg}), respectively. Possible pathways of propagation ere also shown.

Keywords: geomagnetic pulsations; interplanetary magnetic field; solar wind; upstream waves

Introduction

At the present time, the pathways of penetration of the upstream wave energy from the solar wind, as well as their input into pulsation activity within the magnetosphere and at the ground surface, are still unclear. In clarification of this persisting uncertainty, it is very important that we deal only with trustworthy experimental results as such data will provide the bases for selection of the correct model of energy transfer. In this connection, it is critical to determine whether the pulsations recorded on board stationary satellites within the magnetosphere do or do not display the following dependencies: a) is the frequency of occurrence of the pulsations inside the magnetosphere controlled by the orientation of Interplanetary Magnetic Field (IMF) as it is for upstream waves (Russell and Hoppe 1983) and pulsations on the ground (Bolshakova and Troitskaya 1968, Plyasova-Bakounina 1972, 1993, Webb and Orr 1976, Greenstadt and Olson 1977), and b) are the frequencies (F) of pulsations within the magnetosphere controlled by the magnitude B of IMF in the same manner as exists for upstream waves ($F = 6B$) in the solar wind (Fairfield 1969, Plyasova-Bakounina et al. 1978, Russell and Hoppe 1983) and on the ground (Troitskaya et al. 1971, Plyasova-Bakounina 1972, 1993, Gul'elmi et al. 1973, Verő 1986).

Arthur and McPherron (1977) and Takahashi et al. (1981, 1984), when analyzing pulsation data of ATS-6, found clear negative correlation between the angle of the IMF measured from the sun-earth line and the amplitude of pulsations. The frequency of occurrence and the amplitude of both compressional Pc3 and torroidal pulsations Pc3-4, measured at geostationary satellites GEOS-2 and AMPTE, has been associated with low IMF cone angle (Yumoto et al. 1985, Engebretson et al. 1986, Anderson et al. 1991).

[1]Institute of Physics of the Earth, Moscow, Russia

1217-8977/97/$ 5.00

On the ground, the amplitude and frequency of occurrence of the Pc2-4 pulsations increases with decreasing cone angle of IMF, a fact demonstrated by Bolshakova and Troitskaya (1968), Plyasova-Bakounina (1972), Webb and Orr (1976), Greenstadt and Olson (1977), Wolf et al. (1980, 1987), Yumoto et al. (1985) and Engebretson et al. (1986). Analyzing this "angle effect" separately for Pc_{sw} (the solar wind-controlled pulsations, which display $F = 6B$ correlation) and for Pc_{mg} (the magnetosphere-controlled pulsations, i.e. standing resonance oscillations on the local field lines, the frequency of which is controlled by magnetosphere structure and plasma parameters), Plyasova-Bakounina (1993) demonstrated that the frequency of occurrence of Pc_{sw} increases when the azimuthal angle of IMF is close to 140° and close to 180° for Pc_{mg} pulsations. Both pulsations are strongly dissipated when the azimuthal angle of IMF is 90°. Another unambiguous indication of the two types of pulsations was made by Miletits et al. (1988).

Several contradictory results have been published relative to the dependence of the periods of the pulsations (T) observed within the magnetosphere on the IMF magnitude (B). Kovner et al. (1976), Plyasova-Bakounina et al. (1982), Plyasova-Bakounina (1993) investigated the $T(B)$ dependence of monochromatic pulsations in the period band 20–150 sec which were recorded simultaneously by ATS-1 in geosynchronous orbit and on the ground at two ground stations. Analysis showed that that portion of the monochromatic pulsations recorded within the magnetosphere with periods coinciding with the periods of ground-based Pc activity displays a strong $T(B)$ correlation ($T = 160/B$, or $F = 6B$). That portion of the monochromatic pulsations with periods that do not coincide with those of ground pulsation activity did not display any dependence of T on B.

When analizing data of ATS-6, Arthur and McPherron (1977) found no pulsation activity which displayed an $F(B)$ correlation. Takahashi et al. (1984) demonstrated that the high-harmonic standing waves within the magnetosphere display weak negative correlation of frequency F on B. When analyzing data from the synchronous satellite AMPTE CCM, Engebretson et al. (1989) found that the central frequency of the envelope of Pc3 power correlates with B.

In contrast to the work cited above, Yumoto et al. (1984) attempted an analysis of compressional waves recorded at geosynchronous orbit by GEOS-2. They reported a dependence of the frequency of the compressional waves on B as $F = 6B$ (see Fig. 1 of their paper). They have reported also that, in contrast with compressional waves, the transverse waves do not display such dependence. Also, they analyzed simultaneously recorded ground pulsation data, concluding that such data display an $F = 6B$ correlation. Yumoto et al. came to the conclusion that F vs. B correlation is good inside the magnetosphere and worse on the ground.

The main purpose of this paper is to reanalize data of Yumoto et al. and to illustrate the error of their conclusion.

Analysis and discussion

Statistical analysis free of imposed restraint on the solution shows clearly that the data points of Fig. 1 of Yumoto's paper do not fit the relationship $F = 6B$ (see Fig. 1 of this paper). Yumoto et al. forced their solution to go through the point ($F = 0$, $B = 0$). If this restraint is removed, the regression equation is $F = 13 + 4B$ with standart deviation of the slope of 0.4. The center of mass of the point distribution (CM) is located in point (6.67, 39.81). Clearly, the correlation between F and B is weak.

Though Yumoto et al. affirm that the compressional waves within the magnetosphere exhibit $F - B$ dependence, it is our view that the roundish mass of points with no specifiable structure in Fig. 1 of Yumoto et al. demonstrates it to be most probably composed of both types of pulsations, $F - B$ dependent Pc_{sw} and no $F - B$ dependent Pc_{mg}. We believe the fact of lack of $F = 6B$ correlation within the magnetosphere for the compressional waves means that an additional criterion for selection of the $F - B$ dependent pulsations inside the magnetosphere is required. We propose for such a criterion the coincidence of the periods of pulsations recorded

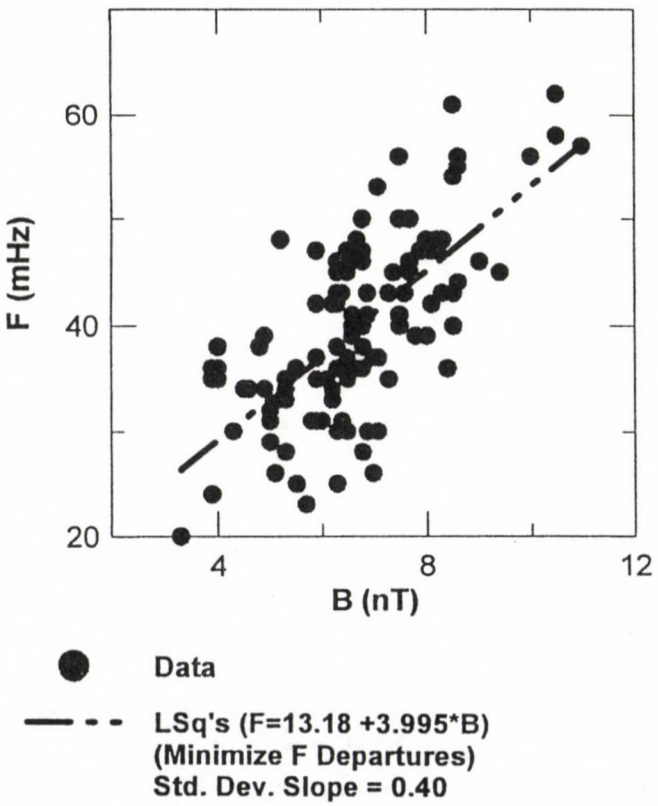


Fig. 1.

both in the magnetosphere and on the ground, as was done in Plyasova-Bakounina et al. (1982). Cases of coincidence between the periods of the compressional wave events inside the magnetosphere and pulsations recorded at the several ground stations — which also correspond to the predicted periods of the upstream waves in the solar wind according to the formula $T = 160/B$ — were demonstrated in the studies of Odera et al. (1994) and Takahashi et al. (1994). These studies support the result of statistical study of Plyasova-Bakounina et al. (1982) and our conclusion that simple selection of compressional waves is an inadequate criterion for defining the $F - B$ dependent Pc_{sw} pulsations inside the magnetosphere.

Supposing that Pc_{sw} are present in the inner magnetosphere, we believe that only one point is important: what percentage of the day-time monochromatic pulsation events occurring within the magnetosphere are Pc_{sw}? According to the estimation of Plyasova-Bakounina (1993), the Pc_{sw} population within the magnetosphere at geosynchronous orbit consists of only 5% of all day-time monochromatic pulsation events, which implies that the majority of pulsation events inside the magnetosphere are composed of the magnetosphere-controlled pulsations Pc_{mg}.

At the ground surface, the situation is quite the opposite. More than 50% of all day-time monochromatic pulsation events recorded at the middle-latitude station Borok in the Pc2-4 frequency range were composed of Pcsw pulsations, the remainder being Pc_{mg} (Plyasova-Bakounina et al. 1986). At the subauroral and high-latitude stations, Pc_{sw} pulsations (range of periods is 5–100 sec) are dominant (Plyasova-Bakounina 1993).

Yumoto et al. also conclude that F vs. B correlation is better inside the magnetosphere than on the ground. By comparing the results of processing two sets of data which cannot be compared they reached this incorrect conclusion. While selecting only compressional waves (from their point of view $F - B$ dependent pulsations) when processing data of geochronous satellite, Yumoto et al. made no selection as regards wave type when processing ground pulsation data but used all pulsation events which were recorded whether Pc_{sw} or Pc_{mg}. The fact that Yumoto et al. did not restrict analysis to pulsations simultaneously recorded at GEOS and at ground stations is shown by the different numbers of points presented on Fig. 1 of their paper: for GEOS $N = 110$, for SGC and OWN stations $N = 201$ and $N = 173$, respectively.

If they had performed their statistical analysis of pulsation data inside the magnetosphere for both transverse waves (Pc_{mg}) and compressional waves (partly Pc_{sw}) as they did for pulsations on the ground, they would not have achieved even a poor F vs. B correlation. As a result of mixture of both types of pulsations Yumoto et al. calculated a poor correlation of T with B for ground-based data. The result is as expected, because only by use of the special method of selection of Pc_{sw} on the ground surface Plysova-Bakounina et al. (1982) were able to demonstrate that Pc_{sw} periods display a very high correlation with B.

Although Alfven resonances in the magnetosphere play the major role in forming the spectrum of Pc_{mg} pulsations, we can draw three conclusions about the necessary character of the initial source of the pulsations observed inside the magnetosphere:

1. their source is located in the solar wind (as their dependence on IMF orientation requires);

2. their source has no F-B dependence;

3. their source is multichromatic (broad-band) in nature because inside the magnetosphere the resonance pulsations are driven at several L-shells simultaneously (Glassmeier et al. 1984).

A source with such characteristics exists in the vicinity of the bow shock — it is the highly-compressional upstream waves of irregular wave form which are generated by reflected "diffuse" protons (Russel and Hoppe 1983, Yumoto 1988). They are multichromatic (broad-band) waves. They do not display any $F - B$ dependence. Their appearance depends upon the orientation of IMF as they are generated by reflected protons, the appearance of which in the solar wind strictly connects with the crossing of the IMF field lines of the bow shock. The reflected protons together with upstream waves disappear before the bow shock when the azimuthal orientation of IMF is 90°.

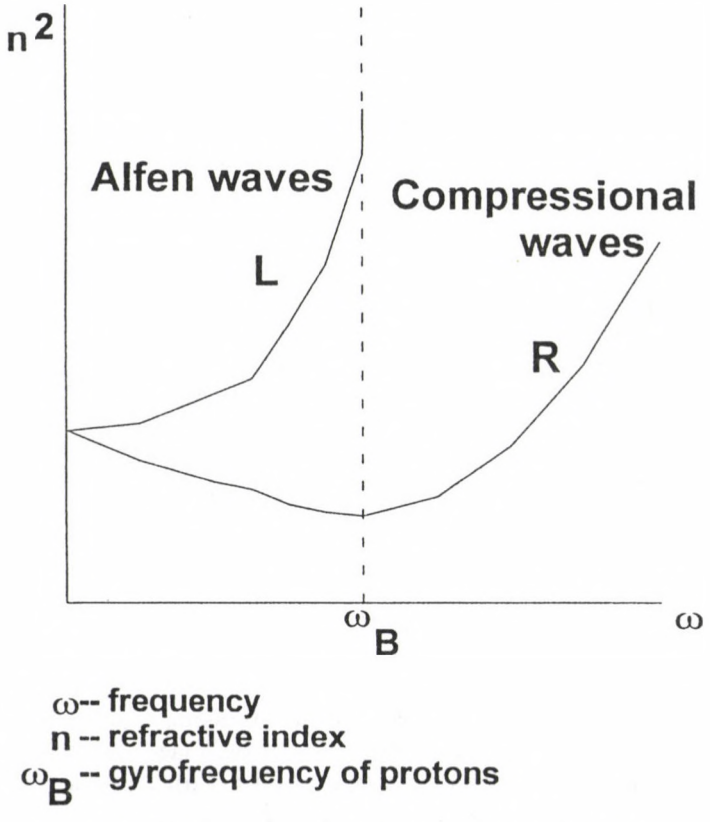


Fig. 2. Dispersion curves of wave spectrum in proton plasma

The dispersion curves of spectra of the excited MHD-waves in the proton plasma of the solar wind show (Fig. 2) that L-waves (Alfven type) are most effectively excited near the gyrofrequency of proton, ω_B (Stix 1968). The phase velocity $V_{ph} = c/n$ (c = speed of light, n = refractive index) takes its minimum value in the vicinity of the gyrofrequency of the protons. Because $\omega_B = 0.1\ s^{-1}$ for typical conditions in the solar wind, Alfven waves are primarily generated in the frequency band of interest to us (low frequeny Pc2-4) although, of course, the compressional waves also are excited.

Upstream waves in the vicinity of the bow shock display a great number of monochromatic waves with frequencies following the formula $F = 6B$ (or $\omega = 0.4\omega_B$), these waves being generated by the mechanism of the two-beams cyclotron instability (solar wind and reflected protons with "intermediate" function of the velocity distribution) (Russell and Hoppe 1983). Their appearance strongly depends on the orientation of IMF. They are mainly transverse waves, but sometimes they display a strong compressional component.

Given their presence within the solar wind, we require an explanation of how these waves enter the magnetosphere. This issue, which is very complicated, cannot be detailed here but some of the mechanisms supporting our hypothesis will be outlined.

The compressional upstream waves

The compressional upstream waves responsible for Pc_{mg} can penetrate into the magnetosphere due to their provoking motion of the boundary of the magnetosphere. The penetration of normally incident compressional waves into the subsolar magnetopause (tangential discontinuity) has been considered both theoretically (Verzariu 1973, McKenzie 1970, Yumoto et al. 1984, Yumoto et al. 1985, Southwood and Kivelson 1990) and experimentally (Wolf and Kaufman 1975, Greenstadt et al. 1983, Song et al. 1993, Engebretson et al. 1991). In contrast with Versariu (1973) and Greenstadt et al. (1983), who reported a transition coefficient of incident power of compressional waves of 1–2%, Song et al. (1993) obtained an observed coefficient of 18%. Inside the magnetosphere, these broad-band, highly-compressional upstream waves couple with Alfven and cavity resonance oscillations at different L-shells, forming Pc_{mg} pulsations. The theory of such coupling is developed by Chen and Hasegava (1974), Southwood (1974), Southwood and Kivelson (1990), Fedorov (1992). Because sometimes the monochromatic $F - B$ dependent upstream waves display a strong compressional component, it must explain the existence of only a small portion of such $F - B$ dependent compressional waves inside the magnetosphere, which penetrate from the solar wind through the subsolar magnetopause.

Considering the magnetopause as a rotational discontinuity and under the assumption that the reconnection of field lines takes place, Kwok and Lee (1984) considered the propagation of ULF waves of all polarizations across the magnetopause. They demonstrate that in this case (in contrast with Versariu's model) their wave power even can be amplified.

The transversal upstream waves

"High latitude" entry has been suggested for penetration to the ground of the $F - B$ dependent mainly transverse upstream waves (Plyasova-Bakounina et al. 1984, Engebretson et al. 1989, 1991). We believe two possible modes of penetration of the energy of the monochromatic transverse waves: one might be a non-linear mechanism of the exchange of energy with the surface waves on the magnetopause, as the Pc_{sw} pulsations appear when the condition for development of a Kelvin-Helmholz (K-H) instability on the magnetopause is fulfilled (Golikov et al. 1980). We believe that the corresponding frequencies in the wide spectrum of surface waves being driven by the K-H instability on the flanks of the magnetosphere amplified by interaction with the monochromatic upstream waves (which display $F = 6B$) by some as yet unknown mechanism. Such waves can observed at the ground surface as Pc_{sw}. Another possibile mode for penetration of the energy of transverse waves is reconnection of the field lines (model of Kwok and Lee 1984). The fact demonstrated by Plyasova-Bakounina et al. (1978) that Pc_{sw} pulsations are recorded on the ground when the radial component of IMF (B_x) is directed away from the sun implies that the reconnection of B_x with magnetospheric field lines takes place in the point which is projected into the cusp/cleft region. The transverse waves are guided along cusp field lines to the ground surface. Such a process explains very nicely that:

1. Pc_{sw} display their maximum of intensity in the cusp region (Plyasova--Bakounina et al. 1984, 1986, Bolshakova and Troitskaya 1984, Engebretson et al. 1986, Morris and Cole 1987);

2. Pc_{sw} on the ground displays a pronounced $F = 6B$ dependence.

3. The amplitudes of Pc_{sw} pulsations depend on the orientation of IMF.

The other possibility of "indirect" entry of wave energy into the magnetosphere is discussed by Southwood and Kivelson (1991) and Kivelson and Southwood (1991), who developed the mechanism of the non-linear transport into the cusp latitudes of the magnetosheath pressure impulses impinging on the magnetopause. We believe that this model qualitatively fits better for explanation of excitation of pulsations connected with sub-storms (Pi1, Pi2) than for monochromatic Pc_{sw} pulsations, which do not display any correlation with the B_z component (Plyasova-Bakounina 1993). Finally, we note that Russell and Elphic (1978) proposed a much different model of penetration of the energy of transverse waves by creating on the magnetopause mini-flux transfer events (FTE's). Though their mechanism may quite possibly be correct for upstream wave penetration into the magnetosphere, presently available data cannot determine the relationships between "FTE events" and "Pcsw pulsations".

It is still a question how the Pcsw energy reaches the middle and low latitudes. Supposedly, the energy of these waves can be transferred through the ionosphere current system, which is created due to the modulation of the ionosphere conductivity. The modulation can be affected by electron precipitation accompanied by

pulsation events (Chernous et al. 1986). This idea was developed by Engebretson et al. who suggested the creation of the 3-dimensional current system, which transfers the energy of the waves not only to low latitudes but also to the inner magnetosphere.

We believe the lack of Pc_{sw} inside the magnetosphere testifies that the Pc_{sw} current system is located only at ionospheric heights. A similar situation exists for the DP2 current system which is a response to the southward component of IMF (Nishida 1978). The observation of the magnetic field at satellites at the inner magnetosphere demonstrates the absence of variations which are synchronous with IMF (Θ) variations.

Conclusion

The fundamental difference between our scenario of penetration of the wave energy and that of others (Yumoto et al. 1984, 1985 and Engebretson et al. 1989, 1991) is that these authors believe that inside the magnetosphere the pulsations display the $F - B$ correlation exactly as upstream waves in the solar wind and Pc_{sw} on the ground. Hence, they must construct scenarios which explain such a phenomenon, which actually is a secondary phenomenon, as we have demonstrated in this paper.

Yumoto suggested "direct entry" of the monochromatic $F - B$ dependent upstream waves into the magnetosphere through the subsolar point of the magnetopause. Then, because "direct entry" is preferable for compressional waves, Yumoto declares that the monochromatic $F-B$ dependent upstream waves are of compressional mode. However, such a conclusion is contradicted both by the theoretical estimation of the exitation of the waves in the proton plasma and by the experimental observations of the monochromatic upstream waves which have demonstrated that such waves are primarily transverse waves. Also, there are difficulties for such a model in explanation of the location of the maximum of the intensity of the $F-B$ dependent pulsations at the cusp area.

In order to explain his observation of the maximum of $F - B$ dependent pulsations in the cusp region, Engebretson et al. (1989) have suggested penetration of monochromatic $F - B$ dependent upstream waves through the cusp ("high latitude entry"). (They did not specify whether the waves were of compressional or transverse type.) But believing Yumoto's statement of the presence within the magnetosphere of $F - B$ dependent pulsations, Engebretson must try to explain this supposed fact. He hypothesizes a scenario involving the creation of a 3-dimensional field-aligned current system (cusp ionosphere – magnetosphere). The problem of all this development is that Yumoto's conclusion is not correct and Engebretson hypothesis is striving to explain a secondary phenomenon.

Our model of the penetration of the upstream wave energy into the magnetosphere and to the ground has a merit not found in other models in that we explain the two observed types of pulsations observed on the ground (the $F - B$ dependent solar wind-controlled Pc_{sw} and the no $F - B$ dependent magnetosphere-controlled Pc_{mg}) via the two main types of upstream waves known to exist in the vicinity of

the bow shock. Each type of upstream waves has its own preferred mode of entry into the magnetosphere. The broad-band highly-compressional upstream waves penetrate through the subsolar point of the magnetopause coupling into Alfven and cavity resonances of the magnetosphere resonator, while the $F - B$ dependent transverse upstream waves penetrate through the cusp forming the $F - B$ dependent Pc_{sw} pulsations on the ground with maximum intensity in this area.

The energy of Pc_{sw} pulsations transfers to low latitudes through the pulsating current system. The lack of Pc_{sw} in the magnetosphere testifies that currents do not spread into the magnetosphere. A similar situation exists for the DP2 current system.

References

Anderson B J, Potemra T A, Zanetti L Z, Engebretson M J 1991: In: Physics of Earth Plasmas (1990) SPI Conference Proceedings and Reprint Series, 10, 419, Scientific Publishers Inc., Cambridge, Mass.

Arthur C W, McPherron R L 1977: *J. Geophys. Res.*, 82, 5738.

Bolshakova O V, Troitskaya V A 1968: *Dokl. Akad. Nauk USSR*, 180, 343.

Bolshakova O V, Troitskaya V A 1984: *Geomagn. Aeron.*, 24, 633.

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

Engebretson M J, Ching-I Meng, Arnoldy R A Jr, Cahill L J 1986: *J. Geophys. Res.*, 91, 8909.

Engebretson M J, Anderson B J, Cahill L J, Arnoldy R L Jr, Newell P T, Meng C I, Zanetti L J, Potemra T A 1989: *J. Geophys. Res.*, 94, 17143.

Engebretson M J, Cahill L J, Arnoldy R L Jr, Anderson, B J, Rosenberg T J, Carpenter D L, Inan U S, Eather R H 1991: *J. Geophys. Res.*, 96, A2, 1527.

Fairfield D H 1969: *J. Geophys. Res.*, 74, 14, 3541.

Fedorov E N 1992: Doctoral thesis, Inst. Phys. Earth, Moscow (in Russian)

Glassmeier K H, Volpers H, Baumjohann W 1984: *Planet. Sp. Sci.*, 32, 1463.

Golikov Yu V, Plyasova-Bakounina T A, Troitskaya V A, Chernikov A A, Pustovalov V V, Hedgecock P C 1980: *Planet. Space Sci.*, 28, 535.

Greenstadt E W, Olson J W 1977: *J. Geophys. Res.*, 82, 4991.

Greenstadt E W, Mellott M M, McPherron R L, Russell C T, Singer H J, Knecht D J 1983: *Geophys. Res. Lett.*, 10, 659.

Gul'elmi A V, Plyasova-Bakounina T A, Chepetnov R V 1973: *Geomagn. Aeron.*, 13, 331.

Kivelson M G, Pu Z Y 1984: *Planet. Sp. Sci.*, 32, 1335.

Kivelson M G, Southwood D J 1986: *J. Geophys. Res.*, 91, 4345.

Kivelson M G, Southwood D J, 1991: *Geophys. Res.*, 96, A2, 1661.

Kovner M S, Kuznetsova V A, Lebedev V V, Plyasova-Bakounina T A, Troitskaya V A 1976: *Annales Geophys.*, 32, 3, 189.

Kwok Y C, Lee L C 1984: *J. Geophys. Res.*, 89, 10697.

Lanzerotti L J, Medford L V, MacLennan C G, Hasegava T, Acuna M H, Dolce S R 1981: *J. Geophys. Res.*, 86, 5500.

McKenzie J F 1970: *Planet. Space Sci.*, 18, 1.

Morris R J, Cole K D 1987: *Planet. Space Sci.*, 35, 1437.

Nishida A 1978: Geomagnetic diagnosis of the magnetosphere. Springer-Verlag, New York, Heidelberg, Berlin

Odera T J, Van Swol D, Russell C T 1994: Solar wind sources of magn. ULF Waves Geophys. Monogr. 81, Copyright by AGU

Plyasova-Bakounina T A 1972: *Geomagn. Aeron.*, 11, 4.

Plyasova-Bakounina T A 1993: Doctoral Thesis. Inst. Phys. Earth, Moscow (in Russian)

Plyasova-Bakounina T A, Golikov Yu V, Troitskaya V A, Hedgecock P C 1978: *Planet. Sp. Sci.*, 26, 547.

Plyasova-Bakounina T A, Stuart W F, Charchenko I P, Troitskaya V A 1982: Issledov. po Geomagn. Aeron. i Fysiki Solntsa, Irkutsk, 62, 103 (in Russian)

Plyasova-Bakounina T A, Münch J W, Gauler H, Troitskaya V A 1984: In: Program and theses of "All-Union Seminar, on Geomagnetic Pulsations, Irkutsk, 1984", Proc. SibIZMIR of Acad. Sci. of USSR, Irkutsk (in Russian)

Plyasova-Bakounina T A, Münch J W, Gauler H F, Troitskaya V A 1986: *Acta Geod. Geoph. Mont. Hung.*, 21, 143.

Russell C T, Elphic R C 1978: *Space Sci. Rev.*, 22, 681.

Russell C T, Hoppe M M 1983: *Space Sci. Rev.*, 34, 155.

Song P, Russell C T, Strangeway R J, Wygant J R, Cattelli C A, Fitzenreiter R J, Anderson R R 1993: *J. Geophys. Res.*, 98, A1, 187.

Southwood D J 1974: *Planet. Sp. Sci.*, 22, 483.

Southwood D J, Kivelson M G 1990: *J. Geophys. Res.*, 95, 2301.

Southwood D J, Kivelson M G 1991: *J. Geophys. Res.*, 96, A1, 67.

Stix T H 1962: The theory of plasma waves. McGraw-Hill, New York

Takahashi K, McPherron R L, Greenstadt E W, Neeley C A 1981: *J. Geophys. Res.*, 86, A7, 5472.

Takahashi K, McPherron R L, Terasawa T 1984: *J. Geophys. Res.*, 89, A5, 2770.

Takahashi K, Anderson B J, Newell P, Yamamoto T, Sato N 1994: Solar Wind Sources of Magn. ULF Waves, Geoph. Monogr. Copyright by AGU

Troitskaya V A, Plyasova-Bakounina T A, Gul'elmi A V 1971: *Dokl. Akad. Nauk USSR*, 197, 1312.

Verő J 1986: *J. Geophys. Res.*, 60, 106.

Verzariu P 1973: *Planet. Space Sci.*, 21, 2213.

Webb D, Orr D 1976: *J. Geophys. Res.*, 81, 5941.

Wolf A, Kaufman R L 1975: *J. Geophys. Res.*, 80, 1764.

Wolf A, Lanzerotti L J, Maclennan C G 1980: *J. Geophys. Res.*, 85, 114.

Wolf A, Kamen E, Lanzerotti L J, Maclennan C G, Bamber J F, Venkatesan D 1987: *J. Geophys. Res.*, 92, A1, 168.

Yumoto K 1988: *Geomag. Geoelectr.*, 40, 293.

Yumoto K, Saito T, Tsurtani B T, Smith E J, Akasofu S I 1984: *J. Geophys. Res.*, 89, 9731.

Yumoto K, Saito T, Akasofu S I, Tsurtani B T, Smith E J 1985: *J. Geophys. Res.*, 90, 6439.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 203–223 (1997)

ELECTROMAGNETIC INDUCTION PROFILE (PREPAN95) FROM THE EAST EUROPEAN PLATFORM (EEP) TO THE PANNONIAN BASIN

A ÁDÁM[1], T ERNST[2], J JANKOWSKI[2], W JOZWIAK[2], M HVOZDARA[3], L SZARKA[1], V WESZTERGOM[1], I LOGVINOV[4], S KULIK[4]

[Manuscript received April 28, 1997]

Complex electromagnetic measurements along a profile crossing different great tectonic units in East-Central Europe (East European Platform, Paleozoic Europe and Carpatho-Pannonian region) were carried out by Hungarian, Polish, Slovak and Ukrainian institutes. Beside recent MT sounding curves, geomagnetic induction vectors, polar diagrams, apparent resistivity and phase pseudosections, a short description of earlier magnetotelluric (MT) and magnetovariational (MV) experiments is also given. For interpretation of these data 1D and 2D inversions were used after separation of the quasi E and B polarized curves. The Carpathian conductor clearly appears in the magnetovariational profile. MT soundings indicate that the mantle conductive basement — presumabely the asthenosphere — steeply deepens from the Neogene Pannonian Basin characterized by high heat flow towards the much colder EEP through the Paleozoic area.

Keywords: asthenosphere; Carpathian conductivity anomaly; induction vector; magnetotellurics

Introduction

In the area of East-Central Europe — in Hungary, Slovakia and Poland — within a relatively short distance (300–400 km) very different great tectonic units touch each other, among them the Neogene Pannonian Basin, the folded belts of the Carpathians with their foredeep molasse, Paleozoic and Precambrian Platforms and between the last two ones the Teisseyre Tornquist Lineament (TTL). This very special and curious geologic-tectonic situation calls the attention of all the geoscientists who are interested in the deep Earth's structure to study and compare the crustal and mantle structure of these units. The first magnetotelluric (MT) and magnetovariational (MV) measurements were carried out in the frame of an earlier co-operation (so-called KAPG) along profiles such as the international KAPG III which started from the Pannonian Basin and after crossing the Carpathians ended in the Ukrainian shield (e.g. Zhdanov et al. 1986) already 2–3 decades ago. Since then

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

[2]Institute of Geophysics of Polish Academy of Sciences, 01–452 Warszawa, ul. Ks. Janusa 64, Poland

[3]Institute of Geophysics, Slovak Adademy of Sciences, 842 28 Bratislava, Dúbravská 9, Slovak Republik

[4]Institute of Geophysics, NAS of Ukraine, 252142 Kiev, Palladin Pr. 32, Ukraine

1217-8977/97/$ 5.00

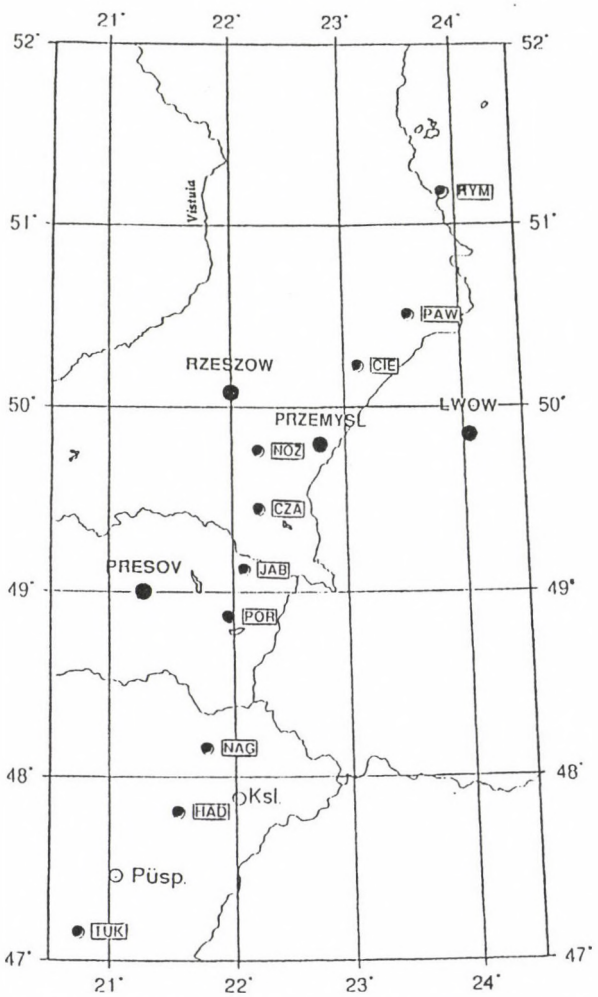


Fig. 1. Geographic map of the measuring sites RYM, PAW, NOZ, CZA, JAB, POR and NAG are the sites of the synchronous field campaign in summer of 1995 ('PREPAN95'), Ksl, HAD, Püsp and TUK are further (non-synchronous) MT sites in Hungary

the magnetotelluric measurement and data processing techniques and interpretation reached a much higher level due to sensor developments, digital data acquisition systems and computer interpretation techniques. It seemed to be promising to carry out new MT and MV measurements along a profile crossing the mentioned great tectonic units using up-to-date techniques. The time difference between the new and old measurements brought not only advantages but disadvantages, too, due to the increased level of the civilization noise which limits resolving power. (See e.g. the MT station Porubka in the Carpathians.)

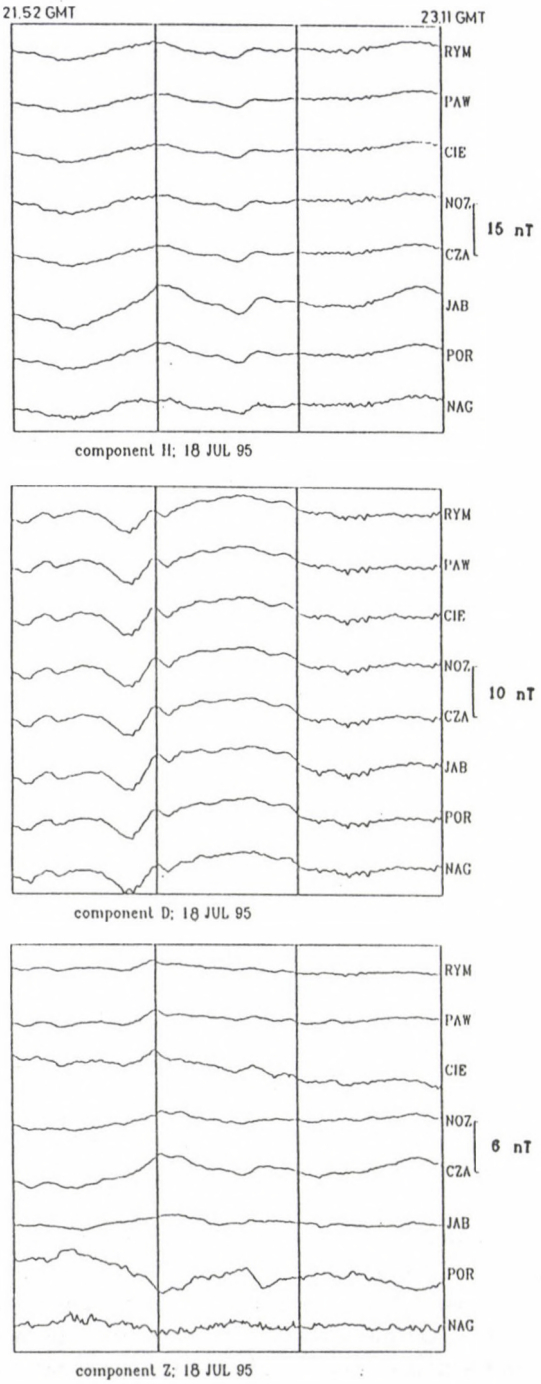


Fig. 2. Synchronous recordings of geomagnetic components (H, D, Z) on July 18, 1995

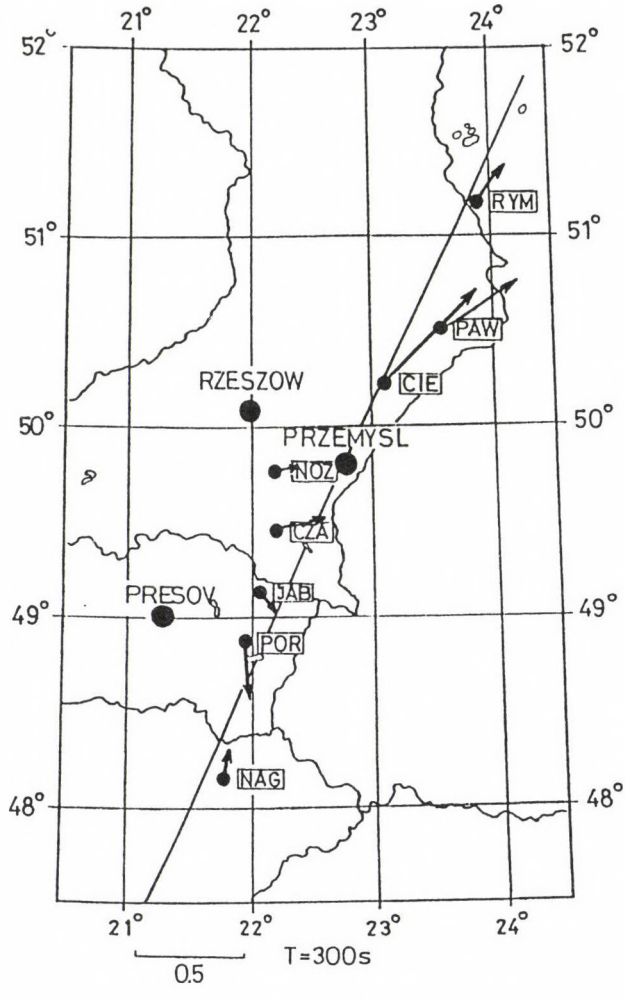


Fig. 3a. Induction vectors for period T = 300 s

Field measurements

Similarly to the Tatra profile published quite recently (Ernst et al. 1997), in this new experiment Polish, Hungarian, Slovak (and partly Ukrainian) geophysicists took part.

Synchronous measurements were carried out at 8 sites along a N NE–S SW profile. Five electromagnetic components: E_x, E_y, H_x, H_y and H_z were measured. As the MT site "Nagyhalász" (NAG) in Hungary was electrically distorted by geological noise four other MT stations have been involved into the interpretation from the NE segment of the Pannonian Basin and only the geomagnetic components of the NAG station were used for magnetovariational sounding. The additional Hungarian stations are: Túrkeve (TUK), Püspökladány (Püsp), Hajdúdorog (HAD) and

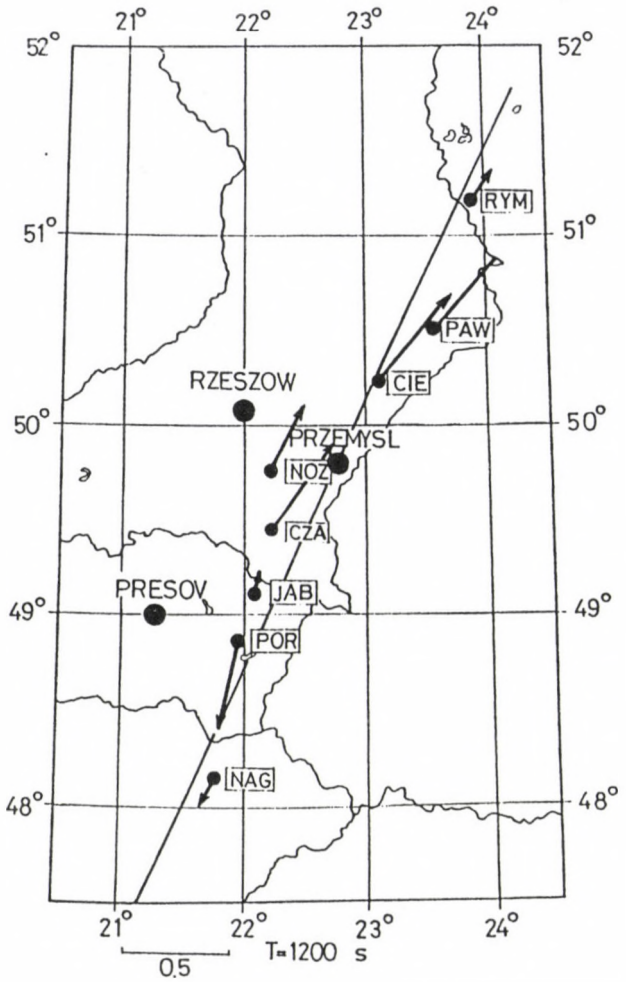


Fig. 3b. Induction vectors for period T =1200 s

Kisléta (Ksl). In the case of the two Slovak stations (Porubka (POR) and Jablonka (JAB)) the noise is similar to NAG. All these sites are shown in Fig. 1.

All measuring groups used the same instruments (DR02) developed at Belsk observatory in Poland (Jankowski et al. 1984). The magnetic variometers allow us to record the field variations in a wide period range starting from 5 s. The sampling rate for digital recording used in field measurements, varied between 2 and 20 s. In the telluric lines of about 100 m length unpolarized electrodes were used. The average duration of continuous recording was about 3 weeks. An example of simultaneous recordings of geomagnetic field variations at the synchronous stations is presented in Fig. 2.

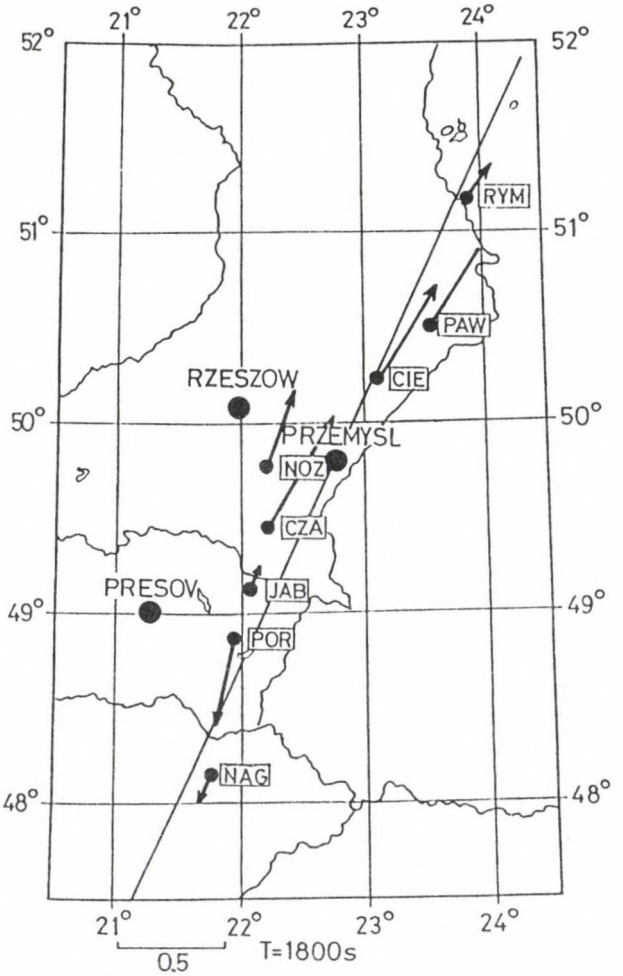


Fig. 3c. Induction vectors for period T = 1800 s

Data processing

The magnetotelluric and geomagnetic transfer functions were calculated by using three different algorithms. The Polish group used two different programs — one works in time domain (Wieladek and Ernst 1977) — the other one works in frequency domain (Semenov and Kaikkonen 1986). The Hungarian group used their own data processing method based on the digital filtering (Verő 1972, Ádám and Verő 1976). All these methods were compared by Ernst et al's (1997) by using magnetotelluric data from Litke station. It has been stated that "the coincidence of all these curves proved to be satisfactory for deep sounding studies."

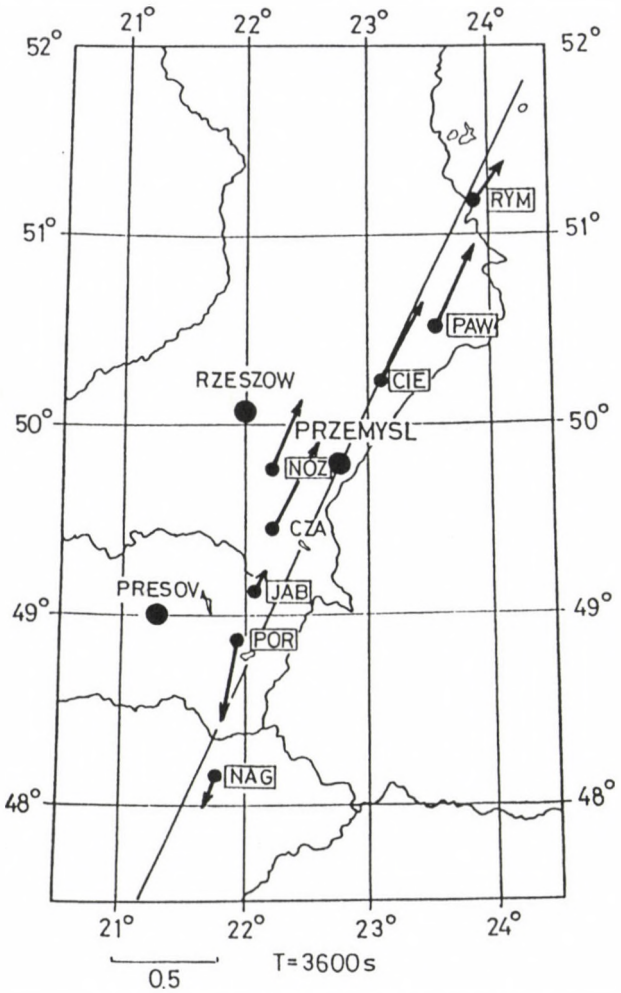


Fig. 3d. Induction vectors for period T = 3600 s

Magnetovariational sounding and distribution of electrical resistivity in the upper crust

The induction vectors have been calculated for several periods (T) on the basis of the relation between the variations of the geomagnetic vertical H_z and horizontal components (H_x, H_y)

$$H_z = X_{zx} H_x + X_{zy} H_y \,.$$

For periods T = 300, 1200, 1800 and 3600 s the induction vectors (arrows) are given in Figs 3a, b, c, and d. The Carpathian conductivity anomaly appears with very small vectors at all periodes: at T = 300 s the vector is directed towards S SE and at periods higher than T = 1200 s it is directed northward. The conductor was

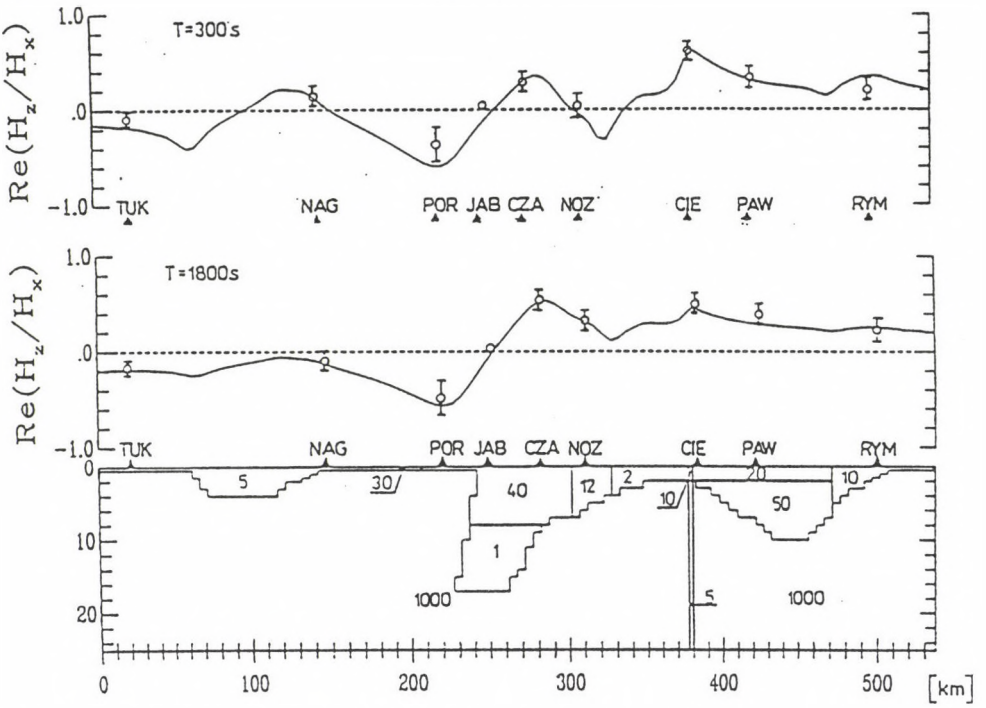


Fig. 4. A model of the geoelectric crustal structure along the profile PREPAN95 based on synchronous magnetovariational measurements

found beneath the MV site JAB (between CZA and POR). At CZA and POR the vectors are longer and have opposite direction similarly to other MV profiles crossing the Carpathian conductivity anomaly (e.g. Jankowski et al. 1991). The vectors are interpreted in sense of Wiese-arrows which point towards the less conductive formations. At the site NAG in the Pannonian Basin, the induction vector between the periods 300 and 1200 s turns by 180° and at longer periods it follows the general tendency of the Wiese-arrows in the centre of the Pannonian Basin (i.e. Hungary) (Ádám et al. 1972).

As the magnetotelluric measurements were strongly disturbed in the Carpathians (one of the most interesting part of the profile), the crustal model of the Carpathian conductivity anomaly (CA) was derived only from the magnetovariational soundings. The crustal model for this area (Fig. 4) was checked at two periods (300 and 1800 s). The most outstanding feature of this model is the Carpathian conductor of 1 Ωm at a depth between 8 and 17 km. It is situated roughly beneath the MV station JAB and limited by stations POR and CZA. Concerning the other parts of the profile, a fault zone with a resistivity of 5 Ωm beneath CIE was assumed. This assumption is based primarily on the experience obtained along other Carpathian MV profiles, including the most recent "Tatra" one (Ernst et al. 1997). Of course, the shape of the anomaly cannot be well determined by MV measurements because of the large distance between the measuring sites.

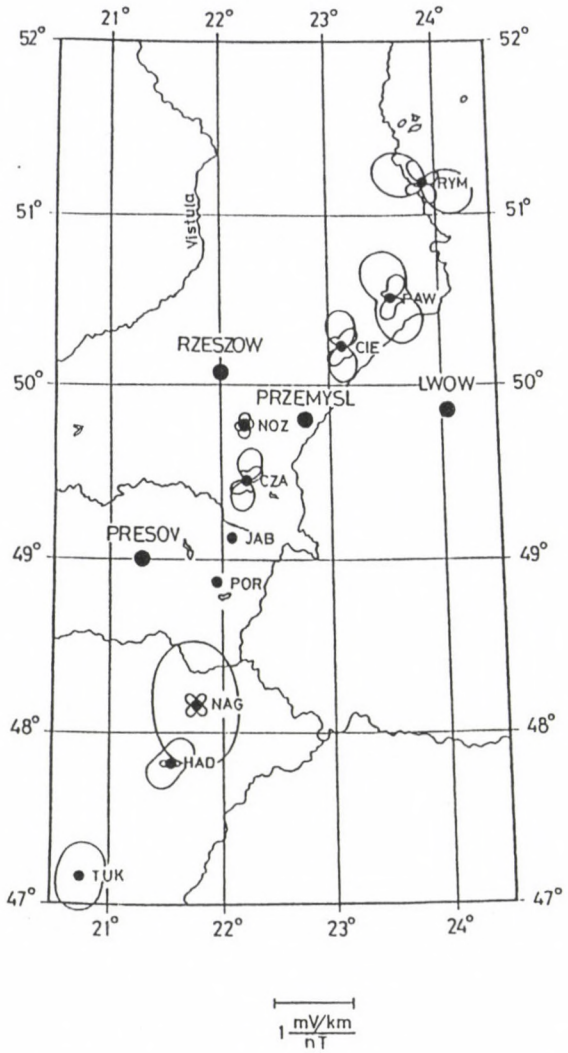


Fig. 5. Z_{xy} and Z_{xx} polar diagrams at the period T = 1600 s

Magnetotelluric investigations

The deep structure of the great tectonic units crossed by the profile should be characterized on basis of as many MT sites as possible. This is the reason why the profile had to be completed by further four MT deep soundings. In case of TUK and HAD, the MT soundings were carried out by the same Polish DR02 MT instrument. At the Püsp and Ksl sites Ádám and Major's (1967) static magnetic variometers were used.

In Hungary the site NAG could not be used because of some unknown geological noise. In Slovak area — especially in the Carpathians — the electric noise caused

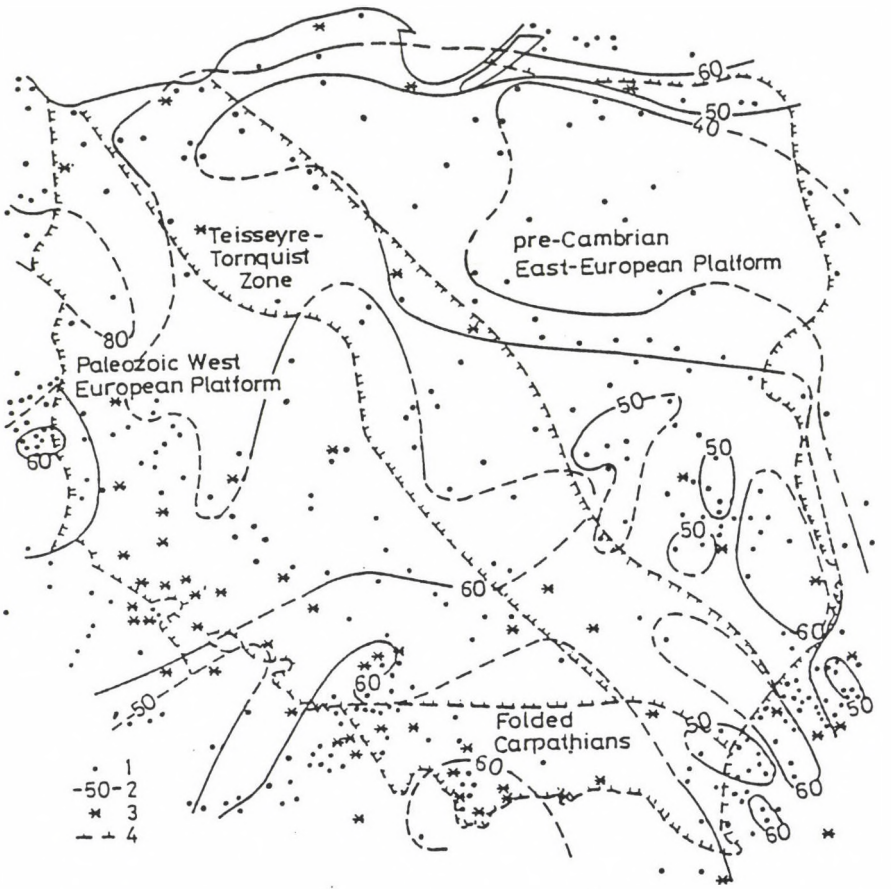


Fig. 6. Heat flow values of Poland (after Gordienko and Zavgorodnyaya, 1996). 1 – points of heat flow determination, 2 – heat flow isolines in mW/m^2, 3 – earthquake epicenters, 4 – boundaries of tectonic zones

mainly by DC powered railways hinders the use of MT records at Porubka (POR). (Due to this noise the resistivity is so strongly increased that 1D inversion gives a depth of about 1000 km for the asthenosphere.)

The extreme values of the resistivity (Rhomax and Rhomin) and the corresponding phase curves are the final products of MT data processing. In Fig. 5 typical impedance polar diagrams are shown for the Polish, Ukrainian and Hungarian MT sites at $T \doteq 1600$ s. The well known relationship between the polar diagrams and the tectonics can generally help in the selection of MT polarizations (or quasipolarizations) in connection with the geological strike. The Z_{xymax} directions of the polar diagrams of this profile lie near N-S (except RYM), nevertheless this direction cannot be connected to the NW-SE direction of the TTL or to the northern boundary of the folded Carpathians (see Gordienko and Zavgorodnyaya's (1996) map in Fig. 6). Local 3D tectonic/geological noises should be assumed.

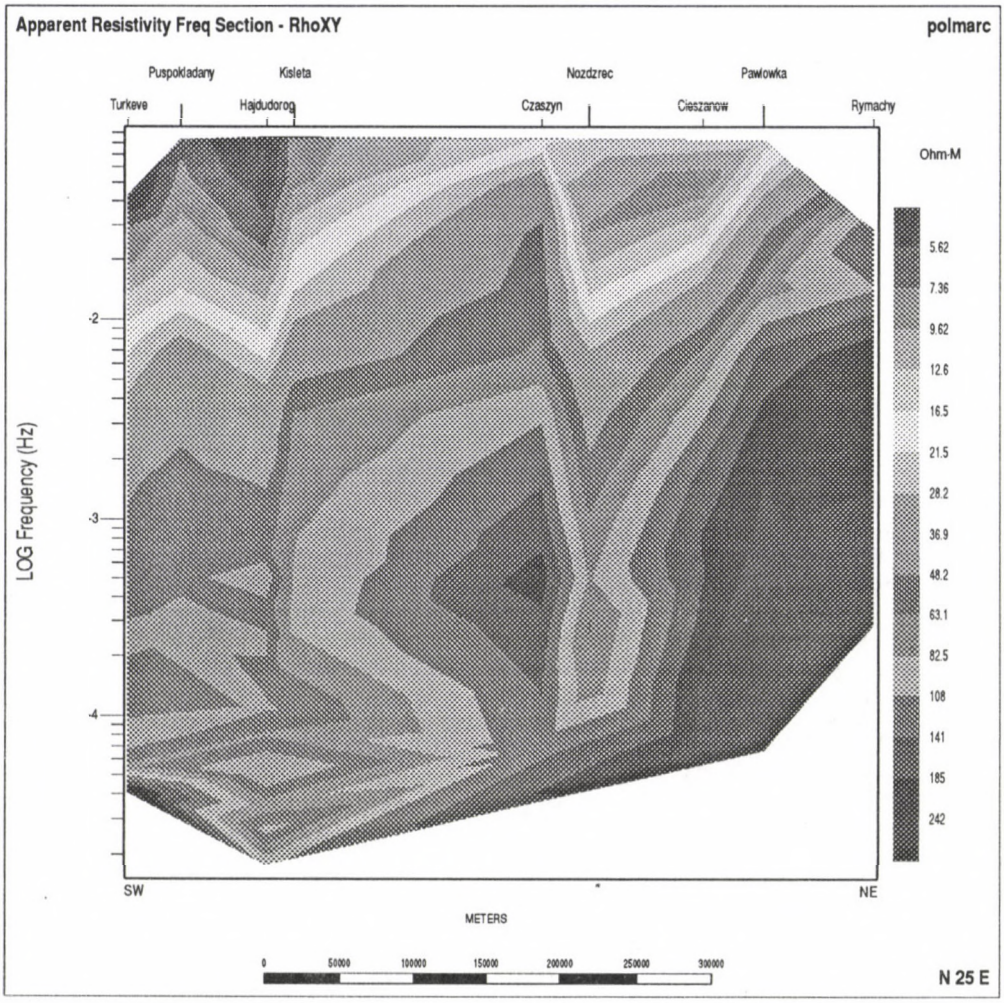


Fig. 7a. MT pseudosections along the profile shown in Fig. 1. Apparent Resistivity Frequency Section Rhomax (xy)

In such uncertain cases, as a first approximation, the extreme apparent resistivity curves are to be analysed and the final conclusion is to be drawn from their combined interpretation.

The first rough analysis of the data starts with pseudosections. These were prepared in different combinations in two versions: in the first one the disturbed soundings (Nagyhalász and Porubka) were included and in the second one they were omitted. In the site Porubka the noise was so high that only the Rhomax curve could determined therefore it appears only in the Rhomax (xy) pseudosection.

In Fig. 7 the aforementioned second version of Rhomax (xy) and Rhomin (yx) apparent resistivity vs. frequency and impedance phase vs. frequency pseudosec-

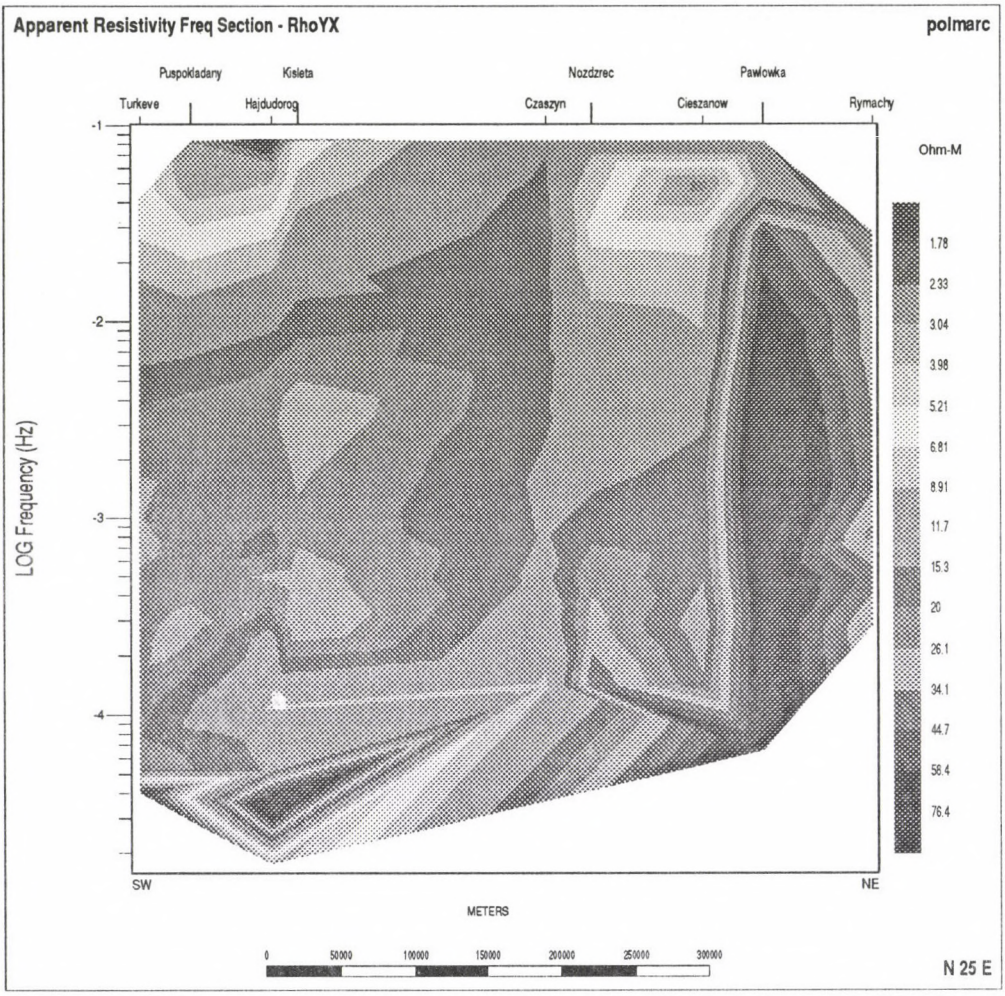


Fig. 7b. MT pseudosections along the profile shown in Fig. 1. Apparent Resistivity Frequency Section Rhomin (yx)

tions are given. These pseudosections clearly show the sedimentary basins both in the Pannonian Basin and in the Carpathian foredeep molasse characterized by low resistivity and low phase values. The EEP appears in the sections with high resistivity.

In the Rhomin (yx) pseudosection at periods greater than 10^4 s the asthenosphere in the Pannonian Basin is indicated by resistivity decrease and high phase values.

As a next step in the interpretation process, 1D layer sequences were calculated by using Marquard inversion from the extreme sounding curves. Typical examples for 1D model-fitting are shown in Fig. 8 for site Túrkeve in the Pannonian

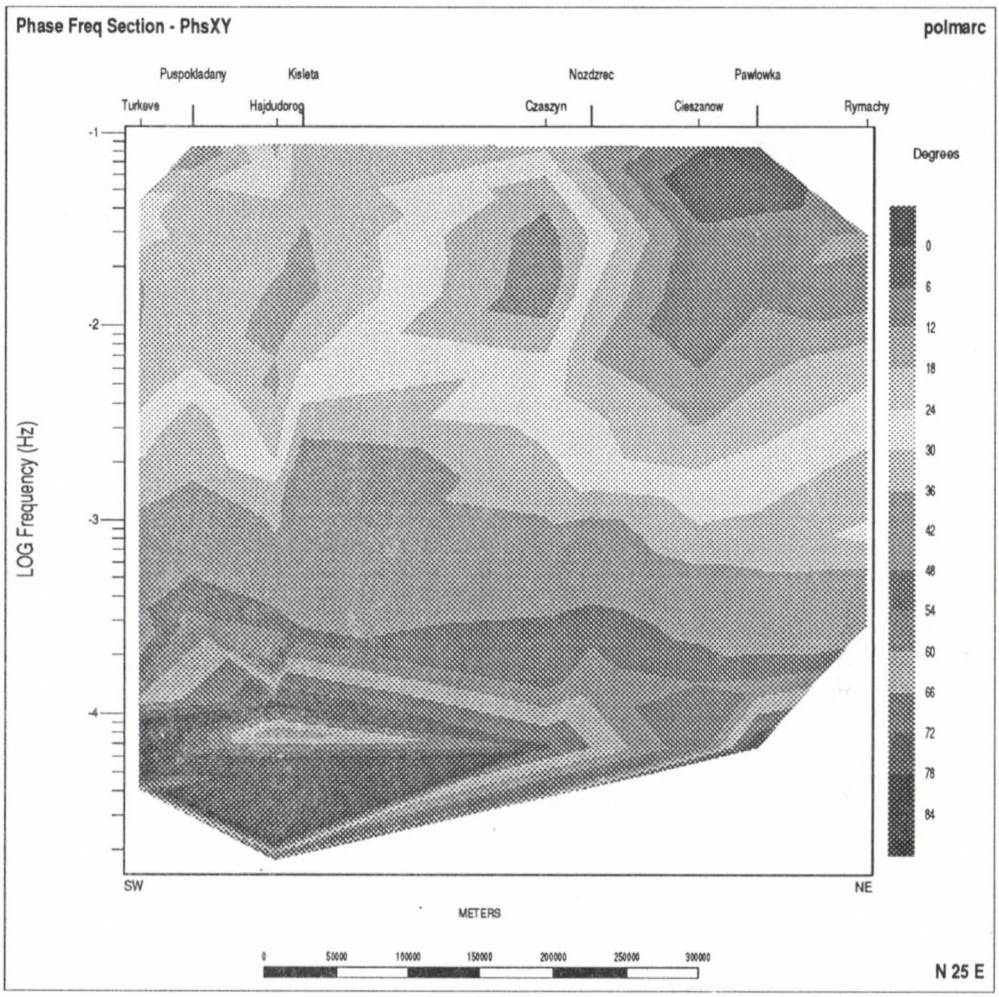


Fig. 7c. MT pseudosections along the profile shown in Fig. 1. Phase Frequency Section (PHSxy)

Basin and and in Fig. 9 for Nozdrzec in the Paleozoic platform. (In both cases the Rhomax curves are illustrated.) The conductance values of the upper (sedimentary) layer (S_1) and the depth values to the conductive basement in the mantle (asthenosphere?) at all sites are given in Table I. While the conductance values are in accordance with the surface geology, i.e. the sedimentary structure, the depth values of the conductive basement — at the first glance — are strongly scattered certainly due to 3D effects, to static shift and to other EM distortions. Some homogenization can be reached according to Berdichevsky and Dmitriev (1976) by computing the geometric mean (approximation to Berdichevsky's determinant, see Appendix) to get an idea about the regional tendencies. If the data measured in the Pannonian Basin and those in the Paleozoic and Precambrian Platforms are

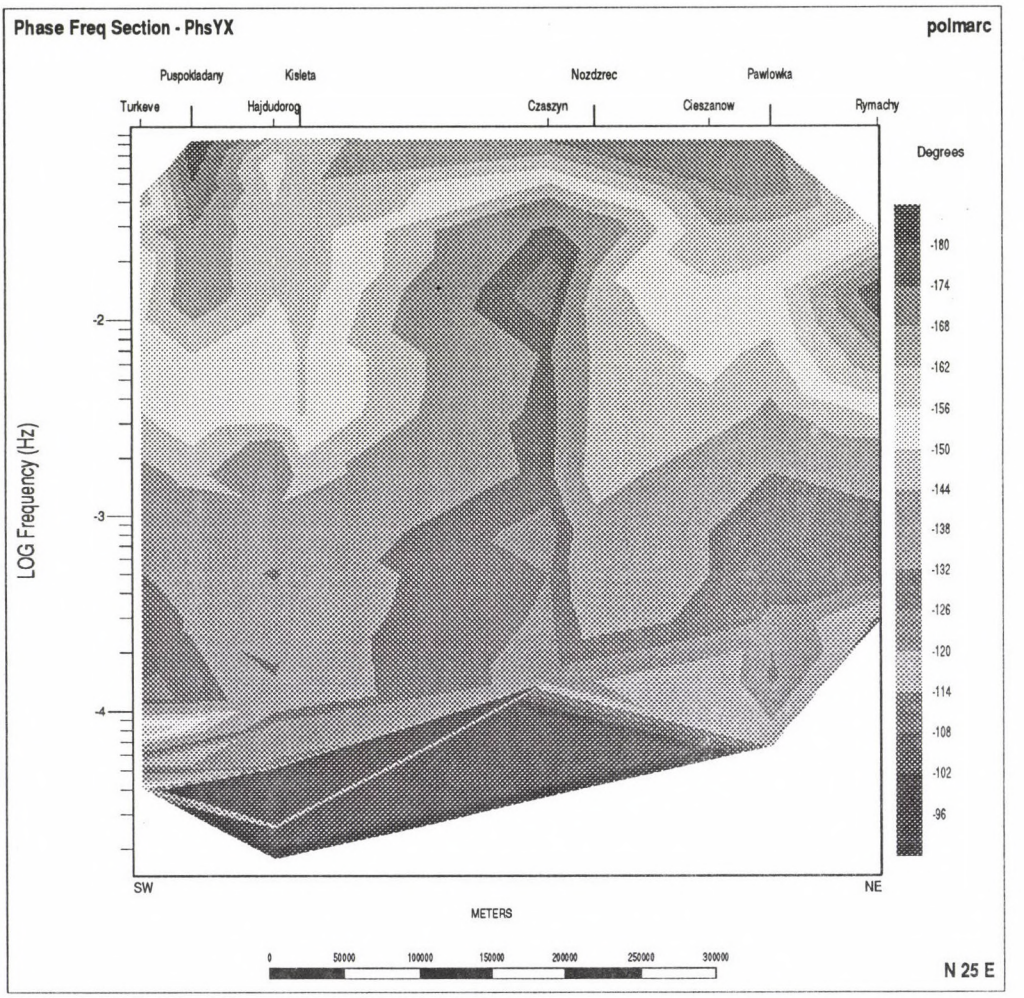


Fig. 7d. MT pseudosections along the profile shown in Fig. 1. Phase Frequency Section (PHSyx)

separated in two groups, the following values are obtained:

The depth to the mantle conducting basement (h) (asthenosphere) calculated from the geometric mean values of h_{Rhomax} and h_{Rhomin}.

Pannonian Basin (4 data): 65.7±9.1 km
Platforms (5 data): 122.7±26.6 km

Conducting basement depth from the Rhomax curves (h_{Rhomax}):

Pannonian Basin (4): 85.7±18.9 km
Platforms (5): 226.4±115.3 km

Conducting basement depth for the Rhomin curves (h_{Rhomin}):

Pannonian Basin (4): 50.9±5.6 km
Platforms (5): 77.1±33.8 km

Table I.

MT site	Sediment			Asthenosphere		
	S_{Rhomax} [Siemens]	S_{Rhomin}	S_{av}	h_{Rhomax} (1) [km]	h_{Rhomin} (2)	$h_{av} = \sqrt{h_1 h_2}$
Túrkeve	929	871	900	66.5	50.7	58.1
Püspökladány	400	700	550	105.2	60.0	79.4
Kisléta	590	420	505	67.0	48.0	56.7
Hajdúdorog	750	684	717	104.0	45.0	68.4
Hagyhalász	425	467	446			
Average for the Pannonian Basin				85.7 ± 18.9	50.9 ± 5.6	65.7 ± 9.1
Czaszyn	247	?	247	415.6	45.4	137
Nozdrzec	666	483	574.5	152.8	141.8	147.8
Cieszanov	581	566	573.5	163.2	61.0	99.8
Pawlowka	400	313	356.5	300.0	76.1	151.0
Rymachy	261	319	290.0	100	61.1	78.3
Average for the Platform				226.4 ± 115.3	77.1 ± 33.8	122.7 ± 28.6

The depth differences in the conductive basement between these two areas have the same character in all the three data sets. The conductive basement is shallower in the Pannonian Basin than in the much older platforms. Which depth value is more probable?

Using the empirical relation between the heat flow and the depth to the asthenosphere (Ádám 1978)

$$h = 155 \cdot q^{-1.45}$$

(q = heat flow in old HFU, 1 HFU $\doteq$ 42 mW/m^2) and Gordienko and Zavgorodnyaya's (1996) heat flow values measured in the Polish area crossed by the PREPAN95 profile (50–60 mW/m^2) and the average heat flow values of the Pannonian Basin (80–100 mW/m^2) the calculated depths are as follows:

in case of 50 mW/m^2 : 120.2 km
80 mW/m^2 : 61.1 km

Comparing these values with the magnetotelluric average depth values, it seems that the the geometric means of h_{Rhomax} and h_{Rhomin} give the best approximation of depth values derived from the regional heat flow.

This example proves that the layer parameters — in this case the depth to the mantle conductive basement — derived by MT measurements can only be statistically correlated to the heat flow in regional sense as both the magnetotelluric and the heat flow data may be disturbed by local effects.

The main tendencies in asthenospheric structures of these great tectonic units can be seen on the Rhomax "Stiched 1D model section", too (Fig. 10). In this figure

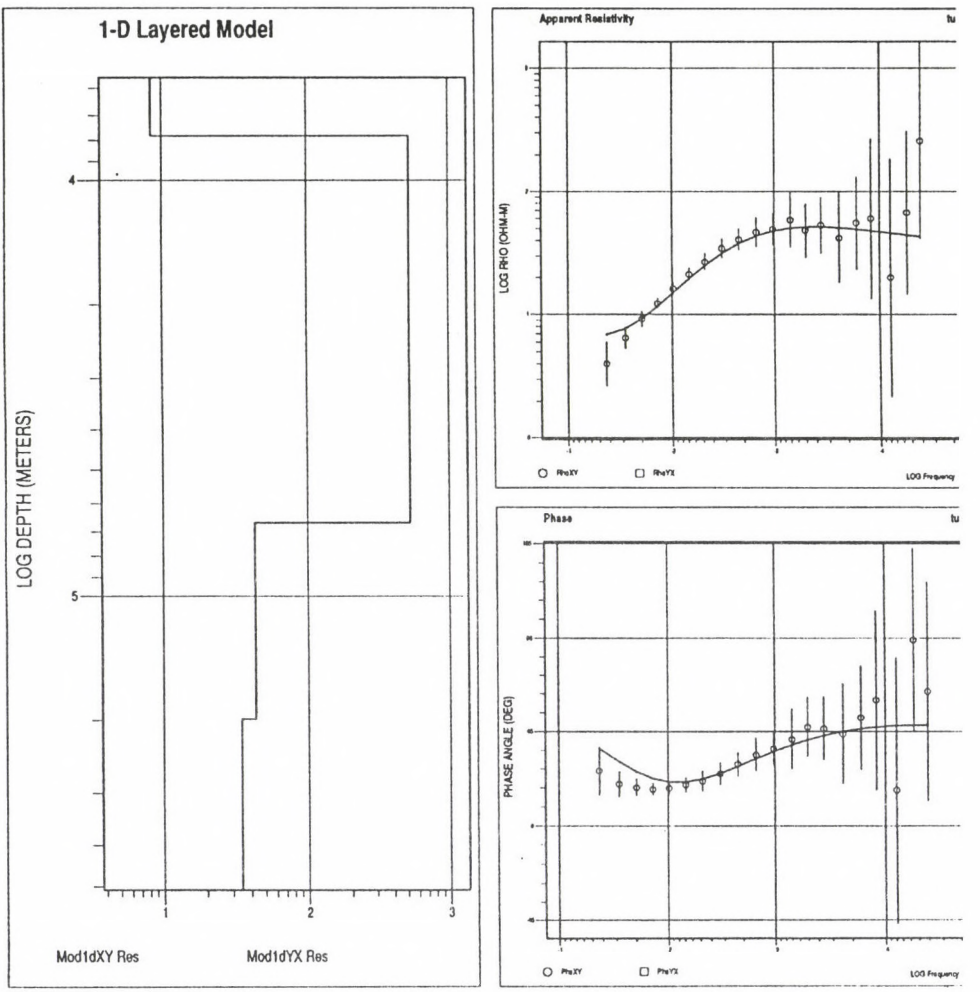


Fig. 8. Rhomax and its phase sounding curves, together with their 1D inversion results in Túrkeve (Pannonan Basin)

asthenospheric depths seem to be larger than those calculated from the geometric mean. In the area north of the TTL, i.e. in the EEP, the Rhomax resistivity both in the whole lithosphere and in the conductive basement has too high values, therefore the partial melting, and hence the existence of an "asthenospheric layer" becomes questionable. (At present stage of the interpretation there is no explanation for the shallower depth of the "conducting basement" in the MT site Rymachy.)

As the next step of the interpretation, different 2D inversions have been carried out partly by using RRI (Smith and Booker 1991), partly by using OCCAM (Constable et al. 1987).

These 2D inversions are obviously rough approximations as the criteria of the 2D structures could not be checked in case of such large distances between the

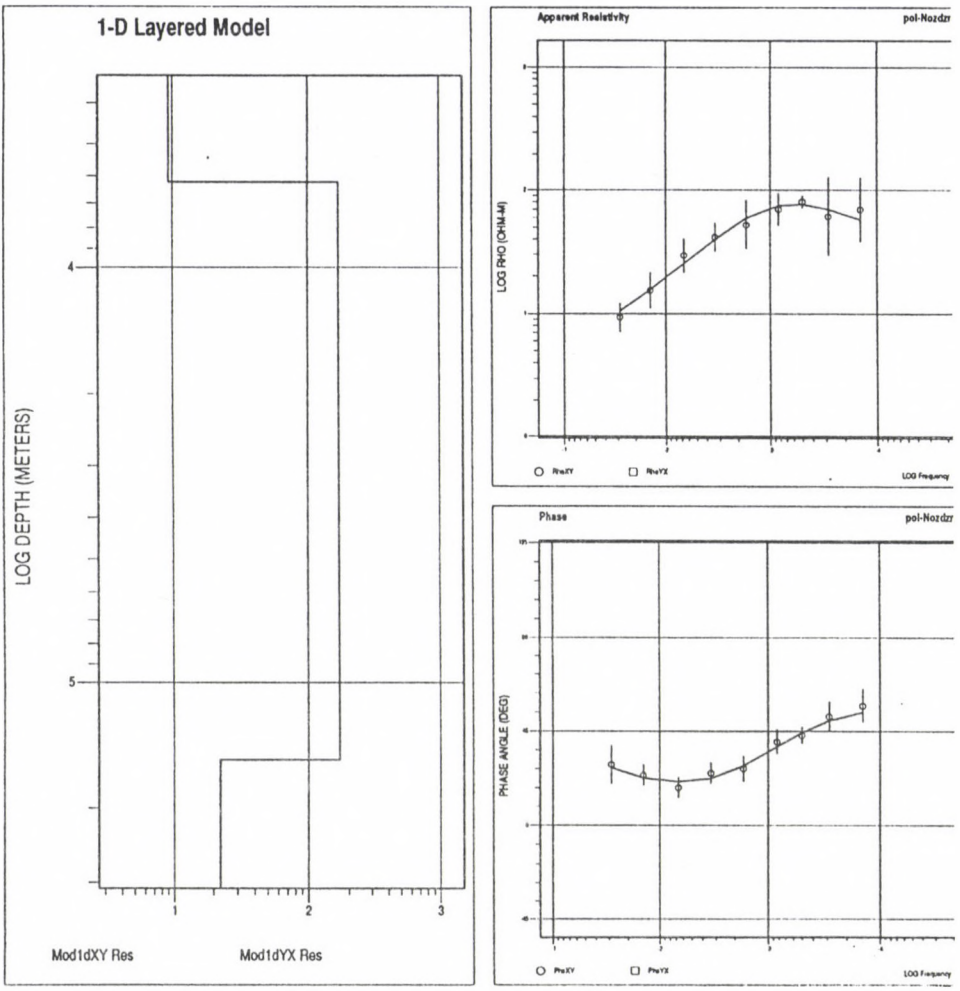


Fig. 9. Rhomax and its phase sounding curves, together with their 1D inversion results in Nozdrzec (Paleozoic platform)

measuring sites lying on very different great tectonic units. The data gap between the Pannonian Basin and the platforms due to electric noise in the Carpathians as a hindering factor should also be mentioned. Without the elimination of the galvanic distortions, the 2D inversion may also lead to very heterogeneous, uninterpretable models. All versions of the different 2D inversions (inversion of Rhomin (yx) alone as TE (E) polarization, and inversion of Rhomax (xy) alone as TM (B) polarization, and their combination) indicate the relative shallowness of the mantle conductive basement (asthenosphere) in the Pannonian Basin and its low resistivity. The high resistivity values of the Precambrian platform can also be seen in the inversion results. Nevertheless, essentially new information to the earlier mentioned one was

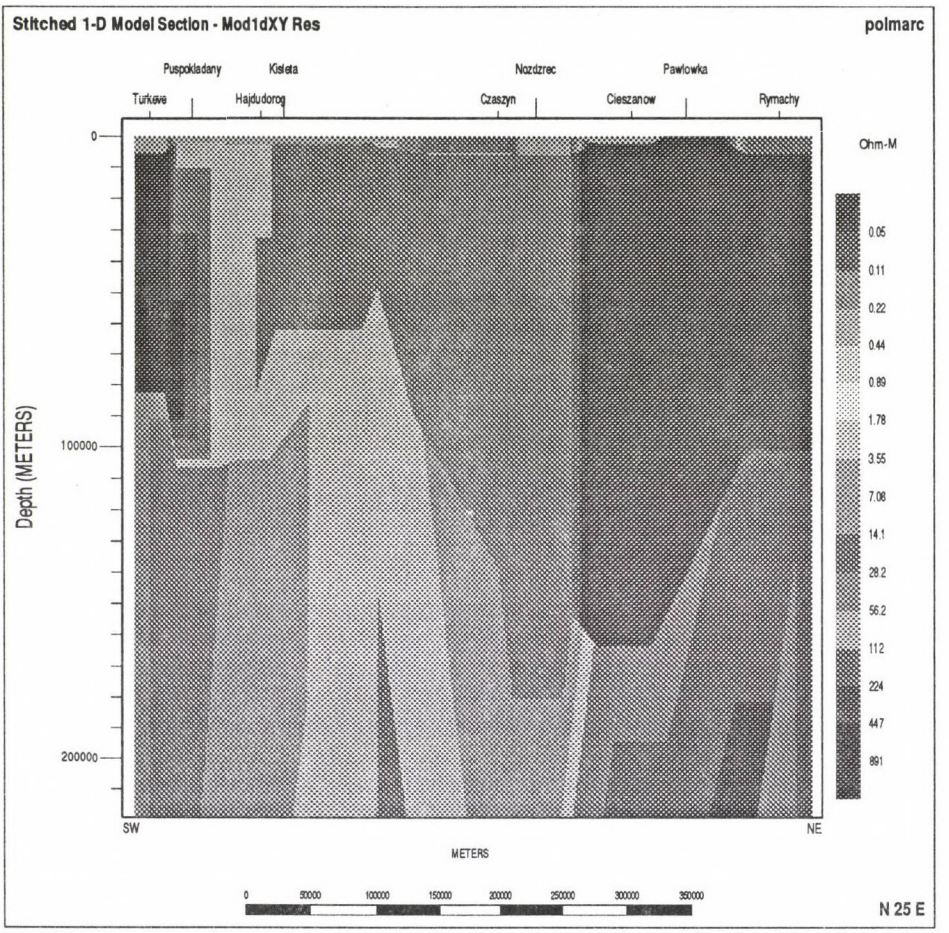


Fig. 10. Stiched 1D model section for Rhomax curves (without Nagyhalász and Porubka)

not supplied by these models. (The final 2D inversion has been made by "trial and error" technique (Fig. 11).)

Conclusions

The main conclusions of the MT and MV measurements along the profile PREPAN95 crossing very different great tectonic units in East Central Europe can be summarized as follows:

— The well-known Carpathian conductivity anomaly was indicated by the magnetovariational measurements at a depth range 8–17 km with a resistivity of 1 Ωm. The shape of the anomaly cannot be exactly determined due to the great distances between the measuring sites. There is no new information concerning its physico-chemical cause (see Jankowski et al. 1991).

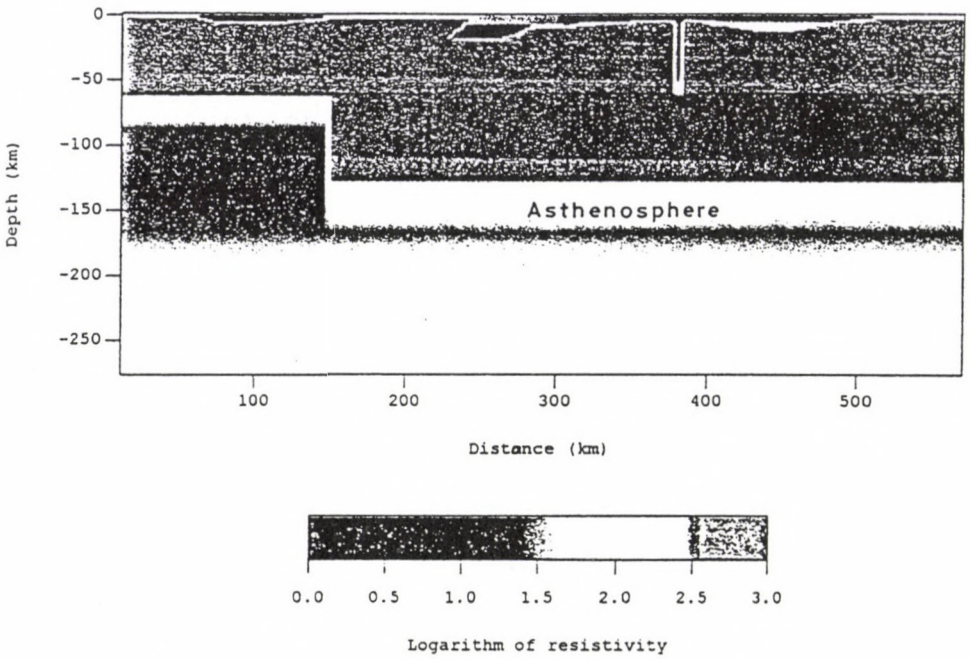


Fig. 11. 2D inversion profile for the deep structure (trial and error)

— In the Neogene Pannonian Basin the conducting basement in the mantle (asthenosphere) is indicated by the magnetotelluric deep soundings at a mean depth of about 65 km and in the area of the Paleozoic and Precambrian platforms at a depth of about 120 km. These depth values are in accordance with the regional heat flow values of the areas under study, i.e. with heat flow of 50–60 mW/m^2 in Poland and 80–100 mW/m^2 in Hungary (centre of the Pannonian Basin). They confirm the earlier MT and MV results (Ádám 1965, Semenov et al. 1993, 1996, 1997).

The weakness of this profile is the magnetotelluric gap in the Carpathians due to the noise caused by DC powered railways currents, therefore the crustal model could be based only on the magnetovariational data.

Certainly the use of remote reference and noise filtering techniques could help in the reliable measurements of the magnetotelluric fields in the Slovak area of the Carpathians to complete this interesting profile. This could be the next step of this co-operation.

Acknowledgement

The authors are indebted to Dr. V Semenov, Mrs J Szendrői and Mr J Rabi for helping the data processing and thank for the Polish Grant (KBN N° 6 P20104) and for the Hungarian National Research Fund (OTKA) (N° T014882).

Appendix

The relation between Berdichevsky's determinant (d) and the geometric mean value of the depths of the highly conductive basement calculated by the extreme MT sounding curves (h_{xy}, h_{yx}):

Berdichevsky's determinant is as the follows:

$$d = \sqrt{Z_{xx}Z_{yy} - Z_{xy}Z_{yx}}$$

if Z_{xx} and Z_{yy} are small, they can be neglegted, and the product of the absolute values of the main impedance elements is the following:

$$d' = \sqrt{|Z_{xy}||Z_{yx}|}\,.$$

The depth of the highly conductive basement:

$$h = \frac{1}{8}\sqrt{10\varrho T} = \frac{1}{8}\sqrt{5T^2|Z|^2}$$

as $\varrho = 0.2T|Z|^2$

$$h_{\mathrm{av}} = \sqrt{h_{xy}h_{yx}} = \sqrt{\left(\frac{1}{8}\right)^2\sqrt{5^2T^4|Z_{xy}|^2|Z_{yx}|^2}} = \frac{T}{8}\sqrt{5|Z_{xy}||Z_{yx}|} = c\sqrt{|Z_{xy}||Z_{yx}|}$$

The geometric mean of the extreme depth values is proportional with the determinant (c = constant).

References

Ádám A 1965: *Gerlands Beiträge zur Geophysik*, 74, 20–40.

Ádám A 1978: *Phys. Earth Planet. Int.*, 17, P21–P28.

Ádám A, Major L 1967: *Acta Geod. Geoph. Mont. Hung.*, 2, 211–215.

Ádám A, Verő J 1976: In: Geoelectrical and Geothermal Studies, KAPG Monograph, A Ádám ed., Akadémiai Kiadó, Budapest, 256–263.

Ádám A, Verő J, Wallner Á 1972: *Acta Geod. Geoph. Mont. Hung.*, 7, 251–287.

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

Constable S C, Parker R L, Constable C G 1987: *Geophysics*, 52, 289–300.

Ernst T, Jankowski J, Semenov V Yu, Ádám A, Hvozdara M, Jozviak W, Lefeld J, Pawliszyn J, Szarka L, Wesztergom V 1997: Electromagnetic soundings across the Tatra Mountains. Submitted to *Acta Geophysica Polonica*

Gordienko V V, Zavgorodnyaya O V 1996: *Acta Geophysica Polonica*, XLIV(2), 174–180.

Jankowski J, Marianiuk J, Ruta A, Sucksdorff C, Kivinen M, 1984: *Geophys. Surv.*, 6, 367–380.

Jankowski J, Pawliszyn J, Józwiak W, Ernst T 1991: *Publ. Inst. Geophys. Pol. Acad. Sci.*, A-19(236), 183–213.

Semenov V Yu, Kaikkonen P 1986: On magnetotelluric data analysis with example for the Kuhmo region in Eastern Finland. Geophys. Dept. of the Oulu University, Finland, Report 13, 1–13.

Semenov V Yu, Ernst T, Jozwiak W, Pawlishin J 1993: *Acta Geophysica Polonica*, 41, 375–383.

Semenov V Yu, Jozwiak W, Ernst T 1996: *Annales Geophysicae*, 14, C-187.

Semenov V Yu, Ádám A, Hvozdara M, Wesztergom 1997: Geoelectrical structures of the Earth's mantle in Pannonian Basin. *Acta Geod. Geoph. Hung.* (present issue)

Smith J T, Booker J R 1991: *J. Geoph. Res.*, 96, 3905–3922.

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

Wieladek R, Ernst T 1977: *Publ. Inst. Geophys. Pol. Acad. Sci.*, G-1 (110), 3–12.

Zhdanov M S, Golubev N G, Varentsov Iv M, Abramova I M, Shneer V S, Berdichevsky M N, Zhdanova O N, Gordienko V V, Kulik S N, Bilinsky A I 1986: *Annales Geophysicae*, 4, B, 3, 335–342.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 225–233 (1997)

FIGURE PARAMETERS OF GANYMEDE

M BURŠA[1]

[Manuscript received April 28, 1997]

The second zonal and second sectorial Stokes parameters of Ganymede determined on the basis of Galileo spacecraft orbit dynamics (June and September 1996) are used to determine the triaxial level ellipsoid of Ganymede. The polar and the equatorial flattening are of the same order in magnitude, about 5×10^{-4}, the secular Love and tidal numbers amount 0.80 and they clarify the equilibrium state of Ganymede. The tidal rotational and orbital dynamics and tidal evolution of the Jupiter-Ganymede system differ significantly from those of the Earth-Moon system.

Keywords: Ganymede; Jupiter; satellites; figure parameters

1. Introduction

The Galileo spacecraft orbit dynamics when passing close to Jupiter's satellite Ganymede in June and September 1996 resulted in determining the second zonal $J_2^{(o)}$ and the second sectorial $J_{2,2}$ Stokes parameters of the largest satellite in the Solar System as follows (Anderson et al. 1996)

$$J_2^{(o)} = -(127.44 \pm 0.64) \times 10^{-6}\,, \tag{1}$$

$$J_2^{(o)} = J_{2,2} = (38.18 \pm 0.21) \times 10^{-6}\,. \tag{2}$$

The above values are conventional, not normalized. They make it possible to treat the triaxiality of Ganymede, the figure of which was believed spherical in the pre-Galileo spacecraft era (Davies et al. 1996). It is just the aim of the paper.

2. Dynamic parameters of the Jupiter-Ganymede system

The rotation axes of Jupiter (J) and Ganymede (G) are nearly parallel. The geocentric equatorial coordinates (declination, right ascension) of the north poles of J and G are (Davies et al. 1996)

$$\delta_{\mathrm{J}} = 64.49^\circ + 0.003^\circ T\,, \quad \alpha_{\mathrm{J}} = 268.05^\circ - 0.009^\circ T\,, \tag{3}$$

$$\delta_{\mathrm{G}} = 64.57^\circ + 0.003^\circ T\,, \quad \alpha_{\mathrm{G}} = 268.20^\circ T \tag{4}$$

(epoch 2000 January 1.5, T in Julian centuries).

The inclination i of the orbital plane of Ganymede is small enough (Burns 1986),

$$i = 0.195^\circ\,, \tag{5}$$

[1]Astronomical Institute, Acad. Sci. Czech Republic, Boční II/1401, 141 31 Praha 4, Czech Republic, E-mail: bursa@ig.cas.cz

1217-8977/97/$ 5.00

to be assumed to equal zero in the figure theory, i.e. G assumed to be orbiting in the equatorial plane of J. The eccentricity of the orbit is (Burns 1986) $e < 0.002$ and it can be assumed to be circular in the figure theory. The Jovicentric gravitational constant of Jupiter is (Campbell and Synnott 1985)

$$GM_{\mathrm{J}} = (126\ 686\ 537 \pm 100) \times 10^9 \mathrm{m}^3\mathrm{s}^{-2}\,, \tag{6}$$

and that of Ganymede (Anderson et al. 1996)

$$GM_{\mathrm{G}} = (9\ 886.6 \pm 0.5) \times 10^9 \mathrm{m}^3\mathrm{s}^{-2}\,; \tag{7}$$

G stands for the Newtonian gravitational constant. The distance Δ_{JG} between the mass centers of the bodies in question (Burns 1986)

$$\Delta_{\mathrm{JG}} = 1\ 070 \times 10^6\ \mathrm{m}\,. \tag{8}$$

The Ganymede's angular velocity of rotation ω_{G} being equal to its mean motion n_{G} is (Davies et al. 1996)

$$\omega_{\mathrm{G}} = n_{\mathrm{G}} = 1.016\ 443\ 41 \times 10^{-5} \mathrm{rad\ s}^{-1}\,, \tag{9}$$

and its mean radius (Davies et al. 1996)

$$\bar{R}_{\mathrm{G}} = (2\ 634 \pm 10)\ \mathrm{km}\,. \tag{10}$$

The parameter in the potential of the centrifugal forces of G

$$q = \frac{\omega_{\mathrm{G}}^2 \bar{R}_{\mathrm{G}}^3}{GM_{\mathrm{G}}} = \frac{n_{\mathrm{G}}^2 \bar{R}_{\mathrm{G}}^3}{GM_{\mathrm{G}}} = \frac{G(M_{\mathrm{J}} + M_{\mathrm{G}})}{GM_{\mathrm{G}}} \left(\frac{\bar{R}_{\mathrm{G}}}{\Delta_{\mathrm{JG}}}\right)^3 \tag{11}$$

comes out, if values (7), (9) and (10) are adopted, as

$$q = 1.909\ 707 \times 10^{-4}\,. \tag{12}$$

Parameters (1), (2) and (7) define the external gravitational potential V of G, smoothed in retaining harmonic terms $n = 2$ only, parameter (12) defines its potential of centrifugal forces Q. Its tidal potential V_t due to Jupiter is as follows

$$\begin{aligned} V_t \;=\; & \frac{GM_{\mathrm{J}}}{\Delta_{\mathrm{JG}}} \sum_{n=2}^{\bar{n}} \left(\frac{\varrho}{\Delta_{\mathrm{JG}}}\right)^n \Big[P_n^{(o)}(\sin\delta_{\mathrm{GJ}}) P_n^{(o)}(\sin\phi) + \\ & + 2 \sum_{k=1}^{n} \frac{(n-k)!}{(n+k)!} P_n^{(k)}(\sin\delta_{\mathrm{GJ}}) P_n^{(k)}(\sin\phi) \cos k(\Lambda + T_{\mathrm{GJ}}) \Big]\,; \end{aligned} \tag{13}$$

$P_n^{(k)}$ denotes the associated Legendre function of degree n and order k, ϱ stands for the centric radius-vector of the potential point P on the physical surface of Ganymede, ϕ is the centric latitude of P, δ_{GJ} and T_{GJ} are equatorial coordinates (declination and hour angle respectively) of J centered at the mass center of G. Because of synchronous rotation of G orbiting in the equatorial plane of J, we may

impose $\delta_{GJ} = 0$, as well as, $T_{GJ} = 0$, the primary meridian of G be that containing the mass center of J. Then all odd zonal harmonic terms in (13) equal zero. Zero are also all the even tesseral harmonic terms if order k is odd, as well as, all the odd tesseral harmonic terms if k is even. Because $\varrho/\Delta_{JG} \approx 0.0025$, we limit ourselves to $\bar{n} = 3$ in (13):

$$\begin{aligned} V_t &= \frac{GM_J}{\Delta_{JG}} \left\{ \left(\frac{\varrho}{\Delta_{JG}}\right)^2 \left[-\tfrac{1}{2} P_2^{(o)}(\sin\phi) + \tfrac{1}{4} P_2^{(2)}(\sin\phi)\cos 2\Lambda\right] + \right. \\ &\left. + \left(\frac{\varrho}{\Delta_{JG}}\right)^3 \left[-\tfrac{1}{4} P_3^{(1)}(\sin\phi)\cos\Lambda + \tfrac{1}{24} P_3^{(3)}(\sin\phi)\cos 3\Lambda\right] \right\}. \end{aligned} \tag{14}$$

3. Triaxial level ellipsoid of Ganymede

The sum of potentials V, Q and V_t represent the total gravity potential W at $P(\varrho, \phi, \Lambda)$:

$$\begin{aligned} W &= \frac{GM_G}{\varrho} \left\{ 1 + \tfrac{1}{3}\mu^{-3} q + \left(\mu^2 J_2^{(o)} - \tfrac{5}{6}\mu^{-3} q\right) P_2^{(o)}(\sin\phi) + \right. \\ &+ \left(\mu^2 J_2^{(2)} + \tfrac{1}{4}\mu^{-3} q\right) P_2^{(2)}(\sin\phi)\cos 2\Lambda + \\ &\left. + \frac{\varrho}{\Delta_{JG}} \mu^{-3} q_G \left[-\tfrac{1}{4} P_3^{(1)}(\sin\phi)\cos\Lambda + \tfrac{1}{24} P_3^{(3)}(\sin\phi)\cos 3\Lambda\right] \right\}, \end{aligned} \tag{15}$$

$\mu = \bar{R}_G/\varrho$. We wish to treat the equipotential surface $W =$ constant $= W_o$ representing the boundary surface of G. Let us select it so that

$$W_o = \frac{GM_G}{\bar{R}_G} + \frac{1}{3}\omega_G^2 \bar{R}_G^2 = \frac{GM_G}{\bar{R}_G}\left(1 + \frac{1}{3} q\right), \tag{16}$$

numerically

$$W_o = 3\ 753\ 694\ \mathrm{m^2 s^{-2}}. \tag{17}$$

This value is considered as fundamental constant in the figure solution.

Having selected the equipotential surface we can treat its figure. We wish to represent it by the best fitting triaxial level ellipsoid E($a > b > c$) centered at the mass center of G, with the largest axis pointing toward the mass center of Jupiter. Its semi-axes and flattenings $\alpha = (a - b)/a$, $\alpha_1 = (a - c)/a$ are defined by five adopted constants: GM_G, W_o, ω_G, $J_2^{(o)}$, $J_{2,2}$, values (7), (17), (9), (1) and (2).

Gravity potential on the surface of E is constant, it is equal to the potential at the basic equipotential surface $W = W_o$. After inserting $\varrho = \varrho_E$ (radius vector of surface E) and $W = W_o$ into Eq. (15), and assuming the total gravity potential be constant on E, one gets, after some tremendous algebra,

$$\begin{aligned} a &= \frac{GM_G}{W_o} \left[1 - \nu^2 J_2^{(o)} + 3\nu^2 J_2^{(2)} + \tfrac{3}{2}\nu^{-3} q + \tfrac{7}{40}\nu^4 (J_2^{(o)})^2 - \right. \\ &- \tfrac{3}{2}\nu^4 J_2^{(o)} J_2^{(2)} \tfrac{1033}{70}\nu^4 (J_2^{(2)})^2 - \tfrac{13}{14}\nu^{-1} J_2^{(o)} q + \tfrac{457}{35}\nu^{-1} J_2^{(2)} q + \\ &+ \tfrac{461}{140}\nu^{-6} q^2 \tfrac{507}{560} F_1^3 + \tfrac{815}{2352}\nu^{-9} q^3 - \tfrac{5279}{3920}\nu^{-6} q^2 F_1 + \\ &\left. + \tfrac{9003}{3920}\nu^{-3} q F_1^2 - \tfrac{301}{392}\nu^{-6} q^2 F_2 + \tfrac{2817}{1960}\nu^{-3} q F_1 F_2 + \tfrac{81}{140} F_1^2 F_2 \right], \end{aligned} \tag{18}$$

$$\begin{aligned}\alpha &= -\tfrac{3}{2}\nu^2 J_2^{(o)} + 3\nu^2 J_2^{(2)} + 2\nu^{-3}q - \tfrac{3}{8}\nu^4 (J_2^{(o)})^2 + \tfrac{69}{28}\nu^{-1} J_2^{(o)} q + \\ &+ \tfrac{9}{2}\nu^4 J_2^{(o)} J_2^{(2)} + \tfrac{15}{14}\nu^4 (J_2^{(2)})^2 + \tfrac{3}{14}\nu^{-1} J_2^{(2)} q - \tfrac{1}{28}\nu^{-6}q^2 - \tfrac{63}{80}F_1^3 + \\ &+ \tfrac{1641}{3920}\nu^{-9}q^3 - \tfrac{8949}{3920}\nu^{-6}q^2 F_1 + \tfrac{13107}{3920}\nu^{-3} q F_1^2 + \tfrac{385}{392}\nu^{-6}q^2 F_2 + \\ &+ \tfrac{3159}{392}\nu^{-3} q F_1 F_2 - \tfrac{45}{7}F_1^2 F_2,\end{aligned} \tag{19}$$

$$\begin{aligned}\alpha_1 &= 6\nu^2 J_2^{(2)} + \tfrac{3}{2}\nu^{-3}q + \tfrac{15}{7}\nu^{-1} J_2^{(2)} q - \tfrac{66}{7}\nu^4 (J_2^{(2)})^2 + \\ &+ \tfrac{9}{8}\nu^{-6}q^2 + \tfrac{364}{49}\nu^{-6}q^2 F_2 + \tfrac{270}{49}\nu^{-3} q F_1 F_2 - \tfrac{90}{49}F_1^2 F_2.\end{aligned} \tag{20}$$

$$\begin{aligned}F_1 &= \nu^2 J_2(o) - \tfrac{1}{2}\nu^{-3}q\,, \\ F_2 &= \nu^2 J_2(o) + \tfrac{1}{4}\nu^{-3}q\,,\end{aligned} \tag{21}$$

$$\nu = \bar{R}_{\mathrm{G}} \frac{W_o}{GM_{\mathrm{G}}}\,. \tag{22}$$

Terms containing q^3, $(J_2^{(o)})^3$ etc. have been retained, however, $n = 3$ tidal terms in (14) neglected.

Numerically,

$$a = 2\ 635\ 057\ \mathrm{m}\,, \tag{23}$$

$$1/\alpha = 1\ 454.5\,, \tag{24}$$

$$1/\alpha_1 = 1\ 939.7\,, \tag{25}$$

and/or

$$b = 2\ 633\ 699\ m\,, \tag{26}$$

$$c = 2\ 633\ 245\ m\,. \tag{27}$$

Consequently, volume τ defined by equipotential surface $W = W_o$, value (17), comes out as

$$\tau = 7.655 \times 10^{19}\mathrm{m}^3 \tag{28}$$

and, in view of (7), the mean density $\bar{\sigma}$ of Ganymede

$$\bar{\sigma} = 1\ 936\ \mathrm{kg\ m}^{-3}, \tag{29}$$

value identical to that by Anderson et al. (1996).

Because of rotational/orbital resonance of Ganymede, we may assume the tidal and rotational distortions due to the tidal friction be responsible for its actual figure during the tidal evolution of Jupiter/Ganymede system. The tidal and rotational distortion $\delta J_2^{(o)}$ and tidal distortion $\delta J_{2,2}$ in the Stokes parameters $J_2^{(o)}$ and $J_{2,2}$ are as

$$\delta J_2^{(o)} = -\frac{1}{3}k_s q - \frac{1}{2}k_t \frac{GM_{\mathrm{J}}}{GM_{\mathrm{G}}}\left(\frac{\bar{R}_{\mathrm{G}}}{\Delta_{\mathrm{JG}}}\right)^3, \tag{30}$$

$$\delta J_{2,2} = \frac{1}{4}k_t \frac{GM_{\mathrm{J}}}{GM_{\mathrm{G}}}\left(\frac{\bar{R}_{\mathrm{G}}}{\Delta_{\mathrm{JG}}}\right)^3, \tag{31}$$

k_s and k_t are secular Love numbers equal 1.5 if homogeneous body and zero if the mass is concentrated at the center. Let the conditions be imposed as

$$\delta J_2^{(o)} = J_2^{(o)}, \quad \delta J_{2,2} = J_{2,2}, \tag{32}$$

i.e. let us assume the tidal and rotational distortions be fully responsible for the actual figure of Ganymede. Then one gets

$$k_s = -3\frac{J_2^{(o)} + 2J_{2,2}}{q} = 0.80, \tag{33}$$

$$k_t = 4\frac{J_{2,2}}{q} = 0.80\,. \tag{34}$$

It means, if (32), the mass concentration toward the center discovered by Anderson et al. (1996) is evident. The same conclusion can be drawn on the basis of the coefficient of inertia (C is the maximum principal moment of inertia)

$$\begin{aligned} \frac{C}{M_G R_G^2} &= \frac{2}{3} - \frac{4}{15}\left[\frac{4-k_s}{1+k_s}\right]^{1/2} = \\ &= \frac{2}{3} - \frac{4}{15}\left[\frac{4q+3(J_2^{(o)}+2J_{2,2})}{q-3(J_2^{(o)}+2J_{2,2})}\right]^{1/2} = 0.311. \end{aligned} \tag{35}$$

This value is practically the same as derived by Anderson et al. (1996).

Note that the ratious

$$\alpha/q = 3.60, \quad \alpha_1/q = 2.70, \tag{36}$$

as well as,

$$-J_2^{(o)}/\alpha = 0.19 \tag{37}$$

correspond well to the equilibrium figure (Burša 1994).

4. Tidal variation within Jupiter-Ganymede system due to tidal friction

The long-term tidal variation in Δ_{JG} can be derived from the Lagrangian planetary equation for the time variation in the semimajor axis of the orbit. The perturbing function in that case is the Jupiter-Ganymede tidal force function ΔV_{JG} which reads, in view of the above $\delta_{GJ} = 0$, $T_{GJ} = 0$, as well as, $\delta_{JG} = 0$ and $T_{JG} = 0$ (δ_{JG} stands for jovicentric declination, T_{JG} for jovicentric hour angle of the center of mass of Ganymede), as follows

$$\begin{aligned} \Delta V_{JG} &= k_J \frac{GM_1^2}{\Delta_{JG}}\left(\frac{\bar{R}_J}{\Delta_{JG}}\right)^5 \left(\tfrac{1}{4} + \tfrac{3}{4}\cos 2\varepsilon_J\right) + \\ &+ k_G \frac{GM_G^2}{\Delta_{JG}}\left(\frac{\bar{R}_G}{\Delta_{JG}}\right)^5 \left(\tfrac{1}{4} + \tfrac{3}{4}\cos 2\varepsilon_G\right), \end{aligned} \tag{38}$$

$$\bar{R}_J = 69\ 911 \text{ km} \tag{39}$$

is the mean radius of Jupiter (Davies et al. 1996), k_J and k_G stand for the Love number for Jupiter and Ganymede respectively, ε_J and ε_G for their phase lag angles. Then the Lagrangian equation gives (Kopal 1978)

$$\begin{aligned} \frac{d\Delta_{JG}}{dt} &= -2\Delta_{JG}^{1/2}\,[G(M_J+M_G)]^{-1/2}\,\frac{M_J+M_G}{M_JM_G}\left[\frac{\partial\Delta V_{JG}}{\partial\varepsilon_J}+\frac{\partial\Delta V_{JG}}{\partial\varepsilon_G}\right] = \\ &= 6\left[\frac{G(M_J+M_G)}{\Delta_{JG}}\right]^{1/2}\left[\frac{GM_G}{GM_J}\left(\frac{\bar{R}_J}{\Delta_{JG}}\right)^5(k_J\varepsilon_J)+\right. \\ &\quad \left.+\frac{GM_J}{GM_G}\left(\frac{\bar{R}_G}{\Delta_{JG}}\right)^5(k_G\varepsilon_G)\right]. \end{aligned} \tag{40}$$

The estimate for the Love number of Jupiter by Yoder (1979) is

$$k_J = 0.5\,, \tag{41}$$

for its phase lag angle by Greenberg (1982)

$$\varepsilon_J = 2.6\times10^{-7}\,, \tag{42}$$

and, if so,

$$k_J\varepsilon_J = 1.3\times10^{-7}\,. \tag{43}$$

As regards Ganymede,

$$k_G\varepsilon_G = 0 \tag{44}$$

at present, because of its rotational/orbital resonance. However, $k_G\varepsilon_G \neq 0$ should be assumed in the past and it should be taken into account in treating the tidal evolution of Jupiter-Ganymede system. Even if a rough estimate, a dynamical model should be adopted. Let the mass, the mean radius and the principal moment of inertia be constant during the tidal evolution, values (7), (10), and (35), or

$$C = 3.20\times10^{35}\ \mathrm{kg\ m^2}\,. \tag{45}$$

Let the angular velocity of rotation $\omega_G(t_o)$ of primeval Ganymede be at Epoch t_o

$$\omega_G(t_o) = 2.181\,662\times10^{-4}\ \mathrm{s^{-1}} \tag{46}$$

i.e. the period of its rotation 8h. Then the primeval spin angular momentum $L_G(t_o)$ comes out as

$$L_G(t_o) = C\omega_G(t_o) = 6.98\times10^{31}\ \mathrm{kg\ m^2 s^{-1}} \tag{47}$$

and it can be compared to its present value

$$L_G(t) = 3.25\times10^{30}\ \mathrm{kg\ m^2 s^{-2}}\,, \tag{48}$$

the decrease comes out as

$$\Delta L_G = -6.65\times10^{31}\ \mathrm{kg\ m^2 s^{-1}}\,. \tag{49}$$

It equals the tidal increase ΔL_{JG} in the orbital angular momentum L_{JG} of the Jupiter-Ganymede system

$$L_{JG} = \frac{M_J M_G}{M_J + M_G} [G(M_J + M_G)\Delta_{JG}]^{1/2} , \tag{50}$$

$$\Delta L_{JG} = -\Delta L_G . \tag{51}$$

Then the tidal increase $\delta\Delta_{JG}$ in the Jupiter-Ganymede system can be estimated

$$\delta\Delta_{JG} = 2\Delta L_{JG}\, \Delta_{JG}^{1/2} \frac{[G(M_J + M_G)]^{1/2}}{GM_J M_G} = 82.5 \text{ km} , \tag{52}$$

and the decrease in the mechanical energy δE_{JG} of the system due to the tidal friction

$$\begin{aligned} \delta E_{JG} &= \tfrac{1}{2} C\, [\omega_G^2 - \omega_G^2(t_o)] - \tfrac{1}{2}\, GM_J M_G \left[\tfrac{1}{\Delta_{JG}} - \tfrac{1}{\Delta_{JG}(t_o)}\right] = \\ &= -6.92 \times 10^{27} \text{ kg m}^2\text{s}^{-2} , \end{aligned} \tag{53}$$

$$\Delta_{JG}(t_o) = \Delta_{JG} - \delta\Delta_{JG} . \tag{54}$$

Assuming tidal evolution time, e.g., as $t - t_o = 1.5 \times 10^9$y, the integral mean value of $(k_G \varepsilon_G)$ can be estimated as

$$\bar{k}_G \bar{\varepsilon}_G = \frac{1}{3} \frac{\Delta L_G \Delta_{JG}^6}{GM_J^2 \bar{R}_G^5 (t - t_o)} = 2.30 \times 10^{-8} . \tag{55}$$

The long-term variation in the J-G distance (40), estimated with the use of present $(k_J \varepsilon_J)$ value (43) and the integral mean hypothetical value (55), comes out as

$$\frac{d\Delta_{JG}}{dt} = 0.003\text{m cy}^{-1} . \tag{56}$$

5. Discussion. Ganymede-Moon comparison

In view of values (33)–(37), the conclusion can be drawn about the hydrostatic equilibrium of Ganymede at present (Anderson et al. 1996). Et least, there is no contradiction among its figure and orbital/rotational parameters from the equilibrium point of view.

A comparison with the Moon suggests itself, because of the orbital/rotational resonance, i.e. the similarity of the rotational/orbital dynamics, and because of the similarity in the size. The second degree lunar Stokes coefficients are (McCarthy 1992)

$$(J_2^{(o)})_{\mathbb{C}} = -202.151 \times 10^{-6} , \tag{57}$$

$$(J_{2,2})_{\mathbb{C}} = 22.302 \times 10^{-6} , \tag{58}$$

the selenocentric gravitational constant

$$GM_{\mathbb{C}} = 4902.7989 \times 10^9 \text{m}^3\text{s}^{-2} , \tag{59}$$

and (Davies et al. 1996)

$$\omega_{\mathbb{C}} = n_{\mathbb{C}} = 2.661\,696\,96 \times 10^{-6}\text{rad s}^{-1}\,, \tag{60}$$

$$\bar{R}_{\mathbb{C}} = 1\,737\,400\text{ m}\,, \tag{61}$$

$$q_{\mathbb{C}} = \frac{\omega^2_{\mathbb{C}} R^3_{\mathbb{C}}}{GM_{\mathbb{C}}} = 7.578\,314 \times 10^{-6}\text{rad s}^{-1}\,; \tag{62}$$

the secular Love numbers of the Moon come out absurd as

$$(k_s)_{\mathbb{C}} = 62.4\,, \tag{63}$$

$$(k_t)_{\mathbb{C}} = 11.8\,. \tag{64}$$

It only supports the well-known fact that the present state of the Moon's body is far from the ideal hydrostatic equilibrium, e.g., the observed ratio $(C - A)/B$ ($C > B > A$) are principal moments of inertia of the Moon) turns out to be 17 times as large as the hydrostatic one (Kopal 1966). If the tidal forces were assumed to be fully responsible for the second sectorial Stokes coefficient (58), then the Earth-Moon distance should be much smaller than its present value 384 400 km, about 192 700 km, and its period should amount 9.74 days. If both, tidal, as well as, rotational distortions were fully responsible for both second order Stokes coefficients, values (57) and (58), than the period comes out 5.5 days at the Earth-Moon distance 131 500 km, i.e. at 20.6 Earth's radii. This relative "equilibrium distance" of the Moon is close to that of Ganymede at present, which amounts 15.3 Jupiter's radii. Also the actual rotational-orbital period of Ganymede, which amounts 7.15 days, is close to the Moon's equilibrium period above.

However, the Earth-Moon system and the Jupiter-Ganymede system differ significantly as regards their tidal friction dynamics: $(k_{\mathrm{J}}\ \varepsilon_{\mathrm{J}})$, Eq. (43), compared to that of the Earth $(k_{\oplus}\ \varepsilon_{\oplus}) \sim 0.013$ is five orders smaller in magnitude.

6. Conclusions

1. The polar and the equatorial flattenings of Ganymede are of the same order in magnitude: 5×10^{-4}.

2. The secular Love numbers of Ganymede, due to the rotational (k_{s}) and tidal (k_{t}) distortions, come out as $k_{\mathrm{s}} = k_{\mathrm{t}} = 0.80$ and support the static equilibrium state of Ganymede.

3. The rotational and tidal distortions can be considered as being fully responsible for the second zonal and the second sectorial Stokes parameters in the gravitational potential, values (1) and (2).

4. The actual figure parameters of Ganymede are close to those defining the equlibrium figure.

5. There is a large difference between the Jupiter-Ganymede and the Earth-Moon orbital/rotational tidal dynamics, even if similarities in some dynamical parameters.

6. The product of the Love number and the phase lag angle of Jupiter amounts only 10^{-5} in relation to that of the Earth.

7. The increase of the Ganymede-Jupiter distance during the tidal evolution of the system due to the tidal friction can be estimated about 82.5 km only.

References

Anderson J D, Lau E L, Sjogren W L, Schubert G, Moore W B 1996: *Nature*, 384, 541–543.

Burns J A 1986: In: Satellites, J A Burns, M S Mathews eds, The University of Arizona Press, Tucson, 1–38.

Burša M 1994: *Studia geoph. et geod.*, 38, 7–22.

Campbell J K, Synnott S P 1985: *Astron J.*, 90, 364–372.

Davies M E, Abalakin V K, Burša M, Lieske J H, Morando B, Morrison D, Seidelmann P K, Sinclair A T, Yallop B, Tjuflin Y S 1996: *Celestial Mechanics and Dynamical Astronomy*, 63, 127–148.

Greenberg R 1982: In: Satellites of Jupiter, D Morrison ed., The University of Arizona Press, Tucson, Arizona, 65.

Kopal Z 1966: An introduction to the study of the Moon. D Reidel Publishing Company, Dordrecht

Kopal Z 1978: Dynamics of close binary systems. D Reidel Publishing Company, Dordrecht

McCarthy D D ed. 1992: IERS Standards. Observatoire de Paris

Yoder C F 1979: *Nature*, 279, 767.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 235–243 (1997)

MAXIMILIAN HELL: AN AURORAL STUDY IN THE 18TH CENTURY*

J VERŐ[1]

[Manuscript received May 26, 1997]

In addition to a biography of M Hell and of his companion, J Sajnovics, the background of his youth, the Selmec Academy and its connection with the Jesuit order is shortly characterized. The circumstances and events of the Northern expedition and the best known result, the discovery of the Finnish-Hungarian linguistic relationship by Sajnovics are mentioned, then Hell's views on auroras before and after the expedition are summarized, together with the description of some observations made during the voyage and in Vardö in connection with auroras and magnetism are discussed.

Keywords: aurora; Hell, Maximilian; history of geophysics

Maximilian Hell was born May 15, 1720 from a family of Bavarian origin which arrived in the then Upper Hungarian mining region centered at Selmecbánya (Hungarian), or Schemnitz (German), currently Banska Stiavnica (in Slovakia) via Bohemia in 1694. It was here that several family members were prominent mining engineers. The first important family member was Maximilian Hell's father, Matthäus Cornelius Hell (or Höll, 1650–1743), a mining counselor. His elder brother Joseph Karl Hell (1713 - past 1789) invented a very ingenious water-raising machine based on his father's concept, and also a pneumatic machine and several other mining tools. Though the family was of humble background, the descendants were recognized by the receipt of a letter of nobility in 1792.

Selmec was then an important center for both metal production, mining and mining education. As one of the first technical institutes of higher education in the world, the Royal Mining Institute was established there in 1735 by King Charles III (Charles VI, as German-Roman emperor). It was hence promoted to Royal Mining Academy by Queen and Empress Maria Theresia in 1770. Among the most prominent professors at the mining academy were Sámuel Mikoviny, Nicolas J Jacquin and Ignaz von Born. Mikoviny was an excellent mathematician and cartographer, Jacquin a French-Dutch botanist, later director of the Vienna Botanical Gardens; and Born, as a leading figure in Hungarian mining, organized the world's first Mining and Metallurgy Congress in Selmec. It was at this congress that 39 scientists from 13 countries studied for several months the new amalgamation method for gold production invented by Born. He was later depicted as Sarastro in Mozart's Magic Flute (Lux, 1988 referring to the Gotha Almanach, 1791, Necrology of Ignaz von Born and Wolfgang Amadeus Mozart). Other "firsts" in Selmec included the subsurface use of explosives for mining (1627), steam engine utilization for mining

*Presented at the EGS XXII. General Assembly in Vienna, April 21-25, 1997

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

1217-8977/97/$ 5.00

(1712, from Newcomer at the suggestion of Leibniz), the first international scientific society. Among the society's members were the eminent Antoine Lavoisier, James Watt and Johann Wolfgang von Goethe.

Miklós Póda became a Selmec professor of mathematics and mining engineering at the invitation of Born. He was a Jesuit (Society of Jesus) of Hungarian ancestry who lived in Graz. Other Jesuits instructed students in Latin. They played significant educational roles at that time in all the country in natural sciences education. Pál Kerekgedei Makó, a colleague of Hell (1723–1793), also an excellent mathematician and physicist was a professor at the universities Vienna and Buda - Pest. He had also published several works on auroras and tried to determine the height of this phenomenon. The artist Johann Nepomuk Fierer made several water-colors of the sky at Trnava (1768–1778), some of which included the aurora (later published by Réthly and Berkes, 1963). The Jesuit order sustained a number of astronomical observatories during the second half of the 18th century at schools in Hungary.

It is not surprising that in this atmosphere M Hell also joined in 1738 the Jesuit order, as he was strongly attracted toward mathematics and the natural sciences. Upon completion of his education he taught physics, mathematics and theology at several Jesuit colleges in Hungary, where he also built astronomical observatories. In 1755 he was invited to be the director of the Imperial Astronomical Observatory in Vienna. Soon after he began the publication of Ephemerides astronomicae, a series that continued from 1757 till 1786. In these yearbooks he also published several scientific papers. He considered this work as very important since he was able to prepare the manuscripts in advance while another Jesuit substituted him during his long absence from Vienna. In 1763 he published *"Anleitung zum nützlichen Gebrauch der Stahlmagnete"* (Introduction to the Useful Application of Steel Magnets).

Between 1768 and 1770 he embarked on the great expedition to Norway and Denmark, to observe the transition of the Venus before the solar disk. Details of this expedition will be covered later. After this expeditionary voyage he presented a series of lectures in Copenhagen on the aurora after which he was honored by being nominated to membership on the Royal Society and was bestowed with a doctorate from the university, becoming the first foreign honoree so rewarded. Upon his return to Vienna he continued the Ephemerides, then published the work De parallaxi Solis. An important work was his plan for an Academy of the Habsburg empire, consisting of sections Mechanics, Physics, Botany, Anatomy and Chemistry, as "only these disciplines can yield new discoveries and observations in science, even if Diplomacy, Archeology and Numismatics may discover some ancient things, too, but such discoveries must not be included into science proper". During his latter years he spent his life in solitude, remaining alone in the observatory. It was during this period that the Jesuit order was suppressed. He fulfilled his duties till the very last days. Even his death on April 14, 1793 was related to a visit at the observatory by a foreign dignitary. Accompanying this guest during cold winter weather he developed pneumonia and shortly thereafter succumbed.

Another participant to the Northern expedition was János Sajnovics, also a Jesuit, mathematician and astronomer. Sajnovics, born on May 12, 1733, was a

colleague of M Hell who selected him for this expedition. Sajnovics kept a rather detailed diary during the expedition which was partly published by Littrow in 1835. Sajnovics was counseled by M Hell to observe the Lapponian (partly Finnish) population and language, since it is similar to Hungarian. At this admonition he prepared the work "Demonstratio idioma Ungarorum et Lapporum idem esse", published in Copenhagen in 1770. This work is considered by linguistic scholars to be one of the firsts of its kind in comparative linguistics. After suppression of the Jesuit order, he became professor of astronomy at Buda university. He published "Idea astronomiae honoribus regiae universitatis Budensis dicata", printed in 1778 in Buda. He died early on March 1, 1785 in Buda.

The best known result of the northern expedition was the Finnish-Hungarian linguistic relationship. In a questionnaire when entering the Jesuit order, M. Hell provided information as to his fluency in the German, Latin and Slavic (most likely Slovak) languages. Nonetheless, he considered himself Hungarian, a term referring to citizenry of the Holy Hungarian Empire, i.e. of Hungary, without nationality distinction. Thus it is somewhat curious that he noticed the similarity of the Lapponian and the Hungarian languages.

This point is clarified in a letter dated April 6, 1769 from Vardö to Pater Höller, confessor of the Vienna Court: "It is to be kept a secret till we return home that we discovered the Lapponian nation, which is spread to the east in all Asia. O God, what a surprise, who could believe that we found in these Lapponians a people who were born from the same fathers as us! They are Hungarians as our brothers, they speak our Hungarian language, they are clad like us Hungarians, their morals and life are the same as of our ancient predecessors. (*Ungari sunt, fratres nostri sunt, linguam loquuntur nostram Hungaricam, veste utuntur nostra Ungarica, mores vitamque vivunt patrum nostrorum antiquorum, verbo, fratres nostri sunt*). O what miraculous, what stupefying thing I have to report to our General Father! (the head of the Jesuit order). But we have to keep secrecy, the Danish king has to know it before all others!" Sajnovics also mentioned that M Hell advised him to observe the "Lapponian" language as it was similar to the Hungarian. Sajnovics learned it quickly so that when they left Norway, he could speak with the local people. It is most interesting that M Hell would later explicitly refer to the common belief of northern people about the aurora. He was pleased with his discovery and also with Sajnovics' zeal in collecting "Lapponian" words. He wrote to Pater Pilgram, his substitute: "I am very content that I chose this companion, as he learned the Lapponian language in such a short time so perfectly. I asked him to include some remarkable things about them in our report. I gave him some criteria and regulations how to make this study and now we are sure nobody will hesitate to accept this our result."

The following provides a description of the expedition, its origin and related circumstances. The Danish king Christian VII (1749–1808) reigning from 1766 was certainly not the ideal nor the best ruler of his country. During the 1770s his administration was led by his minion, Struensee. Family problems led to a palace revolution that eventually his son, Fredrik VI succeeded him on the throne in 1784. However, during Christian VII's early years when M Hell visited, he was in support

of scientific projects, mainly influenced by his main counsel, Count Johann Hartwig Bernstorff (1712–1772).

Bernstorff, a German, served Denmark as diplomat in Saxony, the Regensburg Imperial Diet, and Paris. As minister of foreign affairs (1750), he kept Denmark neutral during the Seven Years' war. He also concluded a pact with Russia, supported industry and commerce, particularly in the West Indies and Mediterranean areas, and abolished serfdom. He provided support and stipends to arts and science such as to Niebuhr's eastern archeological expedition and to the poet Klopstock. He died just before returning from Hamburg at the time of his successor, Struensee's fall and execution.

In a letter dated August 18, 1767 Bernstorff instructed Ludwig Heinrich Bachoff (1725–1792), the Danish ambassador to the Vienna Court, to ask M. Hell whether he would be ready to participate in an expedition to Vardö to observe the transition of Venus, at the time of interior conjunction. Bachoff was a scientist and a poet who had for example published 14 religious songs in 1774. Bachoff quickly reached an agreement with M Hell. M Hell agreed to lead the expedition provided the Vienna Court and his religious superiors gave permission. He was willing to take on the task without pay, on condition that supplies and instruments were provided. On December 10, 1767 preliminary permissions were granted. Thence he strove to complete the next two volumes of Ephemerides so that the mission could start soon after Easter in 1768. February 24 Queen Maria Theresia signed the letter of commission. Accompanied by Sajnovics he was in audience with Emperor Charles VI and the Queen, April 14 and April 26, respectively. The Queen inquired whether they had ample frigid weather clothing. The Emperor presented M Hell his treatise on the Earth's movements, which was to be published. He directed him to report their findings through his confessor.

The itinerary was compiled mainly on the advise of the Trondheim Bishop Gunnerus (1718–1773), also a scientist. After completing his studies in Germany, he lectured in theology in Copenhagen and became bishop of Trondheim in 1758. As a member of the Norwegian Royal Society he studied the Norwegian flora during his official trips and published the Flora Norvegica. The itinerary plan consisted of combined travel by sea and land. Gunnerus offered help in selecting the ship and the necessary crew, as the two-month sea voyage would be rather difficult, but less so than the shorter land route through Sweden, which was outside the jurisdiction of the Danish King. Christian Horrebow (1718–1776), head of the Copenhagen astronomical observatory also provided advice and recommended his student Jens Finns Borchgrevink, as interpreter and companion for the trip. Horrebow also studied sunspots and had some idea about their periodicity. He was less known than his more famous father, the inventor of the Horrebow level that is used by geodesists even to this day.

M Hell accepted these proposals then proceeded with Sajnovics from Vienna on April 28, 1768, and arrived in Copenhagen on June 11th. Mercier, the secretary of the Habsburg embassy in Denmark reported to Prince Kaunitz, the head of the Austrian administration the details on M Hell's arrival. Before the arrival, the two scientists met with the Danish King three times in Traventhal near Lübeck. In the

capital they were accommodated in Horrebow's house where they tried to repair and improve the instrumentation of his observatory. They made the final decision to proceed through Trondheim, and from there by sea, since in Danish territory it appeared easier to ensure an ample food supply and other provisions. Leaving Copenhagen on July 3rd, they arrived in Christiania (Oslo) on July 13th, and in Trondheim on July 30th, where Bishop Gunnerus was most helpful in acquiring food, warm clothing and especially furs.

The most difficult issue was to find an acceptable ship. Halvor Johansen's Anna was selected. The crew consisted of the following: Styrman Joen Johansen, proprietor of the ship; Halvor Gundersen, boadsman; Sten Wirthmann and Ingebrigt Paulsen seaman, and an artillerist from the local garrison, Jean Caspar Müller as cook. They all promised to obey Professor M Hell's commands. Johansen was paid 115 thalers for the journey.

They bought the following provisions: 7 1/2 barrels bread, 13 cords birchwood, 30 dozen boards (planks), 24 empty bottles, 1 ton pea, 4 sides smoked bacon, 2 firearms, 1 barrel gunpowder, 36 pounds lead, 2 sacks shot, 2 coops of chicken, 4 one ton sacks of barley, 2 ovens, 10 iron plates, 2 tons chalk, 400 building stones, 6 pieces salmon, 2 boxes lighting material, 5 tons rye flour, 30 canvases, 10 barrels wheat flour, 6 one ton sacks of malt, 1 sack of oats, 5 cheeses, 4 ankers bear, 1 barrel currant, 1/4 tons butter, 4 chaises, 1 barrel salt, one ton tobacco, 1 basket of pipe, 1 complete tent, 15 ankers wine, (normal and cognac), 19 boxes with signals. The building materials were for the construction of the observatory.

They departed Trondheim on August 22nd with 12 aboard that included Hagerup, Finnmarken's future president. After many adventures they arrived at Vardö on October 11th. Here they commenced immediately to erect the observatory and to carry out astronomical and geographic measurements. They hunted also as a pastime in view of the limited possibilities in this remote region. Other scientific tasks performed by M Hell and his retinue included observations in the sea's phosphorecence, changes in sea level and in relation to auroras. He correctly identified the source of phosphorescence to living organisms.

Besides the astronomical work of M Hell's group, Bellay, an English astronomer attempted to observe the transition of Venus from Kjelvig and Diksen from Rype Klübben near Hammerfest. However, it was only M Hell that succeeded in the observation on June 3, 1769. At the critical moments the sky happened to be clear on the otherwise cloudy day. They also observed the solar eclipse on June 4th. They departed Vardö on July 27th and reached Trondheim on August 30th. At this site the local scientific society elected them members, they sold clothing and metal objects at an auction, thus financial matters could be concluded for this part of the journey. After departure on September 12th they remained in Oslo from September 27 through October 2nd, then reached Copenhagen on October 17th. M Hell presented several lecture series here, including one on the aurora, thence was bestowed as a member of the Royal Society along with Sajnovics. Departing Copenhagen on May 22, 1770, they visited several scientific centers including Göttingen in Germany and returned to Vienna on August 12, 1770.

The primary goal of the expedition was to observe the transition of Venus at

inferior conjunction on June 3, 1769. At the royal audience in Copenhagen, King Christian inquired as to the sky conditions during the critical interval. M Hell replied that an independent scientific plan was also prepared in the event that the main objective would not succeed. Curiously the description of the successful observation became the focus of a lengthy discussion. Firstly Lalande, then Littrow jr., the editor of Sajnovics' report discredited the results. They were dubious about the reliability of the observations, moreover they suspected the plausibility of the results, surmising that they were falsified in order to get the correct value. Much later it was Newcombe who again reviewed the manuscripts preserved in the Vienna observatory and found a correct solution to all the problems doubted by Lalande and Littrow, including the latter's color-blindness. Besides it is disputable whether a scientist with a M Hell's credentials and reputation would falsify the observations.

The following compares M Hell's considerations of auroras prior to the expedition and residence in Vardö. Earlier on April 1, 1761, he wrote to Pater Weiss who observed an aurora in Trnava (Tyrnavia in his letter) that he simultaneously observed. Weiss provided Hell with a detailed description of the event with no exact date.

M Hell's response follows: "I thank you very much for sending me the observations and for the elegant description of the aurora borealis observed in Tyrnavia which corresponded mostly with ours here in Vienna, but here in Vienna the rays were smaller and a radius extending from the northern pyramid was observed; as I started to see the phenomenon later, I could see neither the radius nor the multiple rays; I didn't see those electric fascicles to the left of the binary radius, as there were many vapors near to the horizon, else I have seen three greater radii. I couldn't identify that cloud above the radii from this place, nevertheless, this phenomenon is an electric one, I told visitors that above the greater radii some clouds or denser air have to exist which are produced by the higher or lower electricity which resulted from the top of our telluric mountain peaks from which these electric radii seem to extend. Surely many of us were happy with the observation of that cloud from Tyrnavia which I had seen with my brain and which cloud demonstrated that my sentence: the aurora borealis is a phenomenon of electricity, is miraculously true."

Thus, before the expedition, M Hell was of the same opinion concerning the origin of the aurora as his most prominent contemporaries. He published a book on steel magnets, as previously mentioned with his biography, thus was well acquainted with magnetic properties and their usage in magnetic instrumentation. He had instrument quality magnetic needles including one of English source reputed to be very sensitive and accurate. Sajnovics' fragmentary diary also makes several mentions of magnetic observations. After the observation of the Venus transition they regretted not having made barometric or magnetic observations. At this time, they were simply too engrossed and excited to include them.

In both lectures at Copenhagen and a paper published in the Ephemerides 1777 volume, M Hell admitted his presupposition of a correlation between auroras and electricity. As he wrote, his experiments have shown the relationship between magnetism and electricity. Magnetism is more expressed toward the poles therefore this has also to be true for electricity. It is conjectures that electricity provokes

the luminous phenomena, thus the basic principle of excitation of auroras must be electricity.

After a long lapse in auroral activity in the second half of the 18th century, auroras again became the focus of interest. The viewpoint of its relationship to geomagnetism became the consensus at the time linked to the discovery by Celsius and supported by most scientists. This concept was confirmed by M Hell's experiments (in 1753) proving the magnetism/electricity correlation prior to Oersted's discovery (1820), some fifty years later.

M Hell carried out many experiments during the expedition with the naked eye, with the finest available English electrical machine, and with a magnetic needle. At this time he was very disappointed that he had observed no correlation between auroras and electricity. The magnetic needle did not budge as expected. In detail he had described his futile effort to find even a minuscule movement. It is likely that the needle was much less sensitive than presumed, or the rough polar conditions reduced its sensitivity. These propositions are more likely, as reported by Sajnovics concerning instrumentation incidences, e.g. broken thermometers before they arrived at Vardö.

Sajnovics' diary describes approximately ten auroral events with some detail. The list may be incomplete since Littrow, publisher of the diary, omitted several parts, or about one fourth of the original text according by his own admission. In the case on July 23, 1769, during the trip back to Vienna strong magnetic changes were mentioned, but it is not clear if these refer to temporal or spatial ones.

Quoting the description of an aurora on February 2, 1769 note the following: "On the 2nd we saw a great aurora, the rays were most agile and of changing form and color. A most luminous arc extended from west to east and was about parallel to the equator."

A similarly description from May 27, 1769: "At the evening we continued the angle measurements. At complete light we saw a beautiful aurora during which the sun was covered toward northwest by thick clouds. The rays of this aurora were extraordinarily great, numerous and very vividly colored. The longest extended through the zenith and they were very agile, sometimes they got much shorter then became longer again."

These descriptions also generally characterize their observations in general, too. Longer descriptions were provided during their settlement in Vardö, during travel they were shorter. Weather conditions were indicated, as were the position of the Sun, and sometimes the direction of the rays and arcs.

In Copenhagen M Hell presented many contemporaneous presumptions about auroras. These theories include the following:

De Mairan: aurora is due to a mixing process of the solar and terrestrial atmospheres, then fermentation.

Other mathematicians: aurora is due to optical effects, to breaking of solar radiation at clouds and reflection from polar snowfields.

Physicists: aurora is due to magnetic emanation from the poles; this emanation carries with itself terrestrial vapors; these vapors became fermented or catch fire through friction.

Philosophers explained auroras by light emitted from Icelandic volcanoes.

Northern people thought that aurora is due to tiny frozen particles carried to and fro by wind, without explaining the mechanism of excitation. (This is most likely information collected by Sajnovics from local people.)

His own idea prior to the expedition: aurora is due to electricity/magnetism, being stronger toward the poles, then by light due to electrical discharges.

In the next lengthy portion of his lecture he presented descriptions of auroras observed in different geographic zones, according to (geographic) latitude (60 to 66, 50 to 60, 40 to 50; below 40 auroral observation unknown to him). His "new" theory followed which is based on an observation on January 11, 1769. This is not mentioned in Sajnovics' diary. This theory postulated that vapors rose from the surface, became thicker upon freezing, then aurora appeared in a manner that stars disappeared behind it. Later snow fell on the surface.

He imagined that moisture laden air coming from the south at a certain height becomes frozen in polar regions, then solar radiation causes the aurora. He conjectured this theory as being proven by the fact that the rays followed the sun. Even the aurora due to the moon appeared to him feasible in spite of the prevailing opinion of the northern people (again information from Sajnovics), that auroras do not appear together with the Moon.

His summary is as follows:

"Thus aurora borealis is a phenomenon of our atmosphere of purely optic origin and its material is frozen vapor particles in various configurations, sometimes plane, extremely smooth, light and of variable density and rarity." Following a description of the possible movements of these particles, he continued: "The light of this aurora comes then from the solar radiation, other times from lunar radiation, sometimes from the combined radiation of both celestial bodies and on the surface the frozen particles, the light is reflected in different forms." The aurora can be thus explained by the laws of optics and physics in general.

Then he proceeded to reply to objections. For example is it possible to observe auroras after twilight when the Sun is more than 20 below the horizon? His rebuttal to this query is that it is uncertain whether the atmosphere has limited height, it may be much thicker than presumed at the time. Concerning colors, he refers to the example of a rainbow proposing that refraction and reflection occur simultaneously.

The publication ended here without the inclusion of three parts which were promised. It is possible that the mentioned troubles in the Danish royal family, and the suppression of the Jesuits in Austria inhibited its printing. It is also feasible that M Hell, after publishing the treatise on the aurora in Copenhagen (1770) and the Ephemerides (in 1777), became dubious about his own theory, and refrained from publishing the remaining parts.

M Hell's personality and scope of activity are of interest from two viewpoints. First, his nationality is a paradox, he was a typical representative of a Habsburg monarchy citizen with difficulty in establishing his ethnic or national origin: German, Czech, Slovak, Hungarian? Secondly, from a scientific point of view, he represented a class of scientists with very divergent fields of interest. He was well acquainted with achievements of contemporaries. Though he performed experi-

ments, he lacked the sufficient theoretical background. Unsatisfactory instruments available misled him his evaluation concerning the origin of auroras. This occurred in spite of correct observations and deductions; e.g. the thickness of the terrestrial atmosphere.

References

Faller J 1953: A magyar bányagépesítés úttörői a XVIII. században (Pioneers of Mining Mechanization in Hungary at the 18th Century), Akadémiai Kiadó, Budapest

Křivský L 1997: Pioneers of polar auroras in Hungary 1768-1910 and spectral research at Observatory Ógyalla by Konkoly-Thege. *Acta Geod. Geoph. Hung.*, 32 (present issue)

Littrow C L 1835: P. Hell's Reise nach Wardoe bei Lappland und seine Beobachtung des Venus-Durchganges im Jahre 1769 aus den aufgefundenen Tagebüchern geschöpft und mit Erl. begleitet (P. Hell's Trip to Wardoe in Lapland and His Observation of the Venus-Transit in the Year 1769, Taken from the Discovered Diaries with Comments), Gerold, Vienna

Lux A A 1988: Selmecbánya as the Birthplace of Technical Science in Hungary, Cleveland State university, Lecture Series in Hungarian History, Literature, Arts and Sciences, Lecture XV, (1984).

Pinzger F 1920: Hell Miksa emlékezete (To the Memory of Miksa Hell), Budapest, Hungarian Academy of Sciences

Réthly A, Berkes Z 1963: Nordlichtbeobachtungen in Ungarn (1523–1960). Akadémiai Kiadó, Budapest

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 245–248 (1997)

ESTIMATION OF COEFFICIENTS OF THE 2D-PROJECTIVE TRANSFORMATION WITH THE CAUCHY-FUNCTION

L BATTHA[1] and J SOMOGYI[1]

[Manuscript received June 16, 1997]

The paper deals with the robust estimation of coefficients of the two dimensional projective transformation using the Cauchy-function and shows a solution.

Keywords: Cauchy-function; non linear solution; projective transformation; robust estimation

1. Introduction

The different information systems (GIS) and the computer-aided design systems (CAD) play an inportant part in present time. In these systems facilities must be provided to detect and correct errors in the database. A number of methods have been worked out to detect gross errors in the input data.

For instance, in a CAD-system which is suitable to reconstruct facades from photographs made with metric- or non-metric cameras, the two-dimensional projective transformation is used to get the needed digital data. In case of redundant observations the least-squares method is generally used to estimate the parameters of the projective transformation, because it is supposed that the observation errors are normally distributed. It is known from computational experience that the presence of outliers and contaminations in data sets is possible which could lead to unacceptable results in the least-square method. Therefore robust statistical methods have been developed with the aim of reducing the effects of the outliers and contaminations in cases when the observation errors are assumed to follow a normal distribution.

Several robust methods have been developed. One of them is using the Cauchy probability density (Battha and Somogyi 1996).

This paper gives a robust solution for the 2D-projective transformation using the Cauchy-function.

2. Numerical procedure to find the minimum of the object function $f(x)$ for the case of 2D-projective transformation

The projective relation between two plane can be expressed by the well known fractional linear equations:

$$X_i = \frac{a_1 x_i + a_2 y_i + a_3}{c_1 x_i + c_2 y_i + 1} \quad \text{and} \quad Y_i = \frac{b_1 x_i + b_2 y_i + b_3}{c_1 x_i + c_2 y_i + 1} \quad (1 \leq i \leq n)\,, \qquad (1)$$

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

1217-8977/97/$ 5.00

where
n is the number of points projected and
X_i, Y_i denote the Cartesian coordinates in one plan
x_i, y_i denote the corresponding coordinates in the other plan.

After multiplying the above equations by the denominator, we obtain

$$\begin{aligned} -X_i x_i c_1 - X_i y_i c_2 + x_i a_1 + y_i a_2 + a_3 \qquad\qquad & X_i \\ - Y_i x_i c_1 - Y_i y_i c_2 \qquad\qquad + x_i b_1 + y_i b_2 + b_3 = Y_i \quad . & \end{aligned} \tag{2}$$

There are eight unknowns in Eq. (2), requiring four non-collinear points for a unique solution. In case of additional points an overdetermined solution can be used for the determination of the transformation parameters, (e.g. the least-squares method).

Now we can apply to the above linear system of equations the Cauchy object function, and solve the resulting non linear minimization problem by using the gradient (or steepest descent) method combined with interval halving (Battha and Somogyi 1996).

The object function $\varrho(x)$ which expresses the mathematical connection between the observations x_i and the parameters Θ_i, as well as the influence function $\psi(x)$ for the Cauchy distribution are:

$$\varrho(x) = \ln(1 + x^2) + C \qquad \text{and} \qquad \psi(x) = 2x/(1 + x^2) \tag{3}$$

wich gives a non linear equations' system:

$$\sum_{i=1}^{n} \psi(x_i - \Theta_i) = 0\,. \tag{4}$$

To obtain the solution, we have to find the minimum of $f(y) = \sum_{i=1}^{n} \varrho(u_i) + \sum_{i=1}^{n} \varrho(v_i)$ where

$$\begin{aligned} y &= (c_1, c_2, a_1, a_2, a_3, b_1, b_2, b_3) \\ u_i &= -X_i x_i c_1 - X_i y_i c_2 + x_i a_1 + y_i a_2 + a_3 - X_i \\ v_i &= -Y_i x_i c_1 - Y_i y_i c_2 + x_i b_1 + y_i b_2 + b_3 - Y_i \end{aligned} \tag{5}$$

To start the iteration, a least squares solution y_0 of (2) is computed. Then in each iteration step, the partial derivates $\partial u_i/\partial c_j$, $\partial u_i/\partial a_j$, $\partial v_i/\partial c_j$, $\partial v_i/\partial b_j$ are computed, and used for the next "steepest descent". The coordinates of $g = \text{grad}\ f$ are obtained from

$$\begin{aligned} \frac{\partial f}{\partial c_j} &= \Sigma\Psi(u_i)\frac{\partial u_i}{\partial c_j} + \Sigma\Psi(v_i)\frac{\partial v_i}{\partial c_j} \\ \frac{\partial f}{\partial a_j} &= \Sigma\Psi(u_i)\frac{\partial u_i}{\partial a_j} + \Sigma\Psi(v_i)\frac{\partial v_i}{\partial a_j} \\ \frac{\partial f}{\partial b_j} &= \Sigma\Psi(u_i)\frac{\partial u_i}{\partial b_j} + \Sigma\Psi(v_i)\frac{\partial v_i}{\partial b_j}\,. \end{aligned} \tag{6}$$

Table I.

					Noise values	
P	X_i	Y_i	x_i	y_i	x	y
	m		mm		mm	
1	800	1650	−112.376	−112.859	−0.002	−0.001
2	1000	3400	−83.244	91.812	0.004	0.003
3	1500	3500	−32.966	95.527	−0.005	0.001
4	1300	1900	−48.833	−70.555	−0.009	0.007
5	1850	2550	4.285	5.100	0.003	−0.001
6	2400	1800	59.312	−67.706	0.000	+5.998
7	3000	1700	106.776	−70.898	0.002	−0.006
8	3100	3300	92.241	66.577	0.003	−0.001

Table II.

	Normal distribution		Long-tailed distribution	
P	v_X	v_Y	v_X	v_Y
	m		m	
1	−0.016	0.035	−0.018	0.024
2	−0.022	0.001	−0.039	0.014
3	0.047	0.001	0.049	0.001
4	0.047	−0.062	0.078	−0.08
5	−0.052	0.018	−0.018	0.026
6	−0.015	−0.004	−7.8	−55.0
7	0.018	0.011	−0.076	0.035
8	−0.006	0.000	−0.001	−0.022

To prevent numerical instabilities from stopping the iteration, additional "partial descent steps" are included. In each such step all parameters except one "active parameter" (a_i, b_j, or c_k) are "switched off" (fixed) so that the full descent is effected parallel to the axis of the chosen parameter.

3. Numerical example

In this section we give an example of robust solution based on the method described in the previous section using a problem of 2D-projective transformation.

To demonstrate the applicability of the presented method, it was generated to the x_i, y_i values of a simulated projective transformation consisting of 8 points, then a noise with normal distribution and a gross error was added (contaminated normal distribution). Choosing $\sigma = \pm 0.005$ mm for the standard deviation of the assumed measurements, 0.015 mm was obtained for the acceptable error.

Table III.

Tr. Par.	Normal distribution	Long-tailed distribution	Given values
a_1	8000.524	7999.992	8000
a_2	−1100.445	−1100.370	−1100
a_3	1800.026	1799.992	1800
b_1	−3999.073	−3999.606	4000
b_2	7399.635	7399.460	7400
b_3	2499.997	2499.989	2500
c_1	−1.49966	−1.49988	−1.5
c_2	−1.00014	−1.00024	−1.0

The X_i, Y_i, x_i, y_i coordinates and the x, y noise values are presented in Table I. Table II shows the coordinate discrepancies after the adjustment and Table III collects the given and estimated values of the transformation parameters.

As it appears from the Tables II and III, when outliers may be present in the data set, the use of the Cauchy-function gives a satisfactory robust solution.

Aknowledgement

This work was supported by OTKA (National Science Foundation) under Grant No.014856.

Reference

Battha L, Somogyi J 1996: *Acta Geod. Geoph. Hung.*, 31, 5-9.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 249–252 (1997)

Professor emeritus Ernő Takács 70 years old

Ernő Takács was born on February 1, 1927 in Dorog, his father was smith at the local mining company. He studied at the Benedictine secondary grammar school in Esztergom, then between 1944 and 1946 he had to interrupt his studies, as member of a paramilitary youth organisation he was American war prisoner. Following return in August 1946, he made his final examination at the secondary school, joined Dorog coal mines as draftsman, then from 1947 on he studied at the Mining Prospecting specialisation of the Technical University, Sopron. He graduated in 1952, and simultaneously he participated in the work of the newly established Department of Geophysics as an assistant. Among his duties, the organisation of practical training was one of the most difficult ones due to the lack of up-to-date equipment. He also participated in the earth current measurements in Hungary, including the first experiments in 1952 with Pál Egerszegi, near Sopron at Kópháza. In the years to follow, regular summer field measurements were organised together with the staff of the Geophysics Panel of the Hungarian Academy of Sciences.

The earth current method of exploration was presented in autumn 1955 by him, together with Professor Károly Kántás and Antal Ádám in China. In the next year,

1217-8977/97/$ 5.00

he started such measurements in China as head of the Hungarian group and worked there for three years. During his stay, he invented an apparatus for the processing of earth current measurement with the so-called method of total variations. The earth current crew of the Hungarian geophysical expedition in China supplemented well gravimetric and seismic measurements.

In 1959 he returned to Hungary, but meanwhile the Department of Geophysics was transferred from Sopron to Miskolc. He held lectures in all branches of exploration geophysics to students of geophysics, geology and mining. His research concentrated on the introduction of the magnetotelluric method, he designed an instrument and in 1960 carried out the first measurements in Mezőkeresztes to supplement earth current measurements of the Department made for the Hungarian Oil and Gas Trust. In the following years, he studied the capabilities of the MT method along a North-South profile in Eastern Hungary.

Between 1961 and 1965, he was a postgraduate student (independent aspirant as it was called then), and obtained the degree Candidate of Science based on a thesis on magnetotelluric problems and became assistant professor. The most important result of his magnetotelluric measurements was the discovery of the Transdanubian Conductivity Anomaly (TCA) in 1966, being an object of investigations till present. He drew theoretical conclusions from his results, too, thus he found a connection between the direction of impedance ellipses and the direction of the electric component belonging to the maximum of the vertical magnetic component. He was a leading figure in geophysical exploration of bauxite in Hungary during several decades. Physical model experiments were started by him during his stay in Liverpool to clear the phenomena connected with the TCA and during the seventies, 2D and 3D model experiments gained importance in his activity for the interpretation of earth current- and MT-measurements. He introduced radio frequency sounding and the characteristics of the field of electric high voltage power lines into the geophysical methods in Hungary.

He started research into frequency sounding using artificial power source in 1968 and developed measuring systems and interpretation methods. The first such equipment was used to locate advantageous shotpoint sites. Later this method was also used by him for the detection of faults and cracks due to stress in coal-mines. Numerical modelling helped the interpretation. Programs developed under his guidance and with support from the Hungarian Oil Trust, OKGT, enabled Hungarian geophysicists to describe the magnetotelluric field over two- and three-dimensional structures and the field of an electric dipole in strike direction over two-dimensional models.

He designed several instruments and methods for the measurement of physical rock parameters, partly for educational purposes.

Takács' thesis for the DSc. degree dealt with problems of frequency sounding in mines and got this degree with honour from the Hungarian Academy of Sciences in 1991.

He lectured on geoelectric methods and on mining geophysics for student of geophysics and geology, and on geophysical exploration for future mining engineers and introduced some new courses as well (e.g. geophysical data processing, geophys-

ics of Hungary, special EM- methods etc.). He is the author of 10 lecture notes. Since 1973, he has been professor at the Miskolc University, between 1974 and 1984 he was elected three times Dean of the Faculty of Mining. Between 1983 and 1991, he was head of the Geophysical Department. He has been Professor Emeritus since 1996. Among his memberships in scientific bodies, the most important ones are elected member of the General Assembly of the Hungarian Academy of Sciences, chairman of the Scientific Committee of Geophysics at the Hungarian Academy of Sciences (twice) and membership in the Committee for Scientific Qualification. He organised several successful scientific meeting both in Budapest and Miskolc. He is the author or co-author of more than 100 scientific papers.

He spent relatively long research time in Germany, in the one-time Soviet Union, in Egypt and in the UK. He is bearer of the Chinese Friendship Medal (1959), of the Order of Labour, silver (1978) and golden (1984) grade, of the golden (1976) and diamond (1988) grades of the Mining Service Medal and also of the László Egyed medal of the Association of Hungarian Geophysicists (1989).

At the occasion of his 70th birthday, he was congratulated by lots of foreign and Hungarian friends, earlier students and colleagues, not to speak of his wife (he married in 1956) and of his son, also a geophysicist, who won the best paper award for 1996 just previously. The scientific session to this occasion included about 20 lectures by Hungarian, American, Finnish and Russian authors. They all drank to his health and wished him many more years to work and to teach.

A Ádám, F Steiner, J Verő

References

Csókás J, Takács E 1965: Magnetotellurische Messungen auf der großen Ungarischen Tiefebene. *Freiberger Forschungshefte*, C 174, 83–90.

Csókás J, Takács E 1967: Magnetic well logging by artifical remanence. *Publications of the Technical University for Heavy Industry*, 16, 127–130.

Hursán L, Takács E 1990: Möglichkeiten zur Messung künstlich erzeugter magnetischer Remanenz im Bohrloch. *Freiberger Forschungshefte*, C 447, 146–152.

Steiner F, Takács E 1966: Die Bestimmung der Neigungsverhältnisse der Grundgebirges auf Grund der vertikalen Komponente des tellurisch-elektrischen Feldes. *Acta Geod. Geoph. Mont. Hung.*, 1, 49–60.

Takács E 1968: Anomalous conductivity of the upper crust in the NW foreground of the Bakony mountains. *Acta Geod. Geoph. Mont. Hung.*, 3, 155–160.

Takács E 1969: Possibilities of application of the Earth's electro-magnetic field impedance in the geological mapping of Hungary. *Publications of the Technical University for Heavy Industry*, 28, 175–191.

Takács E 1969: The orientation of the magnetotelluric impedance ellipses. *Acta Geod. Geoph. Mont. Hung.*, 4, 415–423.

Takács E 1971: The role of the vertical pulsations of the magnetic field in magnetotelluric measurements. *Acta Geod. Geoph. Mont. Hung.*, 6, 99–110.

Takács E 1973: Magnetotelluric fields above two-dimensional, two-layered ($\varrho_2 = \infty$) buried structures of monocline and syncline character. *Acta Geod. Geoph. Mont. Hung.*, 8, 229–233.

Takács E 1977: Method for determining the variation in the absorption coefficient of seismic waves with time. *Publications of the Technical University for Heavy Industry, Series A. Mining*, 34, 71–79.

Takács E 1988: In-mine frequency sounding with a buried grounded dipole sources. *Geophysical Transactions*, 34, 343–359.

Takács E 1989: Exploration of coal seams by the measurement of the electric field of a buried vertical AC electric dipole. *Acta Geod. Geoph. Mont. Hung.*, 24, 453–469.

Takács E 1989: Investigations associated with the possibility of the underground frequency sounding using vertical electric transmitter and receiver dipoles in equatorial arrangement. *Annales Univ. Sci. R Eötvös, Sectio G e M*, Volumes 3–4–5, 25–34.

Takács E 1989: Feature of the electromagnetic field in coal measures and their possible use in in-mine surveys. 34th Internation Geophysical Symposium Budapest, Abstracts and Papers of the Technical Program II, 681–691.

Takács E 1991: A comparison between horizontal magnetic and vertical electric fields due to a vertical electric dipole situated within a coal seam. *Acta Geod. Geoph. Mont. Hung.*, 26, 339–350.

Takács E 1991: Exploration of roof and floor layers by the horizontal electric component due to a vertical electric dipole situated within a coal seam. *Acta Geod. Geoph. Mont. Hung.*, 26, 351–363.

Takács E 1991: Effect of lateral inhomogeneities on the frequency sounding curves obtained from the vertical electric components of a buried vertical current dipole. *Geophysical Transactions*, 37, 39–55.

Takács E 1992: Investigation of the feature of the EM field of an underground vertical AC. Electric dipole in respect of mining geophysics. Summary of Dr.Sc. Dissertation. *Annales Univ. Sci. R Eötvös, Sectio G e M* , Vol. 8, 102–103.

Takács E 1993: In-mine exploration of coal seams by the electromagnetic field of a buried vertical AC. Electric dipole. *Annales Univ. Sci. R Eötvös, Sectio G e M*, Vol. 9, 161–208.

Takács E 1995: The normal surface electric field of an E-field excited steel casing. *Acta Geod. Geoph. Hung.*, 30, 331–341.

Takács E, Egerszegi P 1982: Einige Experimente mit elektromagnetischen Verfahren bei der Klärung tektonischer Probleme von Braunkohlenlagerstätten. *Freiberger Forschungshefte*, C 378, 99–108.

Takács E, Egerszegi P 1986: Geophysikalische Untersuchungen der durch den Abbau hervorgerufenen Änderungen im Zustand der Hangenschicht eines Kohlenflözes. *Freiberger Forschungshefte*, C 416, 37–54.

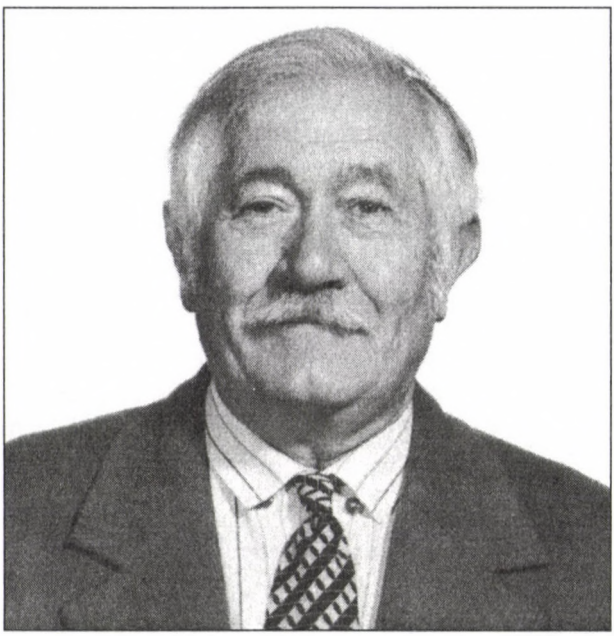
Takács E, Tevan Gy 1973: Numerical method for the computation of magnetotelluric field in inhomogeneous media. *Acta Geod. Geoph. Mont. Hung.*, 8, 55–69.

Takács E, Turai E 1986: Approximative solution of the direct problem of magnetotellurics for two-layered, three-dimensional structures. *Acta Geod. Geoph. Mont. Hung.*, 21, 167–176.

Takács E, Nagy J, Mádai F 1986: Field of a vertical, alternating current, electric elementary dipole in a layered medium. *Geophysical Transactions*, 32, 43–56.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 253–258 (1997)

In memoriam Professor Lajos Stegena (1921–1997)



We wished a last farewell to Lajos Stegena, Professor emeritus of the Loránd Eötvös University, laureate of the Eötvös-wreath of the Hungarian Academy of Sciences in the Rákospalota Cemetery in Budapest on 27 February, 1997 who returned to his Creator after an incurable illness endured with profound patience.

He was a man of wide interest and death finished a curriculum of long span.

He was born on 10 November, 1921 in Keszegfalva, in the present Slovakia as the second son among the three offsprings of a teacher's family. Both his father and mother were teachers of the village. Following the peace treaty after World War I, the family moved to Budapest. He attended the Kálmán Könyves secondary school, then the Budapest Technical University where he received his diploma in chemical engineering in 1942. He joined the Hungarian State Geological Institute and participated in soil research till 1946, when he joined the Loránd Eötvös Geophysical Institute. He carried out research in seismics, geothermics, geochemistry and radiometry, from 1953 on as head of department. In 1957 he received his Candidate of Science degree and in 1963, his Doctor of Earth Science degree. After 1953, he took also part in education at the Faculty of Natural Sciences at the Loránd Eötvös University and since 1962, he had been a Professor of Geophysics there. From 1966 till

1217-8977/97/$ 5.00

1987, through 21 years, he headed the Department of Cartography there, between 1975 and 1984, he led the Committee for Earth Sciences of the Faculty, then till 1987, the Department Group of Environmental Sciences. He lectured during several decades to geophysicists in his favourite geophysical subjects (seismics, gravimetry, geothermics, tectonics), then certain parts of cartography for which he developed a taste, too (projections, history of cartography). He is the author or co-author of six textbooks and many lecture notes. The memory of his educational activity is conserved by a lot of geophysicists, geologists, cartographers. He supervised a dozen postgraduate (and even higher) students.

His activity in creating a scientific school for studying the geonomy of the carpathian Basin is documented by about 150 publications, 23 books and several invitations from foreign universities as guest professor. About 30 universities invited him to give lectures, he was a visiting professor at the universities Freiberg, Clausthal, Kiel and Naples. As UN expert, he was an invited lecturer at the Geothermics World Congress of 1975, San Francisco and the Cartography Seminar of 1983, Moscow. He was invited three times to China for a series of lectures. He spoke, wrote and read four foreign languages.

He was an internationally recognised expert and teacher of earth sciences.

Using seismometers of his own invention and the seismic reflexion method, he was the first to determine the structure of the crust in the Pannonian Basin and found a thinning-out of the crust. This was the first such application of the method in the world. His publications and maps on the geothermal conditions, energy possibilities of the Pannonian Basin are of pioneering importance. His geothermal maps and books constructed in international co-operation belong to the basic works of geothermics.

Plate tectonic research in Hungary was initiated by him, too. Based on studies of crustal structure, gravimetry, geothermics and tectonophysics, he presented together with his group a plate tectonic theory of the development of the Pannonian Basin. Due to his activity, strong schools of seismic crustal study, geothermics and plate tectonics came into being in Hungary.

His books on map history and on the history of Hungarian maps before the Mohács Disaster (defeat from the Turkish army, 1526) are significant additions to the history of Hungarian cartography, and his monographs on Lazarus Secretarius and Antal Reguly highlighted interesting sections of cultural history. He did not live long enough to see the publication of his last book, on the pre-war Hungarian cartography.

His activity of science organisation was exemplary: he participated in IUGG and EGS activities, as member of different committees for a long time. He presided between 1979 and 1983 the International Committee on Geothermics of IUGG and at the same time he was member of the Governing Council of EGS. As such, he organised more than twenty conferences, including the Budapest congress of EGS/ESC in 1980. He was a member of the Association of Hungarian Geophysicist, chairman of the Scientific Committee for Geonomy and member of the Scientific Committee for Geophysics of the Hungarian Academy of Sciences.

He initiated or led several Hungarian and international projects, including as

the most prominent the NSF-MTA project together with MIT, Boston. He also led till his death the project Geothermics supported by the State Science Foundation and the study of extraterrestrial materials, in co-operation with four universities.

Memberships in editorial boards included Clausthaler Forschungshefte, Zeitschrift für Geophysik/Journal of Geophysics (1975–1984), and in Hungary, Annales Sectio Geologica for a long time, then Annales Sectio Geophysica et Meteorologica till his death.

In recognition of his educational/research activity in earth sciences carried out for several decades he was awarded medals from the universities Prague, Freiberg, Clausthal, Moscow, Padova and Sofia, and in 1990, the golden medal of the Loránd Eötvös University.

He said in the hospital lying already in bed: we sit for examination and examine during the whole life, at the end the leaving certificate is issued by our students and friends.

Dear Professor Stegena, our tutor and friend, death finished your final certificate, it is issued with honours. Peace to your spirit.

I Klinghammer

Books

Praxis der seismischen Feldmessung und Auswertung. Borntraeger, Berlin, 1977 (with co-authors)

The constant of gravitation. Akadémiai Kiadó, Budapest, 1979 (with co-authors)

Geothermics and Geothermal Energy. Birkhäuser, Basel-Stuttgart, 1979 (with co-authors)

Lazarus Secretarius, the First Hungarian Mapmaker and His Work. Akadémiai Kiadó, Budapest, 1982 (with co-authors)

Antony Reguly's Northern Ural Map 1846. Stud. Cart., 9., Térk. Tanszék, Budapest, 1983

Paleogeothermics. Springer, Berlin ...Tokyo, 1986 (with co-authors)

Rudimenta Cosmographica Johannes Honterus Coronensis, Coronae, 1542. Stud. Cart., 10, Térk. Tanszék, Budapest (with co-authors)

Handbook of Terrestrial Heat-Flow Density Determination. Kluwer, Dordrecht-Boston-London, 1988 (with co-authors)

Hydrological Regimes and their Subsurface Thermal Effects. Geophysical Monographs 47, Am. Geophys. Union, Washington, 1989 (with co-authors)

and fourten books in Hungarian and other languages!

Publications

On problems of seismometer construction (in Hungarian). *MTA Műsz. Tud. Oszt. Közl.*, 5, 1952, 93–105.

Deep reflections from the Hajdúszoboszló area (in Hungarian). *Geofiz. Közl.*, 4, 1955, 37–40. (with co-authors)

Geothermal maps of Hungary (in Hungarian). *Geofiz. Közl.*, 13, 1965, 221–230.
Gravity research of the deep structure of Hungary. *Ann. Univ. Sci. Budapest, Sec. Geol.*, 8, 1965, 153–159. (with co-authors)
Information theoretical bases of coding in maps (in Hungarian). *Stud. Cart.*, 3, 1972, 77–82.
15th and 16th century maps of Hungary in Italian collections. *Ann. Univ. Sci. Budapest, Sec. Geogr.*, 7, 1971, 157–167.
Geothermal maps of Hungary. Geothermal temperatures at depths 0.5, 1.0, 1.5, 2.0, 3.0 km. 1:1.5 M (in Hungarian). Magyarország Tudományos Térképei, I/1–I/5. 1972
Tabula Hungariae ad quatuor latera. Lazarus studiosus' map of Hungary. *Magyar Tudomány*, 10, 1978, 890–993.
Hungarian contribution to the geophysical mapping of the world. In: Contributions to the History of Geological Mapping. Akadémiai Kiadó, Budapest, 1984, 251–259.
Role of sedimentations and thermal history in oil generation (in Hungarian). *Kőolaj és Földgáz*, 16, 1983, 1–6.
Map of geothermic, hydrothermic and balneotherapeutic areas of Greece. Geogr. Medica. *Int. Journ. of Geogr. of Health*, 14, 1984, 230–234. (with co-authors)
Is geothermal energy usable in Hungary (in Hungarian)? *Magyar Tudomány*, 7, 1991, 892–894.
Illustration of the river Danube on old maps and orientation of the Lazarus map (in Hungarian). *Geod. és Kart.*, 40, 1988, 354–359.
Tiefreflexionsversuche in Ungarn zum Studium der kontinentalen Aufbauung. *Geol. Rundschau*, 146, 1957, 26–29. (with co-authors)
On the principles of geochemical oil prospecting. *Geophysics*, 26, 1961, 447–451.
On the basic principles of the direct geochemical and radiometric exploration. *Geophys. Prosp.*, 11, 1963, 368.
Il principio dell' attualismo e la superficie Moho. *Boll. di Geofis. Teor. ed Appl.*, 8, 1966, 232–239.
Spreading tectonics investigated by magnetotelluric anisotropy. *Nature*, 5, 1971, 5203. (with co-authors)
Geothermal map of Eastern Europe. *Geothermics*, 1, 1972, 140–141.
Late Cenozoic evolution of the Pannonian Basin. *Tectonophysics*, 26, 1975, 71–90. (with co-authors)
Ensinatic and Ensialic Interarc Basins. *J. Geoph. Res.*, 80, 1975, 281–283. (with co-authors)
The Tyrrhenian and the Pannonian Basin: A comparison of two interarc basins. *Tectonophysics*, 1976. (with co-authors)
Geoscientific world mapping: facts and representation. *World Cartography*, 15, 1979, 77–89.
The formation of the intra-Carpathian basins as determined from subsidence data. *E. P. Sci. L.*, 51, 1980, 139–162. (with co-authors)
Determination of paleotemperature by vitrinite reflectance data. *Earth Evol. Sci.*, 3/4, 1981, 292–300. (with co-authors)

Water migration influences on the geothermics of basins. *Tectonophysics*, 83, 1982, 91–99.

Minoische kartenähnliche Fresken bei Acrotiri, Insel Thera (Santorini). *Kart. Nachr.*, 4, 1984, 141–143.

Heat flow determination in the Strymon basin. *J. Geoph. Res.*, 54, 1984, 106–109. (with co-authors)

Velocity structure and geothermics of the Earth's crust along the European Geotraverse. *Tectonophysics*, 121, 1985, 87–96. (with co-authors)

Geothermics and seismicity in the Pannonian Basin. In: Terrestrial and space techniques in earthquake prediction research. Viehweg, Braunschweig, 1979, 466–471.

The low enthalpy resource of the Pannonian Basin, Hungary. In: Geothermal Systems, Wiley, New York, 1981, 221–246. (with co-authors)

Applications of geothermics in petroleum researches. In: O.G.S. Silver Anniversary, Vol. 1958–1983, Trieste, 1984, 189–196.

High temperature geothermal reservoir possibilities in Hungary. Terra Nova, Oxford, 6, 1994, 282–288. (with co-authors)

Models of Mediterranean back-arc basin formation. Extensional tectonics associated with convergent plate boundaries. The Royal Soc. London, 1981, 165–184. (with co-authors) Reprint: Phil. Trans. Royal Soc. London, A, 300, 1454, 383–402.

Der Beitrag der Geothermie und Magnetotellurik bei der Erkundung des oberen Mantels. *Z. geol. Wiss.*, 10, 1982, 349–356.

Water migration influences on the geothermics of basins. *Tectonophysics*, 83, 1982, 91–99.

Stress determination in the Aegean area using volcanic lineaments. Proc. Int. Symp. H.E.A.T., Theophrastus, Athens, 1982, 309–317. (with co-authors)

Evolution of the Pannonian Basin system. *Tectonics*, 2, 1983, 91–137. (with co-authors)

Heat generation of the crust. Sitzungsberichte FEKAPE, TU Clausthal, 1984, 25.

Methods in paleogeothermics. *Lect. notes in Earth Sci.*, 5, 1986, 5–39. (with co-authors)

Statistical determination of class intervals for maps. *The Cartographic Journal*, 24, 1987, 142–146. (with co-authors)

Vitrinite paleogeothermics: some aspects. Rev. Brasil de Geofis., Sao Paulo, 1988.

Magnetotelluric sounding and geothermics in Central and Eastern Europe: a comparison. *Phys. Earth Planet. Int.*, 60, 1990, 229–230.

The 14 January 1810 earthquake at Mór, Hungary. The first map with isoseism. In: Historical earthquakes in Central Europe, Vienna, 1988 (with co-authors). Reprint: *Tectonophysics*, 193, 1991, 109–115.

Isolinie, Barometer, Lazarus, Ungarn, Kompass, Nadelabweichung, Deklination, Bussole. Entries in: Lexikon zur Geschichte der Kartographie, Wien, 1986.

The first European documentation from Japan's Geology. 29th Int. Geol. Congress, Kyoto, Vol. 3, 1992, 1212.

An atlas born before 450 years in Brasso: a Rudimenta Cartographica. *Cartographica Hungarica, Novafeltria*, 2, 1992, 15–18.

Fundamentals of geothermics. In: Handbook of terrestrial heat-flow density determination. Reidel, Dordrecht, 1988, 9–57. (with co-authors)

Heat production and seismic velocity of crustal rocks. *Terra Nova*, 2, 1990, 87–90. (with co-authors)

Correlation between seismicity and horizontal variation of heat flow density in Central Europe. *Tectonophysics*, 179, 1988, 55–61.

Palaeogeothermics. In: Handbook of terrestrial heat-flow density determination. Reidel, Dordrecht, 1988, 391–419.

Lithosphere and evolution of the Pannonian Basin. In: The Pannonian Basin. AAPG Memoir 45, Tulsa-Budapest, 1988, 147–152. (with co-authors)

Thermal effect of hydrogeology in closed basins. In: Hydrogeological regimes and their subsurface thermal effects. Geophys. Monograph, 47, AGU-IUGG, Washington, 1989, 81–86.

Acta Geod. Geoph. Hung., Vol. 32(1–2), pp. 259–262 (1997)

Book reviews

K KRAUS: *Photogrammetrie.* Band 2. Verfeinerte Methoden und Anwendungen. Mit Beiträge von J Jansa und H Kager. Dümmler Verlag, Bonn (1996, 488 pp, 228 figs, 40 tables)

The book is the third edition of Volume 2 published in 1987. This book is as compared to the second edition strongly rewritten and expanded.

The fundamental chapters are:

— Orientation photogrammetric images and photogrammetric point determination

— Photogrammetric collection and visualization of dynamic phenomena

— Calibration of photogrammetric systems.

In relation to the second edition this volume is enlarged with new special fields:

— Supplementation of the mathematical model for the general curves and surfaces in the object space

— Alternative mathematical models, like the linear transformation

— Determination the datum of photogrammetric blocks and free net adjustment

— The use of GPS data for the photogrammetric point determination

— Photogrammetric point determination with digital images and production of digital orthophotos

— Three dimensional digital photogrammetric models and their visualization

— Surface determination with mechanical vision

— Evaluation and production of picture series including animation

— Calibration of photogrammetric systems.

This up-to-date text book gives a very useful help to the students and practitioners in the field of photogrammetry but also gives information to those working in neighbouring disciplines.

J Somogyi

P FÜLSCHER ed.: *Photogrammetrie in der Schweiz, Geschichte-Entwicklung.* Schweizerische Gesellschaft für Photogrammetrie, Bildanalyse und Fernerkundung. Dümmler Verlag, Bonn, 144 pp, 98 figs, 8 colour maps

This book offers a brief survey of the technical development of photogrammetry in Switzerland until 1980.

Accordingly an inquirer can obtain information about the history of photogrammetry, plotter productions in Switzerland, photogrammetric applications, international cooperations, education, scientific work and so on.

The book is recommended to all who are interested in the history and development of the photogrammetry.

J Somogyi

1217-8977/97/$ 5.00

H WILHELM, U ZÜRN, H-G WENZEL eds: *Tidal Phenomena.* Lecture Notes in Earth Sciences 66, Springer 1997, pp 398

The book contains essentials of lectures given at the seminar on "Tidal Phenomena" organised and sponsored by the German Geophysical Society (Deutsche Geophysikalische Gesellschaft) in Oberwolfach (Schwarzwald, Germany), October 1994. The book gives an exceptionally broad view on tidal effects occurring in different parts of the Earth, moreover in the planetary system and in the universe.

Chapter I describes problems related to the earth tides (tide generating potential, tidal response of the solid earth, analysis of earth tidal records, earth tide observations and interpretation, the nearly-diurnal free-wobble resonance).

Chapter II gives information about ocean tides and about their influence on the solid earth (load) and on axial rotation. The reader finds here discussion on the Chandler wobble and pole tides, too.

Chapter III contains papers on atmospheric and geomagnetic tides. The next chapter on "Tidally Induced Phenomena" deals with tides in water-saturated rocks, with the importance of tides in triggering seismic and volcanic events.

The last part of the book describes tides in the outer space (satellite orbit perturbations, tides of Io, tidal effects in binary star systems and interactions between galaxies).

The significance of the book lies in its exceptionally broad coverage of the tidal phenomena. Therefore it contains new tidal information even for the experts active in this field.

P Varga

W SCHRÖDER, H J TREDER eds: *Ertel's Potential Vorticity.* Interdivisional Commission on History of the IAGA and History Commission of the German Geophysical Society, Bremen-Roennebeck, 1997, pp. 175, ISSN 0179–5658, 20 USD

The vorticity is always a very important vector governing motion in geophysical fluids. It is defined by $\zeta \equiv \nabla \times v$, where v is the velocity of a fluid parcel. An equation for $d\zeta/dt$, the rate of change in the vorticity, can be derived by taking the curl of the equation expressing the momentum conservation. Although the vorticity equation is useful because it deals directly with the vector character of vorticity, it is more or less a description of how vorticity is changed, not a useful constraint on the change. However, there does exist an unusually useful and elegant constraint on vorticity, called the *Ertel theorem*, from Hans Ertel's 1942 work published in the German journal *Meteorologische Zeitschrift.* Consider some scalar fluid property Θ, which satisfies an equation of the form $d\Theta/dt = Q$, where Q is an unspecified source of the property Θ. The Ertel theorem states that if a) Θ is a conserved quantity for each fluid element, b) the frictional force is negligible, and c) the fluid is barotropic, i.e., $\nabla\varrho \times \nabla p = 0$ (or Θ can be considered a function of pressure p and density ϱ only), then the quantity $\Pi = \varrho^{-1}(\zeta + 2\Omega) \cdot \nabla\Theta$ (where Ω is the earth's angular velocity) is conserved by each fluid element, that is, $d\Pi/dt = 0$. Quantity Π is *Ertel's potential vorticity.*

The booklet consists of three parts. On the first 35 pages, following a short introduction by the editors, reproductions of some selected papers by Professor Ertel are found. In addition to the famous potential vorticity theorem, the eminent German theoretical meteorologist derived other important theorems well-known today in geophysical fluid dynamics, including his commutation formula, and flow theorem which may be regarded

as a generalization of the corresponding circulation theorem for closed material curves in autobarotropic fluids.

The second part of the booklet is a collection of reflections on Ertel's potential vorticity. Opinions of outstanding theorists as H Reiter in the FRG, C Truesdell in the USA, R Hide in the UK, and H Pichler in Austria, just to mention a few, can be compared here with each other.

While collecting the material of the volume, the principal editor W Schröder made inquiries about Ertel and his work. Responses from some ten leading scientists, as well as some additional papers acknowledging Ertel's achievements, are reproduced in the third part of the booklet.

The book will be of value to researchers interested in the subject either from the scientific or from a historical point if view.

G Götz

W SCHRÖDER ed.: *Physics and Geophysics with Special Historical Case Studies (A Festschrift in Honour of Karl-Heinz Wiederkehr).* Interdivisional Commission on History of IAGA and History Commission of the German Geophysical Society, Bremen-Roennebeck, 1997, 410 pp., 20 USD

The Festschrift contains a rather great part of articles devoted to geophysics/history of geophysics, including historical geophysical data. Among the latter, the following should be mentioned:

Bartha: Data on early magnetic variation in the Carpathian Basin

Wang Shen: Chinese auroral reports during Maunder minimum

Both contain informative data on early events in terrestrial/solar-terrestrial physics. It is interesting to note that the number of auroral reports peaked even during the Maunder-minimum in supposed solar maximum years, nevertheless, they were only local maxima, as the number of observations was some years after the maximum in average higher.

History of geophysics (and physics) is represented by a number of biographical papers:

Silverman: Gustav Schübler: an early nineteenth century geophysicist

Schröder: Ein handschriftlicher Lebenslauf von Emil Wiechert

Unterreitmeier: Rudolf Straubel und Beginn der seismischen Registrierungen in Jena

Debarbat: Laplace and the system of the world

Meyenn: Einstein und die Emigranten in Princeton

Thieme: Nestor der Weberforschung (Karl-Heinrich Wiederkehr)

Meyenn's paper deals also with the group of Hungarian physicists and mathematicians in the US, called then the "Hungarian phenomenon".

Other historical contributions deal with the history of science in an institution or in a country:

Oni: Pioneering physics, geodesy and geophysics research in Nigeria

Kopeczky: 90 years since the first astronomical observation at the Ondrejov Observatory

Some papers deal with geophysics and geology from a non-historical point of view.

Ádám: Mantle diapir - mantle plumes in the Pannonian Basin

Filling: Über die leuchtenden Nachtwolken

Jehl and Barszczus: Über die Entstehung der ozeanischen Phosphatlagerstätten und die Herkunft des in diesem gebundenen Phosphors

Two papers deal with past and future climatological changes:

Landscheidt: Klimavorhersage mit astronomischen Mitteln?

Gregori: Historical data and global change

A rather strong section of the book is devoted to education in a broad sense:

Hörz: Von der Idee zum Weltbild - Anmerkungen zum deutschen Abituraufsatz von Helmholtz

Berg: Die Physik als unverzichtbarer Bestandteil unserer Kultur und Bildung

Moritz: Der Lorenz-Attraktor als didaktisches Modell für die magnetischen Polsprünge der Erde

Klein: Froebels Spielgaben und die romantische Naturphilosophie

Küveler: Der naturwissenschaftliche Roman

Hörz analyses Helmholtz's thesis for the maturity examination about Lessing's Nathan der Weise, together with his father's and professor's remarks, and traces his later humanistic ideas in this early work. The comparison with the father's remarks is especially interesting: two persons of similar moral attitude, but separated by a generation met in this discussion, some 150 years before present.

Some historical works from other branches of science should also be mentioned.

Schreier: Strukturen in der deutschsprachigen Historiographie der Physik zwischen 1800 und etwa 1920

Nicolson: Humboldtian plant geography and its legacy for ecology

Mersich: Einführung des Schilaufs in Mitteleuropa

Zott: Über Preise in der Wissenschaft

Poggemann: Das idealistisch–romantische Werk "Anfangsgründe der Stöchiometrie" und dessen philosophische Fundamente in der christlich-platonischen Physiktheologie

Gregori: Renaissance and history of science, The case of construction theory

A number of papers deal with physical problems.

De Sabbata: Quantum gravity in real space time

Sofonea: Mutations idéatiques dans l'histoire de la physique marquées par l'introduction et l'interprétation des constantes physiques universelles

Liritzis: Elementary particles expansion

Schalk: Kosmos und Planckion zur Algebraisierung physikalischer Probleme

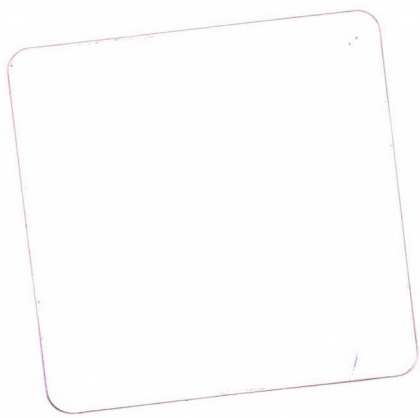
Fischer: Sind wir in einem schwarzen Loch?

Saint-Guily: Trkalian fields and gyroscopic waves in rotating fluids

All these papers, together with Gregori's praise and Neubauer's congratulating words, Thieme's mentioned contribution on one side of Wiederkehr's activity and the list of Wiederkehr's publications make this volume an interesting piece of reading for the representatives of several fields of science and are a worthy homage to a significant scientist.

J Verő

PRINTED IN HUNGARY
Akadémiai Nyomda, Martonvásár



INSTRUCTIONS TO AUTHORS

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

Manuscripts should be typewritten in duplicate or on disk (LATEX*.tex files or Word for Windows Documents or text-only ASCII files) with a written copy, or via E-mail (LATEX*.tex files or text-only ASCII files).

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

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

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

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

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

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

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

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

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

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

5. *Legends* should be short and clear. The place of 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.

7. All the *illustrations* should contain the figure number and author's name in pencil on the reverse. The most important point with figures is clearness. Photographs and half-tone illustrations should be sharp and well contrasted. Colour photographs will be accepted, but the extra cost of reproduction in colour must be borne by the authors (approx. US$ 260 per page).

Only original papers will be published and a copy of the Publishing Agreement will be sent to the authors of papers accepted for publication. Manuscripts will be processed only after receiving the signed copy of the agreement. Information is sent to the first author if no other wish is expressed. Fifty offprints are supplied free of charge, for further copies please contact the Editorial Office.

Acta Geodaetica et Geophysica Hungarica

VOLUME 32, NUMBERS 3–4, 1997

PROCEEDINGS OF THE INTERNATIONAL BEACON SATELLITE SYMPOSIUM 97

SOPRON, 30 JUNE–5 JULY 1997
HUNGARY
PART I

GUEST EDITORS
P BENCZE and R LEITINGER

Akadémiai Kiadó, Budapest

AGG 32 (3–4) 263–479 (1997) HU ISSN 1217–8977

ACTA GEODAETICA ET GEOPHYSICA HUNGARICA
A Quarterly Journal of the Hungarian Academy of Sciences

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

AGG is published in yearly volumes of four issues by

AKADÉMIAI KIADÓ
H-1117 Budapest, Prielle K. u. 19–35

Manuscripts and editorial correspondence should be addressed to

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

Subscription information

Orders should be addressed to

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

Subscription price for Volume 32 (1997) in 4 issues US$ 112.00, including normal postage, airmail delivery US$ 20.00.

Proceedings of the International Beacon Satellite Symposium 97 Sopron, 30 June – 5 July, 1997 Hungary

Part I

Editors

P Bencze and R Leitinger

CONTENTS

New methods, experiments and ideas

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 263–263 (1997)

Preface

The International Beacon Satellite Symposium 97 was organized by the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences in Sopron, from June 30 to July 5, 1997. It was sponsored by the International Union of Radio Science (URSI) through its Working Group G2. It was supported by the Hungarian Academy of Sciences and by the Hungarian National Committee of URSI.

Scientists from 14 countries participated at this Symposium and more than 50 papers were presented in course of the 5 days. There were also poster presentations and two disscussions were organized, one with the topic GPS as the most important source of TEC data now and in the immediate future, the other entitled: Satellite situation, new methods, experiments and ideas. An excursion for showing the sights of the surroundings of Sopron and a lady's programme completed the organization.

The proceedings of this Symposium are published according to an agreement of the Programme Committee and the editor-in-chief of Acta Geodaetica et Geophysica Hungarica in this journal. The papers presented at the Symposium are published in two successive volumes of the journal.

Many peoples contributed to the success of the Symposium, whose help is acknowledged with thanks. We are grateful for the support of the Geodetic and Geophysical Research Institute and whish to thank for the contribution to the success of this conference to the United Air Force, European Office of Aerospace Research and Development.

P Bencze and R Leitinger

1217-8977/97/$ 5.00

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 265–269 (1997)

THE OBSERVED VARIATIONS OF IONOSPHERIC ELECTRON CONTENT

F H HIBBERD[1]

A brief review is given of the behaviour of the electron content of the ionosphere and its variations that has been obtained from beacon satellite studies. This includes the daily variation, the day-to-day and hour-to-hour variability, the 27-day variation, the annual(seasonal), semi-annual and solar cycle variations and variations associated with geomagnetic disturbance and storms. It is remarkable that we are still far from a complete understanding of the causes of most of these phenomena.

Keywords: electron content; ionosphere

1. Introduction

It is appropriate at this time to review briefly the knowledge of the electron content of the ionosphere and its variations, in low and middle latitudes, that have been obtained from radio beacon studies and described by many workers. These variations will be illustrated using a long series of southern hemisphere data obtained at Armidale, (30S, 151E) and $L = 1.8$. Except for the summer-winter variation noted below, there are no great differences between the behaviours in the northern and southern hemispheres.

2. Daily variation and variability

Four typical plots of hourly values of electron content in various seasons and levels of solar activity are given in Fig. 1. Plotting successive hourly values for each month as in these diagrams gives a clear illustration of the daily variation and its variability. The marked day-to-day and hour-to-hour variability characteristic of the F region of the ionosphere is apparent.

The daily variation of the electron content usually attains a maximum in the early afternoon and a minimum in the early morning. A distinct change of slope in the morning increase is commonly evident in mid-morning. Subsidiary transient enhancement may sometimes occur in the late afternoon or evening and in the vicinity of midnight. A number of such increases near midnight are to be seen in the data for Jan 1978.

3. 27-day variation

There is a 27-day variation in the midday values of electron content, associated with the rotation of long-lived active regions on the sun. This is illustrated in Fig. 2, which shows the electron content varying in parallel with the solar radio noise flux $F_{10.7}$. The correlation is most clearly visible at those times when the day-to-day variability is small.

[1]Physics Department, University of New England, Armidale, NSW 2351, Australia

1217-8977/97/$ 5.00

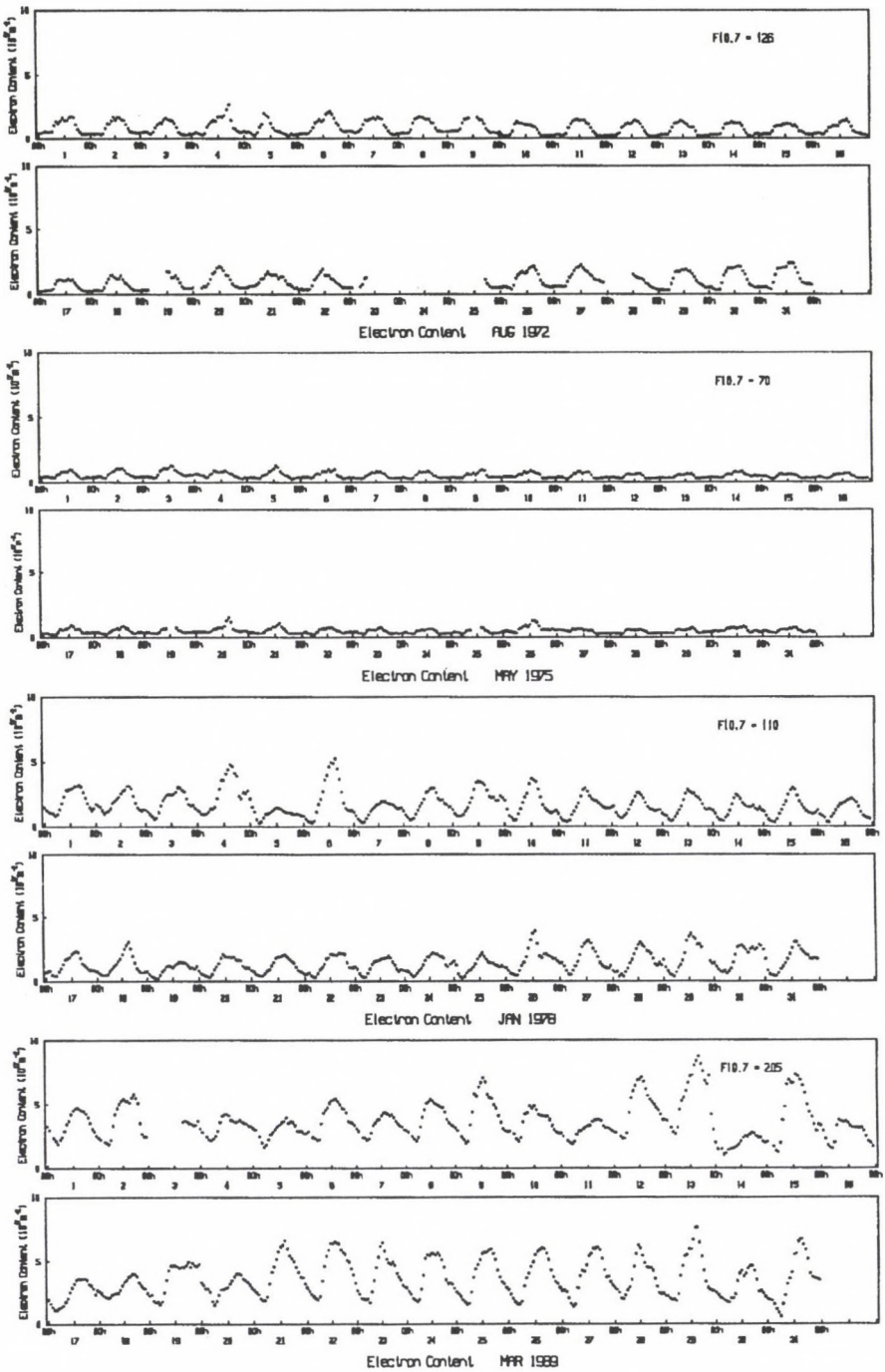


Fig. 1. The daily variation and the day-to-day and hour-to-hour variability of electron content, for different seasons and levels of solar activity

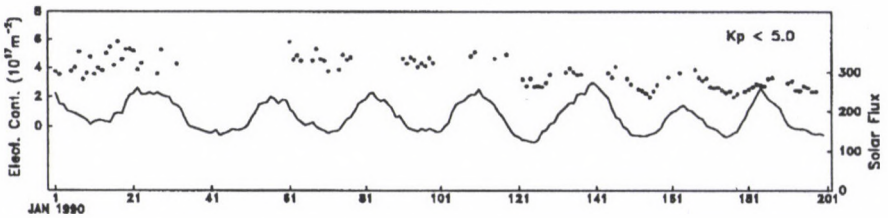


Fig. 2. Example of the 27-day variation in electron content

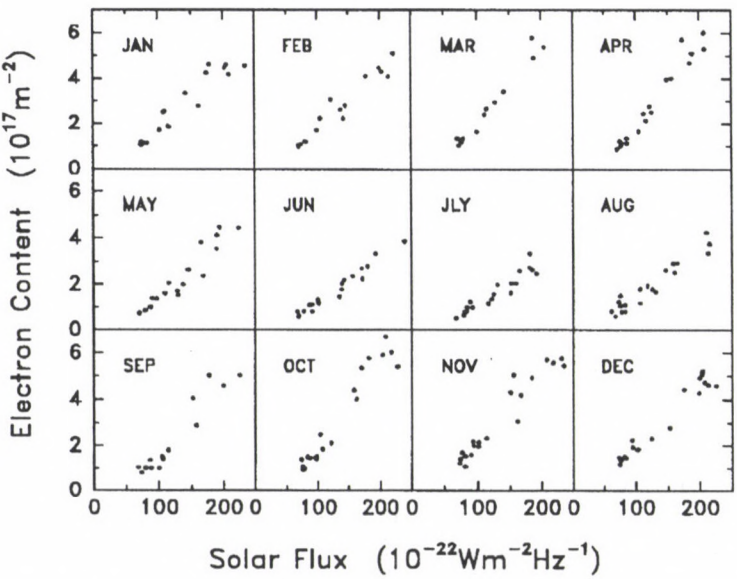


Fig. 3. Variation of electron content over the solar cycle, separately for each month

4. Solar cycle, annual and semi-annual variations

The variation of the midday electron content over the solar cycle is shown separately for each month in Fig. 3. Here the monthly median value of electron content for the same month, for each of the years 1970–1990, (almost two solar cycles) is plotted against the monthly mean value of $F_{10.7}$. The use of monthly medians or means effectively eliminates effects of the 27-day variations in the data. It is seen that for each month the electron content varies almost linearly with the solar flux $F_{10.7}$ over the solar cycle. Also, the slopes of the plots are seen to be steeper in local summer than in local winter and the slope in the equinoctial months is even greater than that in summer. These variations (for the southern hemisphere) correspond to an annual seasonal variation with its maximum in local summer, on which is superimposed a semi-annual variation with maxima near the equinoxes. In the northern hemisphere the annual seasonal variation has its maximum in local winter and minimum in local summer, except at solar minimum.

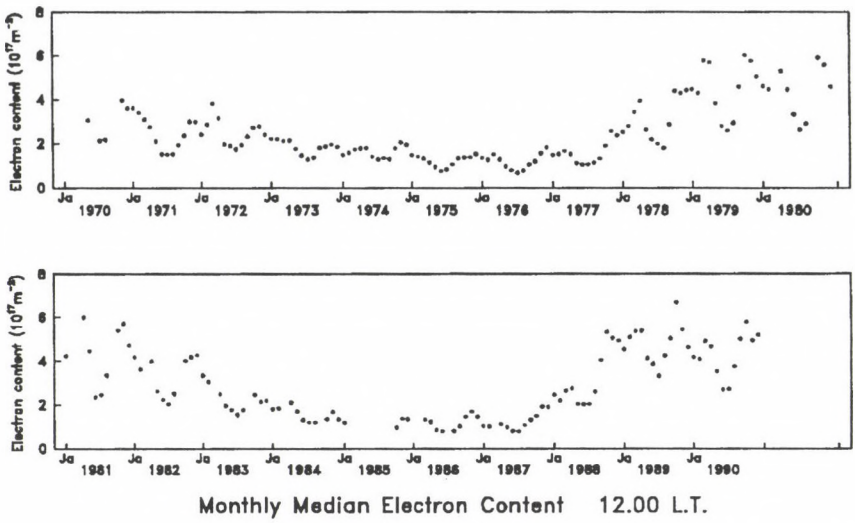


Fig. 4. Variation over the solar cycle of successive monthly medians of electron content

The data in Fig. 3 have been replotted successively month by month in Fig. 4 for the years 1970–1990 in succession. The eleven year variation over the solar cycle is obvious. The equinoxial maxima of the semi-annual variation in each year and the annual variation with the summer values exceeding the winter values in each year can be clearly seen. It is important to note that whereas in the southern hemisphere the midday values of electron content are larger in local summer than in local winter the reverse is true in the northern hemisphere except close to solar minimum.

5. Association with geomagnetic disturbances and storms

Much of the day-to-day variability seen in Fig. 1 is found to be associated with geomagnetic disturbance or storms as indicated by indices such as K_p, A_p or D_st. For instance, in Jan 1978 a sudden commencement occured on Jan 4, followed by high K_p and another s.c. on Jan 5. The midday electron content was increased on the days of the s.c.'s and severely depressed on the following days. Similar behaviour is seen in Mar 1989, with a s.c. and high K_p occurring on Mar 8–9 and a further s.c. occuring on Mar 13. Severe magnetic storms and disturbance occurred over the interval Mar 9–20. This behaviour of the electron content increasing for a day or so at the beginning of a storm followed by a severe depression lasting for one to several days is typical of a storm in summer, especially in years of high solar activity.

In winter an increase in K_p is commonly accompanied by a simultaneous increase in electron content with no subsequent decrease in content. Examples may be seen in 1975 on May 5 and 6, on May 20 and on May 26. Similarly, the increase in the afternoon of Aug 4, 1972 and the elevated content on Aug 5 and 6 are associated with severe magnetic storms and disturbance over the interval Aug 4–9 but are not followed by any severe depression in the content.

6. Comment

The variations that occur in the electron content of the ionosphere in low and middle latitudes will be recognized to be qualitatively very similar to the well-known variations in the peak electron density of the F2 region. This is not surprising. What is remarkable, however, is that the causes of most of these phenomena still remain partly or wholly unexplained.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 271–278 (1997)

LOW LATITUDE TEC VARIATIONS FOR THREE CONSECUTIVE SOLAR CYCLES

B JAYACHANDRAN[1] and T N KRISHNANKUTTY[1]

Hourly values of the total electron content (TEC) for the low latitude station of Hawaii (21.2°N, 157.7°W) during the last three consecutive solar cycles 20, 21 and 22 are used to study the long term and short term variations of TEC for this low latitude station. It is found that for perfectly quiet days, a linear variation for the daytime peak values of TEC (TEC_{max}) with $S_{10.7}$ solar flux exists even for high solar flux values of $S_{10.7}$ during the winter and equinox seasons of the three solar cycles. However, during summer a 'levelling off' effect in TEC_{max} is observed beyond about 180 units of $S_{10.7}$. The long term dependence of TEC_{max} on solar activity is found to be less sensitive during summer compared to winter and equinox for the three solar cycles. In contrast to the behaviour of TEC_{max}, the solar activity dependence of TEC_{min} (morning minimum values of TEC) for the three solar cycles is found to be strongest in summer and weakest in winter. The time difference between TEC_{max} and TEC_{min} remains more or less constant for the entire solar flux range during winter and equinox in the three solar cycles. However, during summer this time difference shows a decreasing trend with increasing solar flux indicating a 'levelling off' effect of TEC_{max} for higher solar flux values in summer. There is good correspondence in the diurnal variation of the day to day variations in TEC for the respective seasons of the solar maximum and minimum phases during the three solar cyles.

Keywords: perfectly quiet days; solar cycle; $S_{10.7}$ solar flux; solar maximum; solar minimum; solar phase; total electron content, TEC

1. Introduction

The TEC of the ionosphere is defined as the total number of electrons in a column of unit area of cross section extending from the ground to the top (nearly 2500 km in altitude) of the ionosphere (Titheridge 1972). Being a measure of the total ionization of the ionosphere, the TEC has been used as a tool for many ionospheric studies and has been measured routinely since 1964 using geostationary satellites (Davies 1980). The short term (day to day) and the long term seasonal and solar activity variations under different geophysical conditions have been discussed by various authors (Balan et al. 1994 and references therein). From these studies it is known that TEC at any location depends on the time of the day, season, solar and magnetic activity conditions. The diurnal and seasonal variations of TEC may be similar from one solar cycle to another but the solar and magnetic activity dependencies can differ from one solar cycle to another, depending on the levels of these activities in the solar cycles, particularly during the solar maximum phases. Hence in order to get a comprehensive picture of the long term variations of TEC it seems essential to study it by using a long data base covering as many solar cycles as possible. Most of the studies conducted todate on the various aspects of

[1]Department of Physics, Sree Narayana College, Kollam 691 001, Kerala, India

1217-8977/97/$ 5.00

TEC variations for low latitudes have been limited to the use of data for one solar cycle or even less (Davies 1980). However, Jayachandran et al. (1992) have given a detailed description of TEC and F region peak electron density variations for the low latitude station of Hawaii for two consecutive solar cycles 20 and 21. In this paper we present the results of a comprehensive study of TEC variations for the above station during the last three solar cycles 20, 21 and 22. While solar cycle 20 is an average one, the 21st solar cycle is the second strongest cycle recorded so far. The 22nd cycle is also nearly as strong as the 21st solar cycle.

2. Data and method of analysis

Hourly TEC data derived at Hawaii from Faraday rotation measurements of VHF telemetry transmission from different geostationary satellites during the period 1964–1989 form the database of the present study (World Data Centre, USA). The solar flux at 10.7 cm ($S_{10.7}$) and the magnetic activity index A_p are used to represent the solar and magnetic activities respectively. In this study while days with $A > 16$ are considered highly disturbed (D) (magnetically), days with $A_p \leq 10$ are taken as quiet days. Further, the days with $A_p \leq 10$ but preceded or followed by days with $Ap > 10$ are considered as disturbed quiet (DQ) and the quiet days which are bounded by quiet days only are called perfectly quiet days (PQ). The months of November, December, January, February; March, April, September, October and May, June, July, August are respectively taken as winter, equinox and summer seasons.

3. Results and discussion

The mean diurnal variations of TEC for perfectly quiet days during the winter, equinox and summer seasons for the three solar maxima (1969, 1981 and 1989) and for the two solar minima (1965, 1985) years in the three solar cycles are shown in Fig. 1. It can be seen that the mean diurnal TEC variation for the same seasons of the three solar cycles are qualitatively similar for both the solar maximum and minimum phases. However, there are quantitative difference in the TEC values for the same seasons of both solar maximum and minimum phases of the three solar cycles, which is more prominent in the daytime peak values of winter and equinox seasons and is in tune with the changes in the $S_{10.7}$ solar flux values.

Figure 2 shows the scatter plots of the day time peak values of TEC_{max} with the corresponding $S_{10.7}$ solar flux values for the D days DQ days and PQ days of the winter season in the 21st solar cycle. The figure clearly shows that for PQ days the variation of TEC_{max} with $S_{10.7}$ is linear during the entire flux range. However, for both DQ and D days this linearity is not evident. For both these groups of days the TEC_{max} values are scattered more for the higher flux range with more values being scattered to the lower side of TEC_{max} values. Similar trends are observed for the equinox season of the 21st solar cycle and also for the winter and equinox seasons of the 20th and 22nd solar cycles.

The solar activity variations of TEC_{max} for PQ days for the three seasons of the three solar cycles are shown in Fig. 3. Each point in the figure represents the mean of all values falling within ten units of $S_{10.7}$. A linear variation of TEC_{max} with

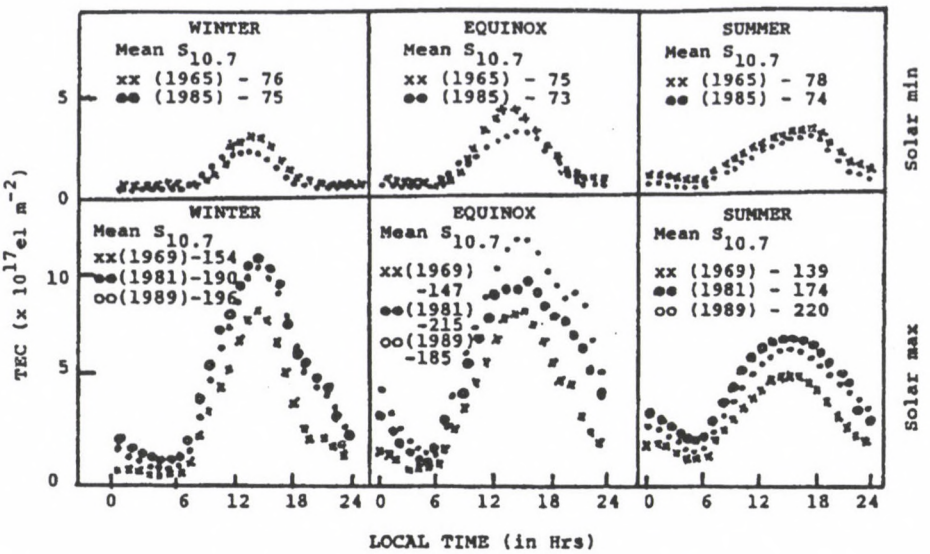


Fig. 1. Mean diurnal variations of TEC during the three seasons of winter, summer and equinox for the two solar minima (1965 × × × and 1985 ● ● ●) and for the three solar maxima (1969 × × ×, 1981 ● ● ● and o o o 1989) years

solar flux exists for the entire flux ranges for perfectly quiet days during winter and equinox for the three solar cycles, whereas in summer a levelling off effect in TEC_{max} is observed beyond about 180 units of $S_{10.7}$. Besides, for about the same solar flux values, during the winter and equinox seasons, the stronger 21st and 22nd solar cycles register higher values of TEC_{max} than the moderate 20th cycle in the higher solar flux ranges. Figure 4a is the same as Fig. 3 but for the morning minimum values. In contrast to the behaviour of TEC_{max}, the solar activity dependence on TEC_{min} is strongest in summer and weakest in winter for the three solar cycles. Figure 4b gives the time difference between the noontime maximum and the morning minimum values of TEC. This time difference remains more or less constant for the entire solar flux range for both the winter and equinox seasons. However, during summer this time difference seems to be reduced for higher flux values, indicating the 'levelling off' effect in TEC_{max} values during the summer season for higher flux values.

The day to day variability of any parameter is best described by the ratio of the standard deviation to the monthly mean value. Figure 5 gives the diurnal variation of such ratios expressed in percentage for the three seasons in the solar maximum and minimum phases. The excellent similarities in the percentage variation of TEC during the respective seasons of both solar phases of the three solar cycles permit us to draw an average picture of the seasonal behaviour of the day to day TEC variability. Table I gives the average of the mean daytime (0800–1600) and nighttime (2000–0400) percentage variations of TEC for the three seasons in the two solar phases, considering the three solar maximum years together and also the three

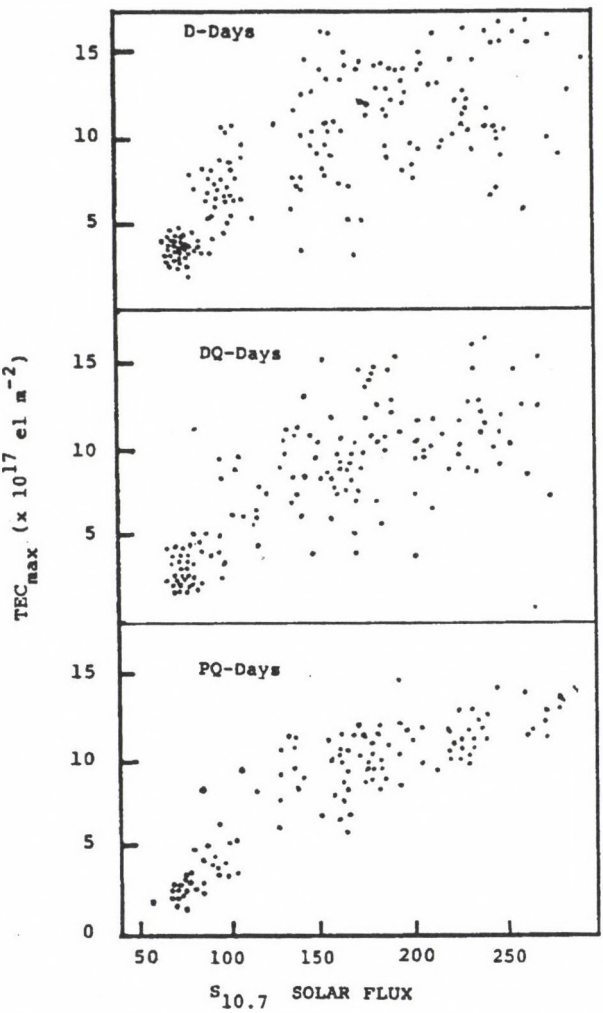


Fig. 2. Scatter plot of TEC_{max} values with $S_{10.7}$ solar flux for magnetically disturbed, disturbed quiet and perfectly quiet days during winter season of the 21st solar cycle

solar minimum years together. The table shows that the mean daytime variability is less than the corresponding mean nighttime variability during all the seasons of solar maximum and minimum phases, except during the solar minimum winter for which the daytime variability is higher than that of the nighttime. During the solar maximum the mean daytime TEC variability does not show any significant seasonal dependence. However, the mean nighttime variabilities are highest in winter and lowest in summer. During the solar minimum, while the daytime variations are highest in winter and lowest in summer, the nighttime variations are highest in equinox and lowest in winter.

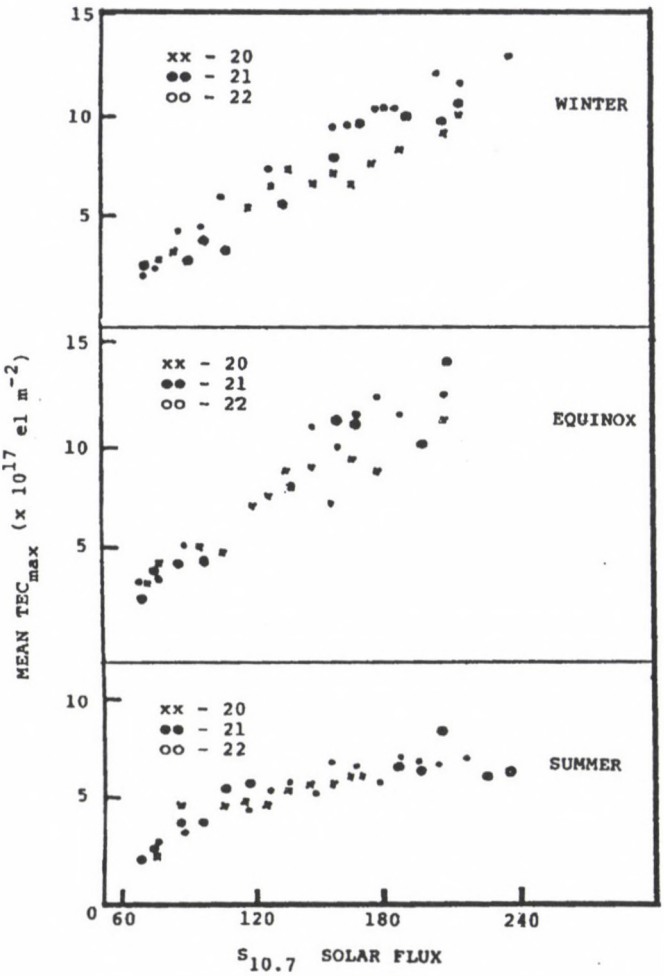


Fig. 3. Solar activity variations of mean TEC_{max} (for perfectly quiet days) during the three seasons of the three solar cycles (× × × – 20, ● ● ● – 21 and o o o – 22)

The TEC variations in the higher flux ranges above 200 units of $S_{10.7}$ have been investigated by many authors. Bhuyan et al. (1983) report that for an Indian low latitude station TEC decreases with $S_{10.7}$ solar flux after increasing linearly upto about 200 units of $S_{10.7}$. Balan et al. (1994) report a saturation in TEC_{max} values for higher solar flux values above 200 units of $S_{10.7}$ for the same low latitude station of Hawaii. Bhuyan and Tyagi (1994) report a saturation effect in TEC_{max} for the higher solar flux values for the low latitude station of Lumping. It is to be noted that these investigators have included all the quiet days in their analysis irrespective of whether they are preceded or followed by disturbed days and so the storm effect has not been eliminated completely. However, Doherty and Klobuchar (1994) observed that for the mid latitude station of Hamilton the saturation effect in TEC_{max} for

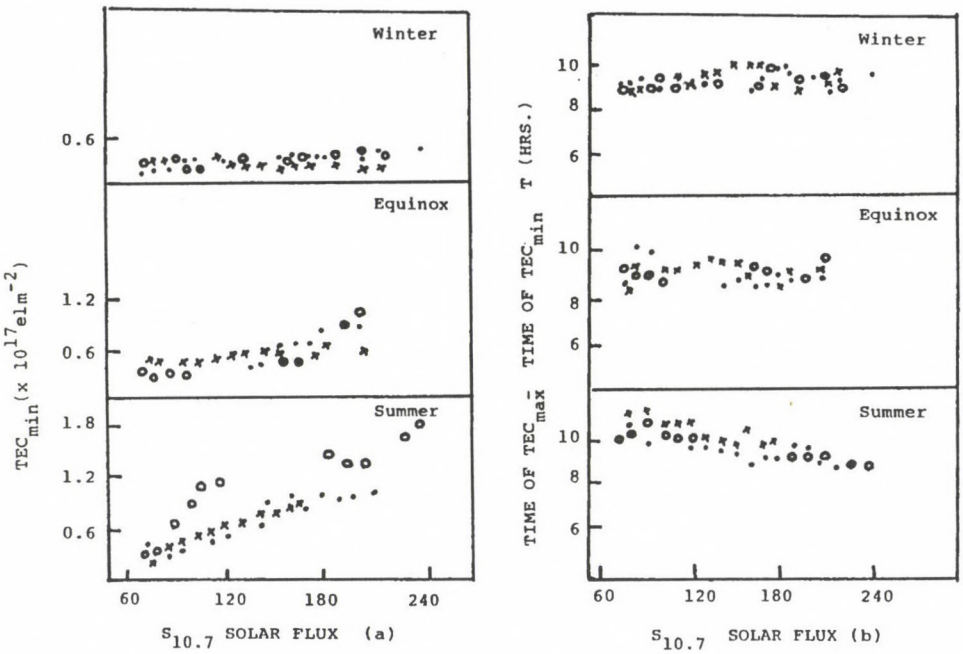


Fig. 4. a) Same as Fig. 3 but TEC_{min}. b) The time difference between the TEC_{max} and TEC_{min} for the three seasons with $S_{10.7}$ solar flux for the three solar cycles

Table I. Mean daytime (0800–1600) and nighttime (2000–0400) percentage variations of TEC for the three seasons in the two solar phases

Seasons	Percentage deviation during			
	Solar maximum		Solar minimum	
	Day	Night	Day	Night
Winter	24	50	30	22
Equinox	22	40	25	40
Summer	23	30	20	30

higher solar flux values is not apparent. In their study to eliminate the storm effects they have excluded disturbed days ($A_p > 10$) and also any quiet day that follows a disturbed day. The present study clearly shows that the exclusion of the D and DQ days from the analysis results in a linear variation of TEC_{max} with solar flux even for higher solar flux values above 200 units of $S_{10.7}$. Thus it is concluded that the magnetic activity plays a dominant role in producing the saturation effect in TEC reported by the earlier investigators.

The higher nighttime variability of TEC compared to daytime found in the present study has been reported from other stations also (Klobuchar 1983). The

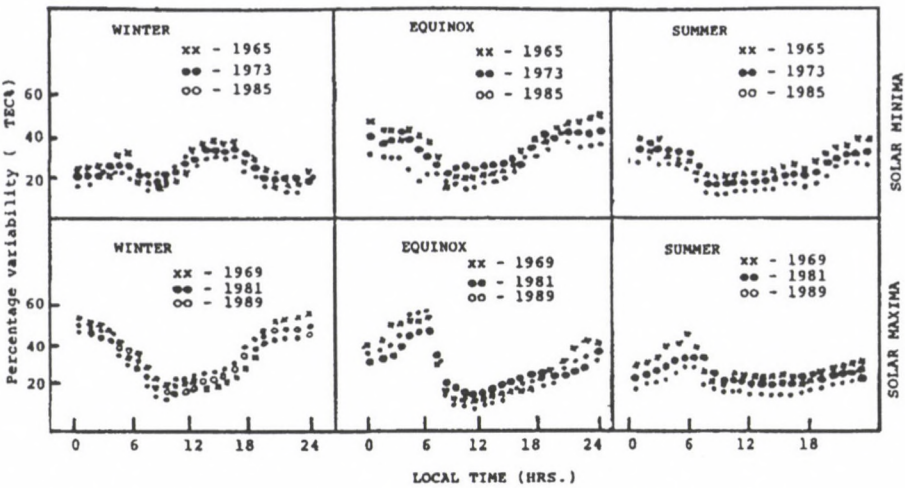


Fig. 5. Mean diurnal variations of percentage standard deviations of TEC (σ TEC%) for the three seasons during the solar maximum and minimum years of the three solar cycles (× × × – 20, ● ● ● – 21 and o o o – 22)

lower winter nighttime percentage of TEC variability observed in the study may be due to the fact that in the winter hemisphere the field aligned plasma flow is from the protonosphere to ionosphere, particularly during post midnight hours and the duration of this flow is reported to be decreasing with increase of solar activity (Evans and Holt 1978). Large downward flow of H^+ during the winter nighttime can decrease the $O^+ - H^+$ transition level, giving large increase in the scale height of the topside ionosphere which could appreciably increase the measured electron content values (Titheridge 1973) and thus lowering the variability ratio.

In recent years fairly complex models have been developed for simulating the ionosphere for any specified geophysical condition. In order to advance the models further for accurate simulation on long term, there are two basic requirements. First is the accurate knowledge of the model inputs such as solar flux, temperature, winds and electric fields for any geophysical condition and the second is an extensive base of ionospheric data against which the models can be tested for their validity. The sort of the results presented here are expected to aid in someway to fullfil the second requirement.

References

Balan N, Bailey G J, Jenkins B, Rao P B, Moffett R J 1994: *J. Geoph. Res.*, 99, 2243–2253.

Bhuyan P K, Tyagi T R 1994: In: Proc. URSI – Int. Beacon Satellite Symposium, University of Wales, U.K., 205–208.

Bhuyan P K, Tyagi T R, Singh L, Somayajulu Y V 1983: *Ind. J. Radio Space Phys.*, 12, 84–93.

Davies K 1980: *Space Sci. Rev.*, 25, 357–430.

Doherty P H, Klobuchar J A 1994: In: Proc. URSI – Int. Beacon Satellite Symposium, University of Wales, U.K., 220–223.

Evans J V, Holt J M 1978: *Planet. Space Sci.*, 26, 727–744.

Jayachandran B, Balan N, Rao P B 1992: In: Proc. URSI – Int. Beacon Satellite Symposium, Cambridge, MA, USA, 220–223.

Klobuchar J A 1983: In: Proc. URSI – Int. Beacon Satellite Symposium, NPL, Delhi, 3–12.

Titheridge J E 1972: *Planet. Space Sci.*, 20, 353–373.

Titheridge J E 1973: *J. Atm. Terr. Phys.*, 35, 981–1001.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 279–285 (1997)

SHORT TERM IONOSPHERIC VARIABILITY IN THE NORTHERN HEMISPHERE FOR TWO CONSECUTIVE SOLAR CYCLES

B JAYACHANDRAN[1], T N KRISHNANKUTTY[1], M RAJENDRAPRASAD[1]

Hourly values of the total electron content (TEC-Faraday content) collected at five stations in the northern hemisphere extending from the low to high latitudes lying in the narrow longitude zone of (270 ± 20)°E during the solar maximum years 1981 and 1989 respectively of the last 21 and 22nd solar cycles and also during the common solar minimum year 1985 are used to study the latitudinal variations of the day to day TEC variability. The effect of solar phase change in the variability is found to be the least for the midlatitude station of Boulder and most during the nighttime for the high latitude station of Goose bay. The magnetic disturbance generally enhances the TEC variability for the five stations during the three seasons of both solar phases and this effect seems to be somewhat higher for the mid and high latitudes during the equinox season of the solar maximum years. A spectral investigation of the TEC variability reveals narrowing of the spectrum from low to high latitudes during the solar minimum phase while the reverse seems to be the case during the solar maximum. While the low latitude TEC variability spectrum is not correlated to either the variability spectrum of $S_{10.7}$ solar flux or to the mean equatorial D_{st} variations, the mid latitude TEC variations show a good positive correlation with the variability of A_p and the higher latitude TEC variations are negatively correlated with the variability spectrum of the daily mean Auroral electrojet (AE) indices.

Keywords: Auroral electrojet; solar cycle; solar maximum; solar minimum; solar phase; $S_{10.7}$ solar flux; TEC variability; total electron content, TEC

1. Introduction

The earth's ionosphere exhibits considerable day to day variability in one of its widely measured parameter viz. the total electron content (TEC). A study of the day to day variability in TEC assumes importance because of its influence in HF and VHF radio communication networks. Day to day variations in TEC under different geophysical conditions have been studied by a number of investigators (Jayachandran et al. 1995 and references therein) and tried to correlate this phenomenon of ionosphere variability to changes in EUV fluxes (Jayachandran et al. 1995), electron production and loss rate (Titheridge 1974), local atmospheric conditions in the thermosphere (Rama Rao et al. 1980), effect of neutral winds (Kane 1980), magnetic activity (Aravindan and Iyer 1990) etc. However, most of the earlier studies were limited to single stations and/or to restricted solar activity variation. In this paper we present the results obtained from a study of the day to day TEC variations using simultaneous TEC data collected at five stations in the northern hemisphere extending from the low to high latitudes in a narrow longitude zone of (270 ± 20)°E during the solar maximum years 1981 and 1989 and during

[1]Department of Physics, Sree Narayana College, Kollam 691 001, Kerala, India

1217-8977/97/$ 5.00

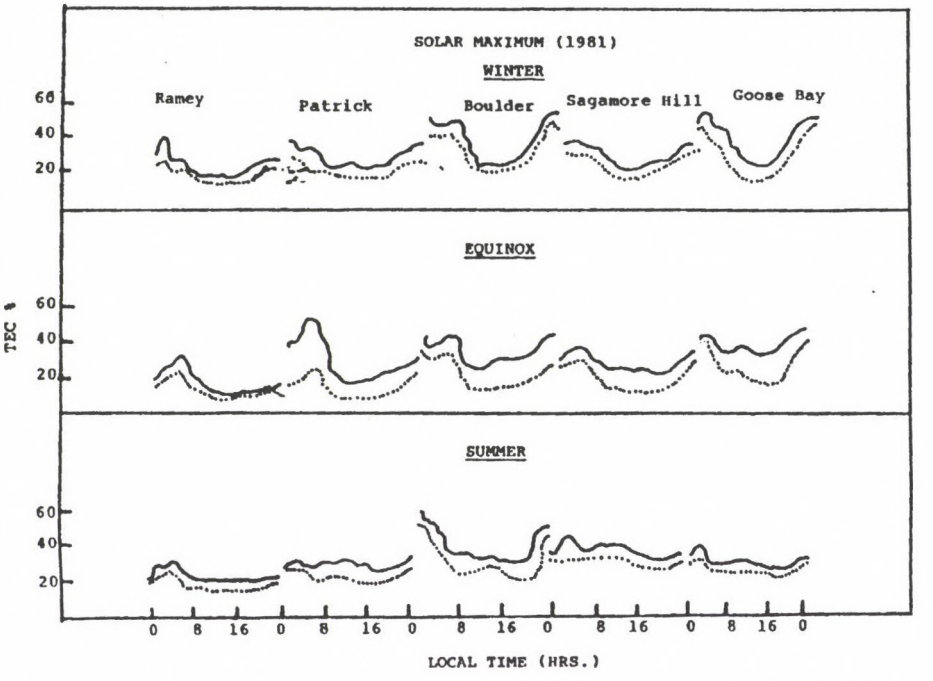


Fig. 1. Mean diurnal variations of percentage standard deviations of TEC (σ% TEC) for the five stations during the winter, equinox and summer seasons of solar maximum year 1981 (— represents for all days and ··· represents for quiet days alone)

the common solar minimum year 1985 of the last two strong solar cycle 21 and 22. The effects of solar and magnetic activities in the latitudinal variations of TEC variability are investigated.

2. Data and method of analysis

The hourly values of TEC (Faraday rotation measurements of VHF telemetry transmission from different geostationary satellites) collected at the five stations of Ramey (28.7°N, 289°E), Patrik (37.2°N, 267°E) Boulder (44.4°N, 251°E), Sagamore Hill (50.0°N, 283°E) and Goose bay (58.6°N, 286°E) during the solar maximum years 1981 and 1989 and also during the solar minimum year 1985 form the data base of the present study (World Data Centre, USA). Day to day variations in TEC at any local time of a given month is described by the variability parameter (in percentage) and is defined as the ratio of the standard deviation of TEC at that time to the corresponding monthly mean value. The months of January, February, November, December, March, April, September, October and May, June, July, August are taken respectively as the winter, equinox and summer seasons. For a comparison of TEC_{max} (noon time peak) variations for the above five stations during the solar maximum and minimum phases with the corresponding variations of $S_{10.7}$ solar flux, mean equatorial D_{st} values, A_p indices and mean AE indices,

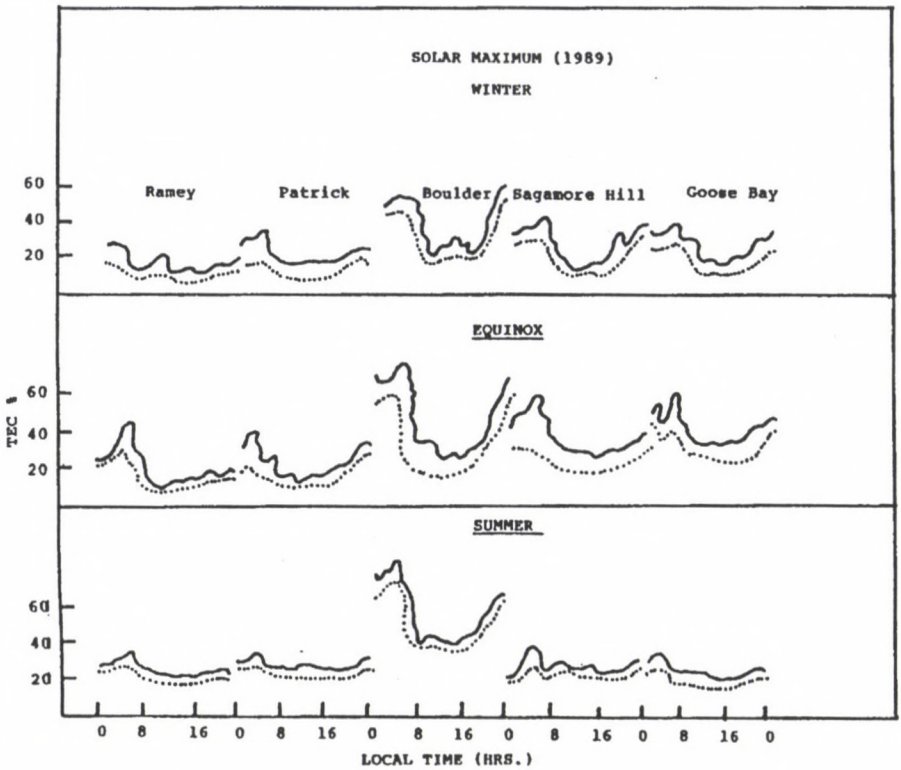


Fig. 2. Same as Fig. 1, but for the solar maximum year 1989

the daily variations of all these parameters from their respective monthly values are computed and subjected to spectral analysis and the prominent periodicities above −6 dB levels are compared.

3. Results and discussion

Figures 1 and 2 represent the seasonal mean curves of the diurnal variation (on hourly basis) of the ratios of the standard deviation to the monthly averages expressed in percentage for TEC during the solar maximum years 1981 and 1989 respectively. The continuous curves represent the variability parameter computed by taking all the days while the dotted curves represent the variability parameter computed by taking only the quiet days. It can be seen that there exists excellent similarities in the diurnal TEC variability curves for the two solar maximum years (1981 and 1989) for the respective stations during the corresponding seasons. During the solar maximum years, the nighttime variability is higher compared to the daytime variability for all the five stations in the three seasons. Both daytime and nighttime variabilities are lower for stations in the low latitude regions (Ramey and Patrick) compared to the mid (Boulder) and high latitude (Goose bay) stations.

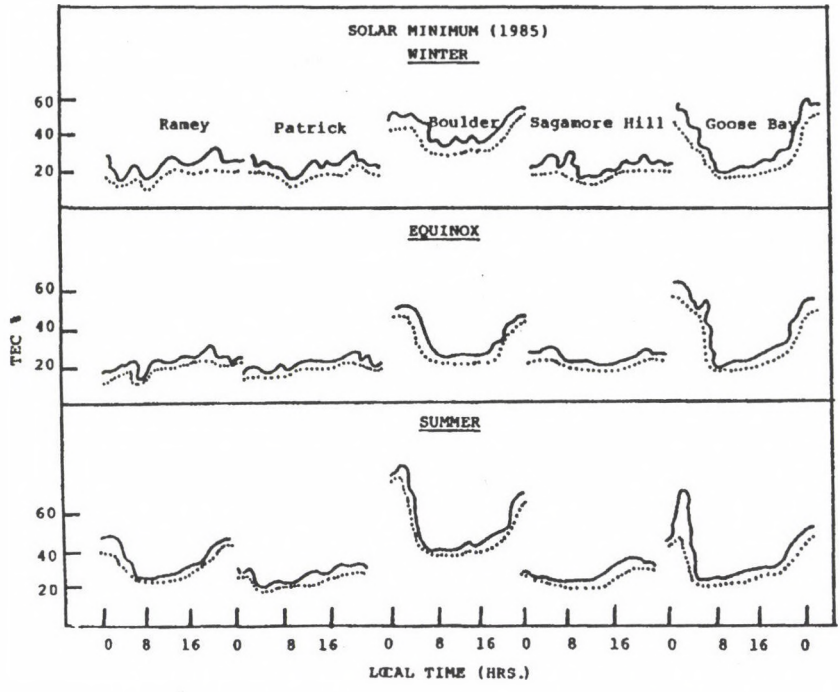


Fig. 3. Same as Fig. 1, but for the solar minimum year 1985

The difference between the daytime and nighttime variabilities is minimum during summer for the different stations except for the mid latitude station of Boulder for which the nighttime variability is much higher compared to the daytime variability during summer also. Figure 3 is the same as Fig. 1 but for the solar minimum year 1985. It is seen that during the solar minimum phase also, in general, the nighttime variability is higher than the corresponding daytime variability for the five stations. However, the difference between the daytime and nighttime variabilities is reduced for the different stations compared to that during the solar maximum phase except for the mid and subauroral stations of Boulder and Goose bay respectively. The effect of solar phase change in TEC variability is the least for the mid latitude station of Boulder for the three seasons and most for the high latitude station of Goose bay during the nighttime. The effect of magnetic disturbance is, in general, to enhance the TEC variability for the five stations during the three seasons of both solar phases and this effect seems to be somewhat higher for the mid and high latitude stations during the equinox season for the solar maximum years.

Figure 4 gives the power spectra of the percentage variations of TEC_{max} (day time peak) from the corresponding monthly mean values for the five stations during the solar maximum (1981) and solar minimum (1985) years. The prominent periodicities above −6 dB power level are shown in the figure. It is interesting to note that for all the five stations, for both the solar phases, the periodicities of the stronger components, in general, correspond to the periods of rotation of the sun (27 or 36

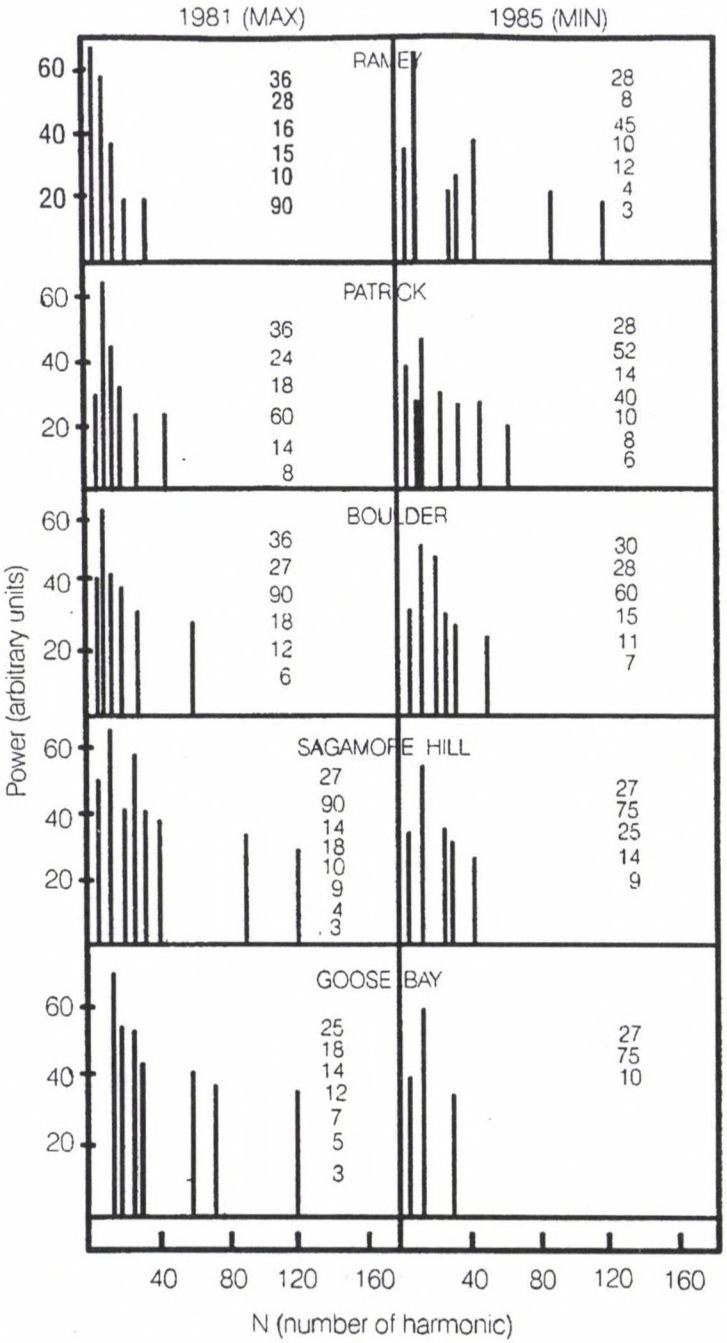


Fig. 4. Power spectra of ΔTEC_{max} for the 5 stations during the solar maximum year 1981 and solar minimum year 1985 (periodicities in days are shown in the figure)

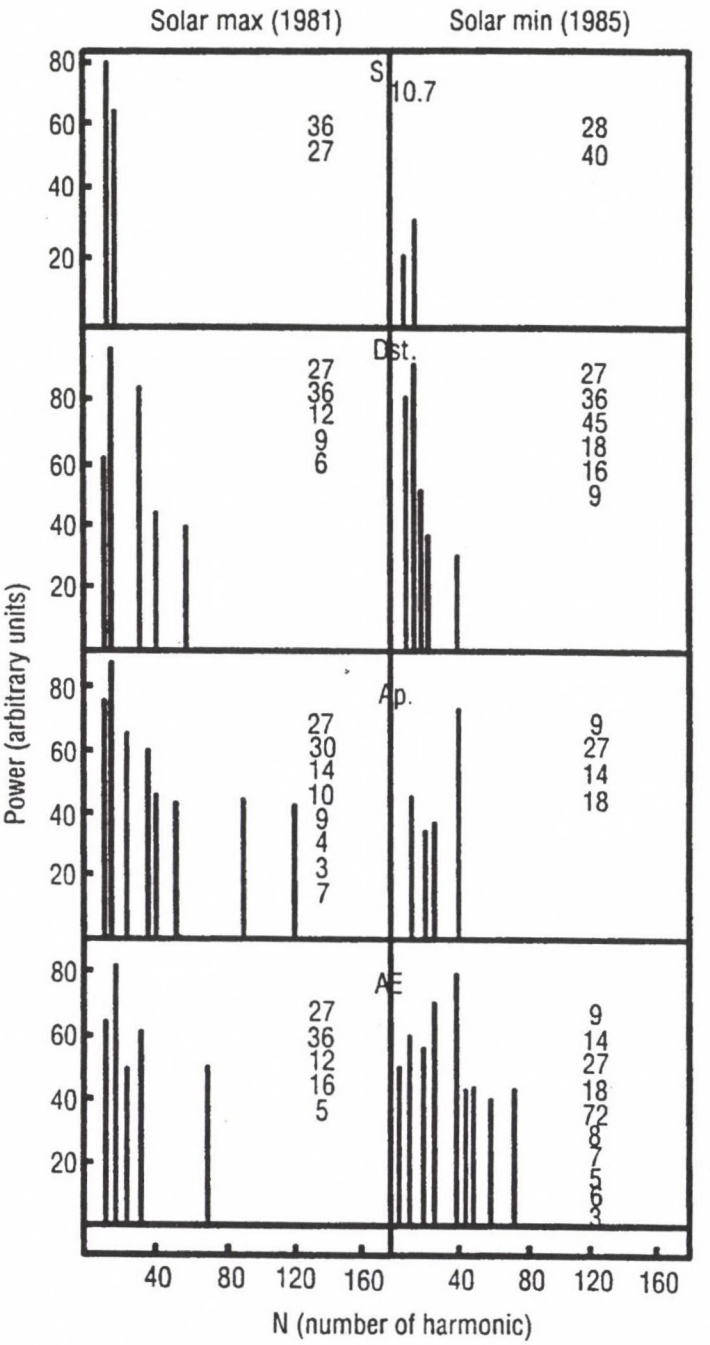


Fig. 5. Power spectra of $\Delta S_{10.7}$, ΔD_{st}, ΔA_p and ΔAE during 1981 (solar maximum) and during 1985 (solar minimum) (periodicities in days are shown in the figure)

days). It can also be seen that the spectrum is being narrowed from low to high latitudes during the solar minimum phase while the reverse seems to be true for the solar maximum year. Figure 5 gives the spectral picture of the variation of $S_{10.7}$ solar flux, mean equatorial D_{st} values A_p and AE indices for the solar maximum (1981) and solar minimum (1985) years. It can be seen that while the low latitude TEC variability spectrum (shorter period less than 15 days) is not correlated either to the $S_{10.7}$ variability spectrum or to the variability spectrum of mean equatorial D_{st} variations, the midlatitude TEC variations show good positive correlation with the A_p variations and the high latitude TEC variations are negatively correlated with the variability spectrum of AE.

The day to day variation for the different latitudes presented in this study are in general agreement with early studies for each latitude sector with relative small data base (e.g. Aravindan and Iyer 1990). But, none of the earlier studies has brought out the solar and magnetic dependencies of the TEC variability for the different latitudes as done in the present study. However, a detailed study of TEC variability involving the day to day changes in the solar EUV fluxes, magnetic activity variations, thermospheric compositions and dynamics is essential for the proper understanding of the TEC day to day variations.

References

Aravindan P, Iyer K N 1990: *Planet Space Sci.*, 38, 743–750.

Jayachandran B, Nair R B, Balan N, Rao P B 1995: *J. Atm. Terr. Phys.*, 57, 1599–1609.

Kane R P 1980: *Radio Sci.*, 15, 837–842.

Rama Rao, Nru D, Rao Srirama M 1980: In: Proc. COSPAR-URSI Symposium, Warsaw, Poland, 175–187.

Titheridge J E 1974: *J. Atm. Terr. Phys.*, 36, 1249–1257.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 287–292 (1997)

INFLUENCE OF SOLAR WIND ON THE TEC VARIATIONS AT MID AND SUB-AURORAL LATITUDES DURING SUNSPOT MAXIMUM

T E GIRISH[1], B JAYACHANDRAN[2], SAJAN P SHAMSUDEEN[1]

One cannot find many studies on the relationship between the amount of energy transferred in to the earth's magnetosphere and ionospheric TEC variations observed at mid or sub-auroral latitudes. In this work we have attempted to relate day-to-day variations of four well known solar wind-magnetosphere energy coupling parameters and daily mean TEC variations at a mid latitude (Sagamore Hill) and a sub-auroral (Goose bay) station during the year 1981, at sunspot maximum. We could obtain statistically significant correlation between the variations of the above solar wind parameters and TEC at Goose bay and Sagamore Hill during the months February (winter) and March (equinox) for the year of study. However we cannot find a similar result for August (summer) 1981. This study demonstrates the usefullness of solar wind-magnetospheric coupling parameters in modelling ionospheric TEC variations during high solar activity periods.

Keywords: mid and sub-auroral latitudes; solar wind-ionosphere relations; sunspot maximum; TEC variations

1. Introduction

There are several studies on the relationship between ionospheric total electron content (TEC) variations and solar 10.7 flux or geomagnetic activity variations (e.g. Aravindakshan and Iyer 1993). There are also studies on the variations of TEC in association with different solar wind conditions such as different range of solar wind speed and directions of Interplanetary Magnetic Field (IMF) components B_y and B_z, observed by earth-orbiting satellites (e.g. Bremer 1996). However one cannot find many studies (Girish et al. 1994) on the relationship between variations in the amount of solar wind energy transferred in to the earth's magnetosphere and TEC variations at mid and sub-auroral latitudes. In this work we have made such a study on the relationship between day-to-day variations of four well known solar wind--magnetosphere energy coupling parameters reported in literature (e.g. Gonzalez et al. 1994) and day-to-day variations in daily mean TEC at the stations Sagamore Hill (mid latitude) and Goose bay (sub-auroral) during the year 1981, at sunspot maximum.

2. Data used and method of analysis

a) TEC parameters

TEC data for the American stations Sagamore Hill (mag. lat. 50.2°N, 294°E long) and Goose bay (mag. lat. 59.2°N, 297°E long) obtained using telemetry transmissions from geostationary satellites (published by World Data Centre – A)

[1]Department of Physics, University College, Trivandrum 695 034, Kerala, India
[2]Department of Physics, S.N. College, Kollam 691 001, Kerala, India

1217-8977/97/$ 5.00

for the months February, March and August in the year 1981 is used for our study. From hourly observations of TEC we have determined daily (24 hourly in UT) mean values of TEC for Sagamore Hill and Goose bay for the above period.

b) Solar wind-magnetosphere energy coupling parameters

Hourly values of the solar wind and IMF parameters observed by earth orbiting satellites for the above months (Couzens and King 1986) is used to calculate hourly values of four well-known solar wind-magnetosphere energy coupling parameters (Gonzalez et al. 1994).

i) $\varepsilon = VB^2 \sin^4(\Theta/2)\, I_0^2$ representing power delivered by the solar wind-magnetosphere MHD dynamo

ii) $K = nV^2$, a parameter proportional to kinetic energy of the solar wind flow near earth

iii) VB_s, a parameter representing solar wind electric field which is convected in to the earth's magnetosphere and

iv) $F = (nV^2)^{1/6} VB \sin^4(\Theta/2)$.

Here V is the solar wind velocity, n is the solar wind particle number density, B_y is the east-west component of IMF in GSM system, B_s is the southward component of IMF in GSM system, B is average IMF magnitude.

Also $I_0 = 7R_E$

$$\Theta = \tan^{-1}(|B_y|/|B_z|)$$

when $B_z > 0$ and

$$\Theta = 180 - \tan^{-1}(|B_y|/|B_z|)$$

when $B_z < 0$.

In addition to this daily mean A_p values and solar 10.7 cm flux values are also used.

c) Correlation between solar wind, TEC and solar 10.7 cm flux variations

We have calculated monthly correlation coefficients between the variations of daily mean values of four Solar wind-Magnetosphere Energy coupling (SME hereafter) parameters, geomagnetic A_p indices and solar 10.7 cm flux with the variations of daily mean TEC values of Sagamore Hill and Goose bay. This is done for the months February (winter), March (equinox) and August (summer) respectively for the year 1981. The values of correlation coefficients are shown in Table I. Correlation coefficients which are statistically significant at the 95% confidence levels are given in bold figures.

Table I. Correlation coefficients between variations of TEC at Sagamore Hill (SMH) and Goose bay (GB) with variations of solar wind parameters, 10.7 cm flux and A_p for February, March and August 1981

Solar wind parameter	February Av. TEC		March Av. TEC		August Av. TEC	
	SMH	GB	SMH	GB	SMH	GB
ε	**0.47**	**−0.45**	**−0.42**	**−0.39**	−0.14	−0.24
K	**0.384**	0.17	−0.18	−0.25	0.13	0.27
VB_z	**0.481**	**−0.56**	**−0.66**	**−0.49**	0.32	0.09
F	**0.52**	**−0.5**	**−0.39**	**−0.49**	−0.23	−0.3
10.7 flux	**0.5**	−0.08	−0.29	**−0.42**	**0.78**	**0.62**
A_p	**0.48**	**−0.59**	−0.34	**−0.48**	0.06	−0.09

3. Discussion

Earlier studies on the ionospheric response to solar wind and IMF variations observed near earth is generally based up on superposed epoch analysis, which can suggest only an average physical picture of the problem (for example see Bremer 1996). But in the present study we have attempted to correlate day-to-day variations of four well known solar wind-magnetosphere energy coupling parameter with the similar variations of TEC at mid latitude (Sagamore Hill) and sub-auroral (Goose bay) stations during sunspot maximum. Our study demonstrates the usefullness of solar wind-magnetosphere energy coupling parameters similar to A_p and $S_{10.7}$ cm flux in modelling ionospheric TEC variations during high solar activity periods.

For the month of February (winter) we can find statistically significant correlation between TEC variations at Goose bay (negative correlation) and SME parameters (ε, VB_z and F). But the correlation of the same with solar 10.7 cm flux variations is not significant. In Fig. 1 we have shown variations of daily mean values of a) TEC at Goose bay and b) solar 10.7 cm flux variations during February, 1981 c) SME parameter VB_s. We find two distinct enhancements of the solar wind parameter VB_s during Feb 6th and Feb 25th respectively. We notice statistically significant decrease of the daily mean TEC values of Goose bay (from its monthly mean at 1σ level) on days Feb 6th and Feb 26th respectively. These TEC decreases occur on geomagnetically disturbed days ($A_p = 48$ on Feb 6th and $A_p = 28$ on Feb 26th) and can be considered to be solar wind induced negative ionospheric storms.

For the month of March 1981 (equinox) we can find statistically significant correlations between TEC variations at Sagamore Hill (negative correlation) and SME parameters (ε, VB_z, F). But the correlation of the same with solar 10.7 cm flux variations is not significant. In Fig. 2 we have shown variations of daily mean values of a) TEC at Sagamore Hill and b) solar 10.7 cm flux variation during February 1981 c) SME parameter VB_s. In Fig. 2 we can notice a statistically significant decrease of mean TEC of Sagamore Hill for the day March 5th in association with

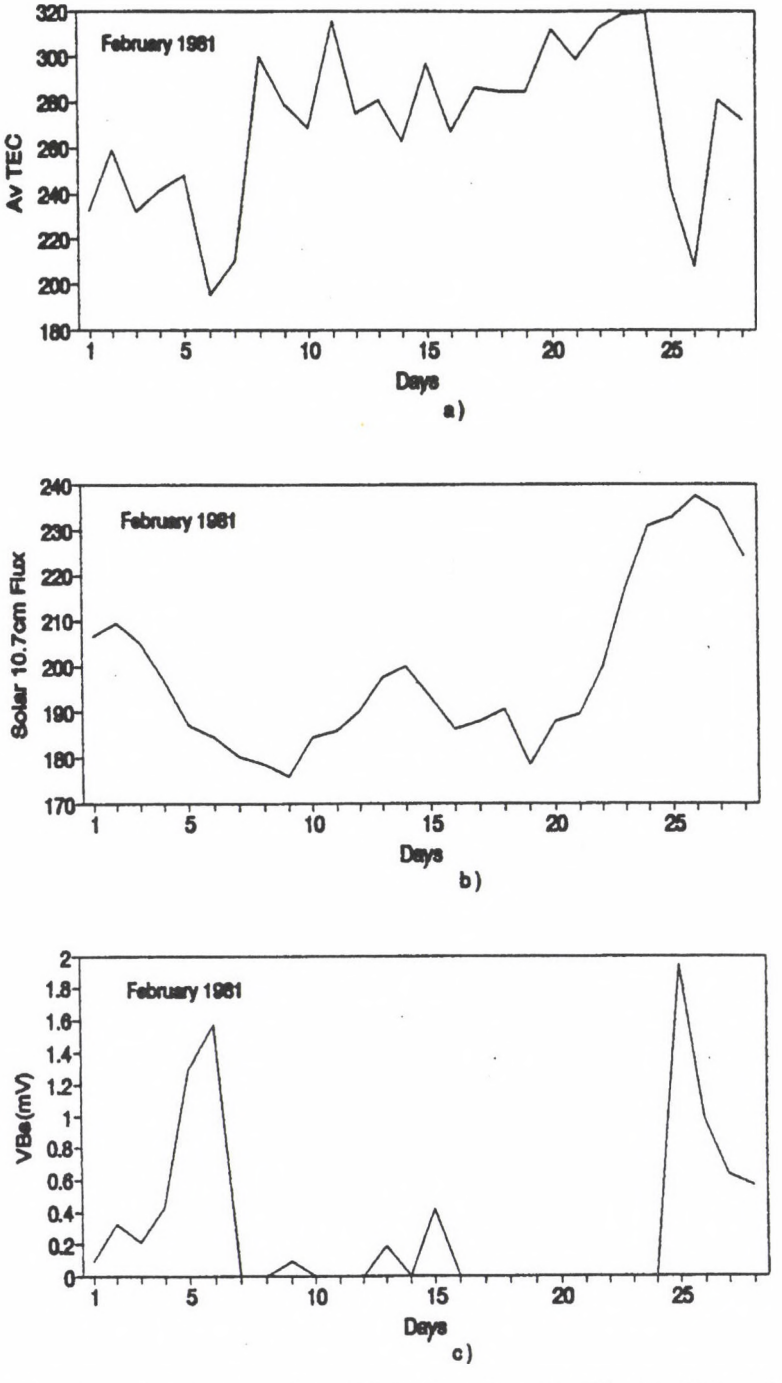


Fig. 1. Variations of daily mean values of a) TEC at Goose bay b) solar 10.7 cm flux c) solar wind-magnetosphere coupling parameter VB_s during February 1981

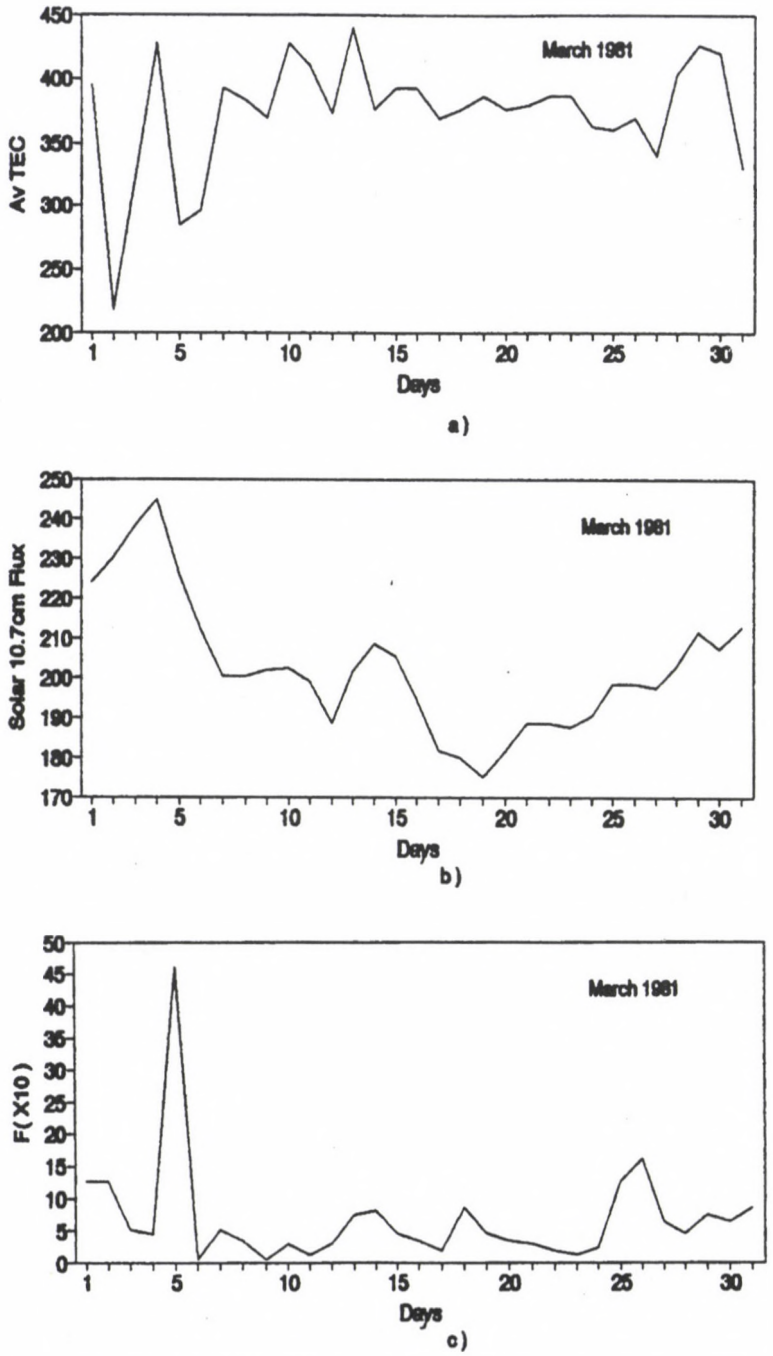


Fig. 2. Variations of daily mean values of a) TEC at Sagamore Hill b) solar 10.7 cm flux c) solar wind-magnetosphere coupling parameter F during March 1981

a significant enhancement in SME parameter F on the same day. It is interesting to find that solar flux variations show a steep increase beyond 230 units during the same period. Thus the decrease of TEC on March 5th at Sagamore Hill ($A_p = 81$) can be considered to a solar wind induced negative ionospheric storm.

During the month August 1981 (summer) we cannot observe statistically significant correlations between TEC variations at both stations (Sagamore Hill and Goose bay) and SME parameters. However, one can find high correlation between the TEC variations with the $S_{10.7}$ cm flux variations during the same month. Thus we infer that solar wind influence on ionospheric physical properties can change depending on the season (Bremer 1996).

4. Conclusions

1. During a sunspot maximum year 1981, inspite of the observation of large values of solar 10.7 cm flux (average values $\sim$ 200 units) we could obtain statistically significant correlations between variations of daily mean values of TEC at the stations Sagamore Hill and Goose bay and solar wind-magnetosphere energy coupling parameters for the months February (winter) and March (equinox). However, we could not obtain a similar result for the month August 1981 (summer).

2. The above results suggest that solar wind-magnetosphere coupling parameters can also be used for modelling TEC variations during winter/equinox at mid and sub-auroral latitudes.

References

Aravindakshan P, Iyer K N 1993: *J. Atm. Terr. Phys.*, 55, 1565–1573.

Bremer J 1996: *J. Atm. Terr. Phys.*, 58, 845–854.

Couzens D A, King J H 1986: Interplanetary medium data book – Supplement 3A, NSSDC, NASA Goddard Space Flight Centre, Maryland

Girish T E, Jayachandran B, Nair R B 1994: In: Proc National Space Science Symposium, IS 26, Trivandrum, India, 393–394.

Gonzalez W D, Joselyn J A, Kamide Y, Kroehl H W, Rostoker G, Tsurutani B T, Vasyliunas V M 1994: *J. Geoph. Res.*, 99, 5771–5783.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 293–299 (1997)

DAY-TO-DAY VARIABILITY OF IONOSPHERIC ELECTRON CONTENT AT LOW LATITUDES DURING MINIMA OF 20TH AND 21ST SOLAR CYCLE

SUSHIL KUMAR[1], S D MISHRA[1], S K VIJAY[1], A K GWAL[1]

The day-to-day variability in ionospheric electron content (IEC) has been studied using hourly values of IEC at a number of low latitude stations in Indian, American and Pacific sectors during the minima of 20th and 21st solar cycle. It is found that variability is higher in American sector as compared to that in Indian and Pacific sectors. The variability is significantly larger during E- and D-months then that during J-months at all the stations. The solar activity (Rz) is found to have significant control on the day-to-day variability whereas magnetic activity (A_p) effect is not systematic. Under the geomagnetic disturbances variability is suppressed in nighttime and is enhanced in daytime in Indian and Pacific sectors whereas in American sector the trend is reversed.

Keywords: day-to-day variability; ionospheric electron content, IEC; low latitude ionosphere; seasonal and solar cycle variation

1. Introduction

The Faraday rotation of satellite radio beacon transmissions has been used for determining the ionospheric electron content (IEC) and its variation for last several years. The IEC thus observed is that which is prevalent at subionospheric points rather than that at the location of receiving apparatus. Kane (1975) suggested that the day-to-day variability in IEC is due to erratic equatorward neutral winds that originate in polar regions intermittently even under quiet conditions, creating convective cells that result in ionospheric irregularities of scale length of about 3000 km and wander slowly around the globe. Soicher and Gorman (1981) concluded that, for mid- and high latitude stations, the day-to-day variability of IEC during day is $< 25\%$ irrespective of location, season, and solar activity while at night it is significantly higher, especially in equinoctial months. Dabas et al. (1984) observed that day-to-day changes in IEC in the form of single day abnormality, alternate day abnormality and long term periodic fluctuations of periodicity of 27 and 45 days, all of which are apparently controlled by the equatorial electrojet and are not always correlated with solar and magnetic activity. Jayachandran et al. (1995) studied the short term variability of IEC and peak electron density (NP) during solar cycle 20 and 21, at low latitude station, and reported a good correspondence in day-to-day variations of IEC and NP from one solar cycle to other for both solar maximum and minimum.

In the present paper we present results on day-to-day variability of IEC at low latitude stations: Guahati and Ahmedabad around the crest of equatorial anomaly in Indian region, Lunping around the crest of the equatorial anomaly in Pacific

[1]Space Plasma Laboratory, Department of Physics, Barkatullah University, Bhopal–462 026, India

1217-8977/97/$ 5.00

Table I. The geographic coordinates and coordinates of subionospheric points corresponding to TEC measurements

	Geographic coordinated of observing stations			Co-ordinates at subionospheric points		
	Geogr. lat.	Geogr. long.	Dip. lat.	Geogr. lat.	Geogr. long.	Dip. lat.
Ahmedabad	23.0°N	72.6° E	17.0°	21.7° N	69.8°E	17.1°N
Guahati	28.9°N	91.8° E	18.5°	23.8° N	83.6°E	18.0°N
Lunping	25.0°N	121.17° E	14.8°	23.03°N	121.9°E	12.1°N
Palehua	20.7°N	203.7° E	19.0°	20.6° N	206°E	19.6°N

sector, and Palehua beyond the norhtern anomaly crest in American sector during the minima of solar cycles 20 and 21. It may be pointed out here that in Indian region, the northern edge of anomaly crest can extend up to $26° - 27°$N geographic latitude when the crest center moves upto $17° - 18°$ from dip equator which is located at 9°N in this sector. The diurnal, monthly, seasonal and solar and magnetic activity variations of day-to-day variability parameter have been studied. The geographic coordinates and the coordinates of subionospheric points corresponding to TEC measurements of the stations are given in Table I. The stations Ahmedabad, Guahati, Lunping are in eastern zone and Palehua in western zone.

2. Experimental data and analysis

The published hourly IEC values derived from the Faraday rotation of VHF telemetry transmissions from geostationary satellite at Indian Stations, Guahati and Ahmedabad during low solar activity period of solar cycle 20 (October 1975 to July 1976) and hourly values of IEC from Lunping and Palehua during 1986 under low solar activity period of solar cycle 21 have also been used. The standard deviation $\sigma/\vec{x}$ of hourly values from monthly median value ($\vec{x}$) is determined from which ratio $\sigma/\vec{x}$ in percentage is computed for each hour of each month of observation. This $\sigma/\vec{x}$ forms the basic parameter for analysis which is a good indicator of day-to-day variability of IEC.

3. Results

3.1 Diurnal and seasonal variation of variability

Figure 1 represents the annual mean diurnal variation of day-to-day variability (σ/x) parameter in IEC in percentage for low sunspot period 1975–1976 (mean $Rz = 14$) at Indian stations and 1986 (mean $Rz = 15$) at American and Pacific sectors. The variability is larger in American sector as compared to that in Indian and Pacific sectors. The variability is low (10–25%) by day and high (40–50%) by night and this day-to-night difference is evident at all the stations. The variability reaches to an especially high peak around sunrise at Indian and Pacific stations and little later morning hours at the American station. These results are consistent with

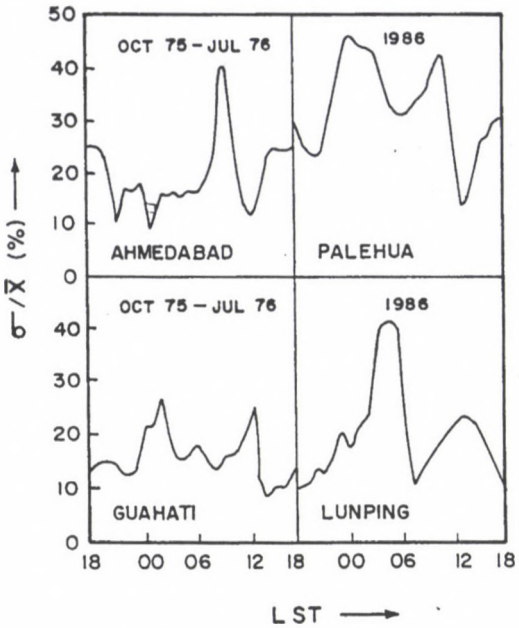


Fig. 1. Diurnal variation of day-to-day variability of ionospheric electron content as a function of local time

that reported by Aravindan and Iyer (1990) for low latitude stations. It is of interest to note that the day-to-day variability shows to maxima in American sector one around midnight hours another after sunrise whereas in Indian and Pacific sectors it shows only one maxima around sunrise. Figure 2 shows the variability in IEC during three different seasons: E-months (March, April, September, October), D-months (January, February, November, December) and J-month (May, June, July, August). It is seen that variability at any station during E- and D-months is higher than that during J-months expect for Guahati where it is smaller during E-months. The day-to-day variability is larger in nighttime during E- and D-months but during J-months it is larger in daytime.

3.2 Solar and magnetic activity effects on variability

To study the solar and magnetic activity effects on day-to-day variability, the month-to-month variation of variability in IEC has been shown in Fig. 3 wherein solar-geophysical parameters: sunspot number (Rz), and magnetic activity (A_p) have also been plotted. The variation is not smooth and ranges to about 65% in American sector and about 40% in Indian and Pacific sectors. It seems here to indicate that changes in day-to-day variability are not systematic with A_p. From the results described in Fig. 3 it is also clear that IEC variations are better correlated with Rz even during low solar activity period. However it is more evident during high solar activity. To bring out the clear picture of magnetic activity effects

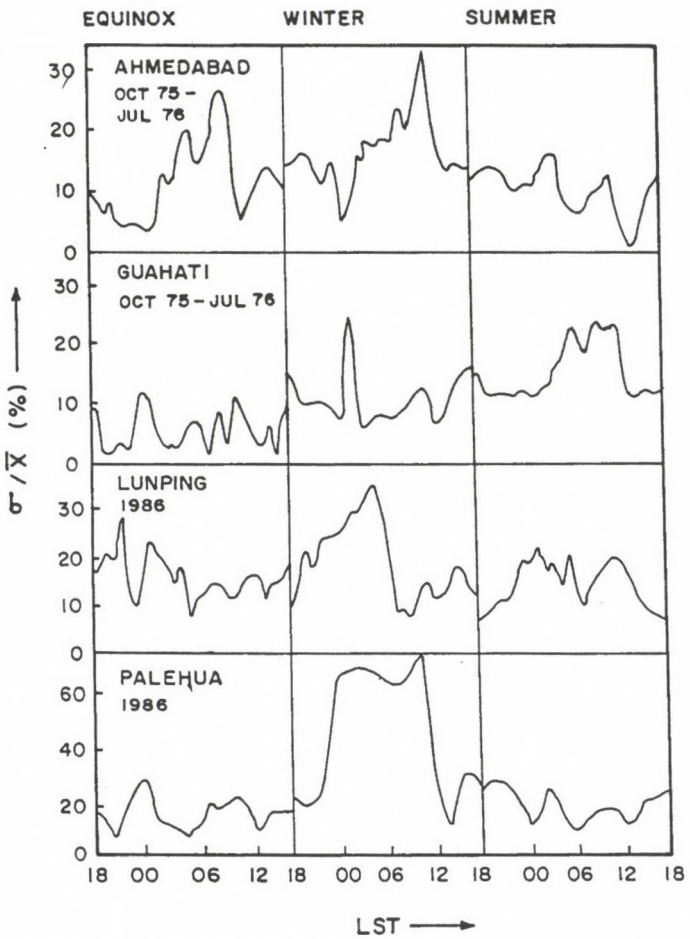


Fig. 2. Seasonal variation of day-to-day variability of ionospheric electron content as a function of local time

on variability, annual mean diurnal variability in IEC separately for almost all disturbed days (D-days) and quiet days (Q-days) for all four locations has been plotted in Fig. 4. The value of nighttime variability is suppressed on D-days in nighttime and is enhanced in daytime in Indian and Pacific sectors. In American sector the trend is reversed where the variability is enhanced on Q-days in nighttime.

4. Discussion

Total electron content (TEC) is one of the important parameters in the study of the ionospheric dynamics. One of the common feature of TEC is its day-to-day variability which is most difficult parameter to predict and affects the performance and reliability of any ionospheric prediction scheme. Many authors (e.g. Titheridge

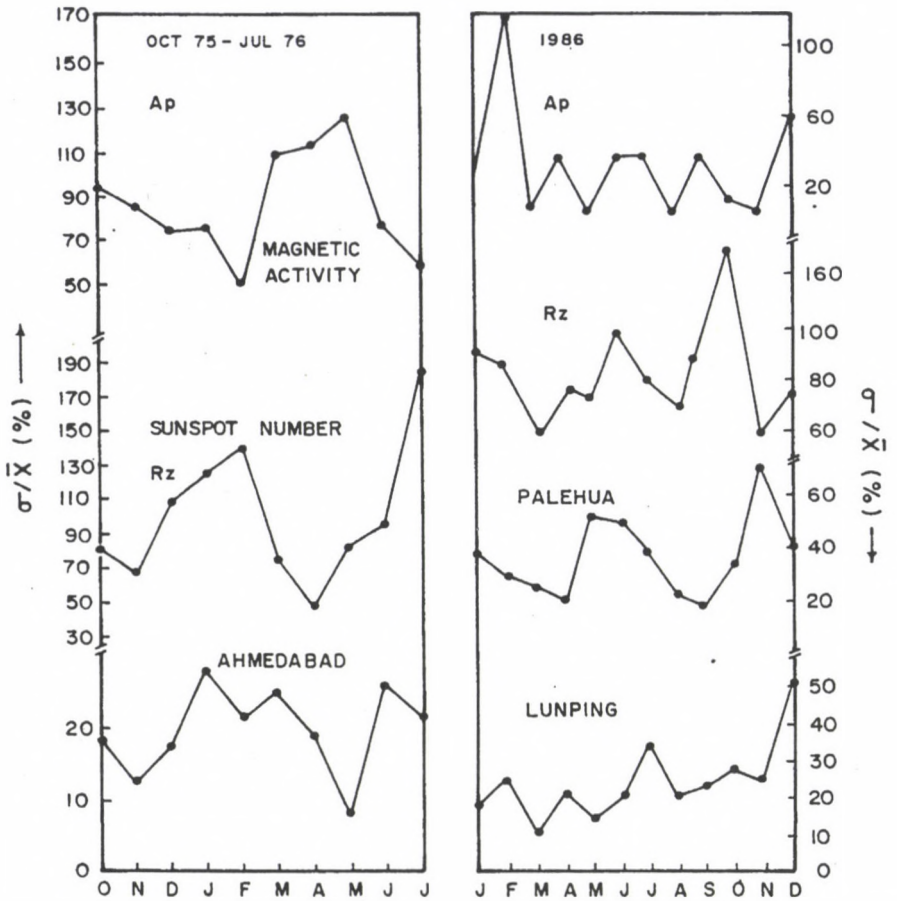


Fig. 3. Month-to-month variation of day-to-day variability against the solar geophysical parameters (sunspot number Rz, and magnetic activity A_p)

1966, Kane 1980, Rama Rao et al. 1981, Dabas et al. 1984, Soicher and Gorman 1985, Aravindakshan and Iyer 1993, Rishbeth 1993, Zarraoa and Sandron 1996) have studied the day-to-day variability. Some of them reported large day-to-day variability whereas some of them tried to correlate the variability with solar geophysical parameters. The solar EUV and X-rays are main ionising radiations that are found to undergo large-solar cycle variations. In absence of solar EUV data, other conventional solar indices such as sunspot number (Rz), 10.7 cm solar flux are normally employed in ionospheric studies. The variability in TEC depends on several factors such as local time, season, neutral wind, geomagnetic activity and solar activity, equatorial electrojet strength, temperature composition and neutral winds. The results presented here bring out that variability is larger in western sector as compared to that in eastern sector. There is a month-to-month variation and variability is higher during E- and D-months as compared to that during J-months.

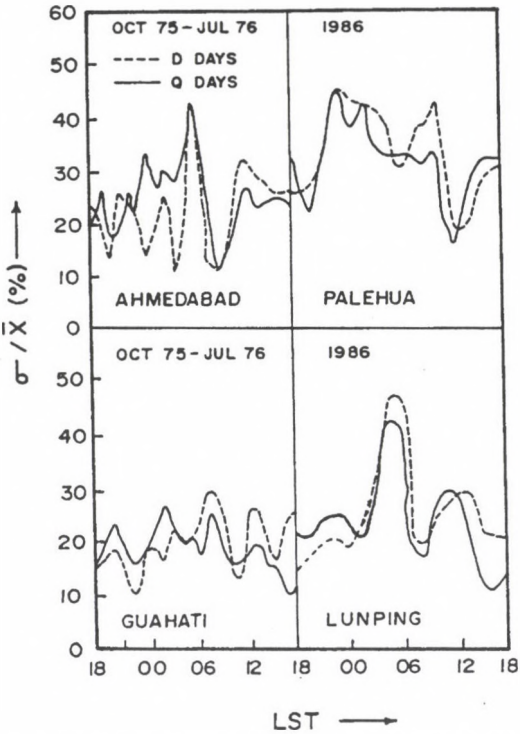


Fig. 4. Diurnal variation of day-to-day variability on five international quiet days and five international disturbed days as a function of local time

The day-to-day variability is better correlated with sunspot number even during low solar activity indicating sunspot. No systematic dependence on magnetic activity is apparent. The day-to-day variability in the equatorial region, which registers highest values of IEC and strongest large scale gradient of IEC, is essentially associated with the variations in the electrojet strenght (Dabas et al. 1984, Rastogi and Alex 1987, Aravindan and Iyer 1990). From the results it can be indicated that except the electrojet there are other sources like sunspot number which can contribute to the day-to-day variability. It is also known that electrodynamical drifts play a major role in affecting the distribution of electron densities at low latitude and its variability and they themselves are found to undergo local time, seasonal and solar and magnetic activity variations (Lakshami et al. 1995). The larger day-to-day variability in western sector may be due to the changes in electric field which result in changes in ExB drifts and subsequently changes in amount of plasma lifted up which diffuses downward to low latitude along the field lines. The difference in the ExB drift and meridional neutral wind velocity between the eastern and western longitudes may also explain the longitudinal difference found here in the day-to-day variability of IEC. Su et al. (1994) have discussed the longitudinal variations of ExB drifts and role of neutral air wind in F region ionosphere.

The variation of IEC as a function of A_p brings out only a gross picture of the effect of magnetic activity. Therefore, to obtain some systematic results the day-to-day variability of IEC has been plotted on Q- and D-days. On the basis of such analysis we can say here that under the geomagnetic disturbances the day-to-day variability is enhanced in post-midnight in Indian and Pacific sectors whereas in American sector nighttime variability is increased on Q-days. The increase in variability on Q-days as seen in Fig. 4 may be due the contribution of variability of equatorward neutral wind which is superimposed on quiet day pattern. During night this may lift the plasma upward and increase the ionisation and variability parameter.

Acknowledgements

Authors are thankful to Prof. K N Iyer for useful discussion suggestions. One of the authors (SK) is highly grateful to the Council of Scientific and Industrial Research (CSIR), Govt. of India, New Delhi, and M P Council of Science and Technology (MPCST), Bhopal, India for financial support for presenting the paper in IBSS 97.

References

Aravindan P, Iyer A 1990: *Planet Space Sci.*, 38, 743–750.

Aravindakshan P, Iyer K N 1993: *J. Atm. Terr. Phys.*, 55, 1565–1573.

Dabas R S, Bhuyan P K, Tyagi T R, Bhardwaj P K, Lal J B 1984. *Radio Sci.*, 19, 749–756.

Jayachandran B, Balachandran Nair R, Balan N, Rao P B 1995: *J. Atm. Terr. Phys.*, 57, 1599–1609.

Kane R P 1975: *Geophys. Res.*, 80, 3091–3099.

Kane R P 1980: *Radio Sci.*, 15, 837–842.

Lakshami R S, Dabas R S, Veenadhari B 1995: *Ind. J. Radio and Space Physics*, 24, 45–49.

Rama Rao P V S, Nru D, Sriram Rao M 1981: In: Proceedings of the COSPAR/URSI Symposium, A W Wernik ed., Polish Scientific Publishers, Warszawa, Poland, 91–94.

Rastogi R G, Alex S J 1987: *J. Atm. Terr. Phys.*, 49, 1133–1137.

Rishbet H 1993: *J. Atm. Terr. Phys.*, 55, 165–171.

Soicher H, Gorman F J 1981: In: Proceedings of the COSPAR/URSI Symposium, A W Wernik ed., Polish Scientific Publishers, Warszawa, Poland, 91–94.

Soicher H, Gorman F J 1985: *Radio Sci.*, 20, 383–387.

Su Y S, Bailey G J, Balan N J 1994: *J. Atm. Terr. Phys.*, 36, 1619–1628.

Titheridge J E 1966: *J. Atm. Terr. Phys.*, 28, 1135–1150.

Zarraoa N, Sandron E 1996: *Ann. Geophys.*, 14, 11–19.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 301–308 (1997)

DYNAMICS OF MEDIUM-SCALE TRAVELLING IONOSPHERIC DISTURBANCES AS DEDUCED FROM TRANSIONOSPHERIC SOUNDING DATA

E L Afraimovich[1], O N Boitman[1], E I Zhovty[1], A D Kalikhman[1], T G Pirog[1]

The diurnal and seasonal variations of medium scale travelling ionospheric disturbances (MS TID) have been studied by means of phase, angle of arrival and scintillation measurements. These investigations enabled the determination of the temporal variation of the MS TID velocity vector, thus increasing the knowledge of the temporal variation of the propagation direction.

Keywords: ionosphere; medium-scale travelling ionospheric disturbances; remote sensing; spaced reception; travelling ionospheric disturbances

1. Introduction

Travelling ionospheric disturbances (TIDs) with a typical size ranging from a few hundred km to a thousand km are of great interest as a matter for scientific enquiry in the ionosphere physics and as a factor that limits the accuracy of modern-day radio engineering systems used in navigation and radio interferometry; a rich variety of publications addressed the issue of TIDs.

This paper presents the diurnal and seasonal statistics for the dynamics of medium-scale travelling ionospheric disturbances (MS TIDs) using data from an annual run of ETS-2 radio signal polarization, angle-of-arrival and scintillation measurements at 136 MHz frequency obtained in Irkutsk (52°N, 104°E) (Afraimovich et al. 1991).

2. Determination of the dynamic characteristics of medium-scale irregularities

We obtained the seasonal-diurnal sample statistics by analyzing a set of parameters calculated from results of an hourly processing, with results presented by Afraimovich (1995).

The range 30–60 min was largely used, which lies within the generally accepted range of TID periods caused by AGW (Georges 1968, Mercier 1996) as most significant in experiments on TID radio interferometry.

In the simplest form, space-time variations in phase $\phi(x, y, t)$ of the transionospheric radio signal that are proportional to TEC variations $\Delta I(x, y, t)$ in the ionosphere, at each given time t can be represented in terms of a flat-Earth approximation of a travelling plane phase front

$$\phi(x, y, t) = k\Delta I(x, y, t) = \phi'_t(t)\Delta t + \phi'_x \Delta x + \phi'_y \Delta y \qquad (1)$$

[1]Institute of Solar-Terrestrial Physics, POB 4026, Irkutsk, 664033, Russia, Fax: +7 3952 462 557, e-mail: afra@iszf.irk.ru

1217-8977/97/$ 5.00

where k is the proportionality factor. Here it is assumed that in the case of small space-time increments (with the separation between the receiving antennas Δx, Δy much smaller than the typical spatial scale of a disturbance, and the time interval Δt between the counts is much shorter than the time scale of a disturbance) the influence of the second derivatives can be neglected.

The propagation velocity $v(t)$ and the direction $\psi(t)$ of the phase front at each given time may be defined as

$$\begin{cases} W_y(t) &= \phi'_y(t)/\phi'_t(t) = |W|\cos\psi \\ W_x(t) &= \phi'_x(t)/\phi'_t(t) = |W|\sin\psi \\ v(t) &= \dfrac{1}{\sqrt{W_x^2(t) + W_y^2(t)}} \\ \psi(t) &= \arctan\left(W_x(t)/W_y(t)\right) . \end{cases} \tag{2}$$

In practice, instantaneous values of v and ψ that are determined every 30 s, were used in our case to construct — on a selected time interval — not only distribution of the azimuth $P(\psi)$ but also the velocity $P(v)$.

In the statistics to be discussed below, to winter belong about 17 total days; to spring, summer and fall belong, respectively, 35, 20 and 43 days.

For comparing the dynamics of MS TIDs and small-scale irregularities (SSIs), we processed all scintillation records taken through spaced-antenna reception, satisfying the criteria of data standard correlation handling.

A typical example of the diurnal dependence of the magnetically quiet variations in space-time characteristics of the ETS-2 radio signal phase in the range of periods 30–60 min for November 15, 1990 is given in Fig. 1a ($K_p = 0.76$). The diurnal dependence of TEC in the vertical column (a) is a curve typical of a magnetically quiet winter day, with a day-to-night oscillation amplitude as large as 35×10^{16}el/m^2; the zero mark corresponds to a minimum value of TEC at night.

A striking difference from such an example is provided by a sample of the diurnal dependence of the signal parameters of our interest on the next day following a strong magnetic disturbance of April 11, 1990 (Fig. 1d; a mean value of the index $K_p = 5.3$). On that day the usual diurnal variation of TEC and its space-time derivatives is greatly affected: a maximum value of TEC during the daytime does not exceed 15×10^{16}el/m^2, although two days before it was on the level 30×10^{16}el/m^2.

As an example, Fig. 1 presents the diurnal dependences of instantaneous values of $\psi(t)$ (c, f) and $v(t)$ (b, e), obtained by Eq. (2) for the series of ϕ'_y, ϕ'_x and ϕ'_t for November 15, 1990 (a) and April 11, 1990 (d). We wish to note only the fact that a magnetically quiet autumn day is characterized by a regular clockwise rotation of the direction of ψ in the range of azimuths 130°–220°.

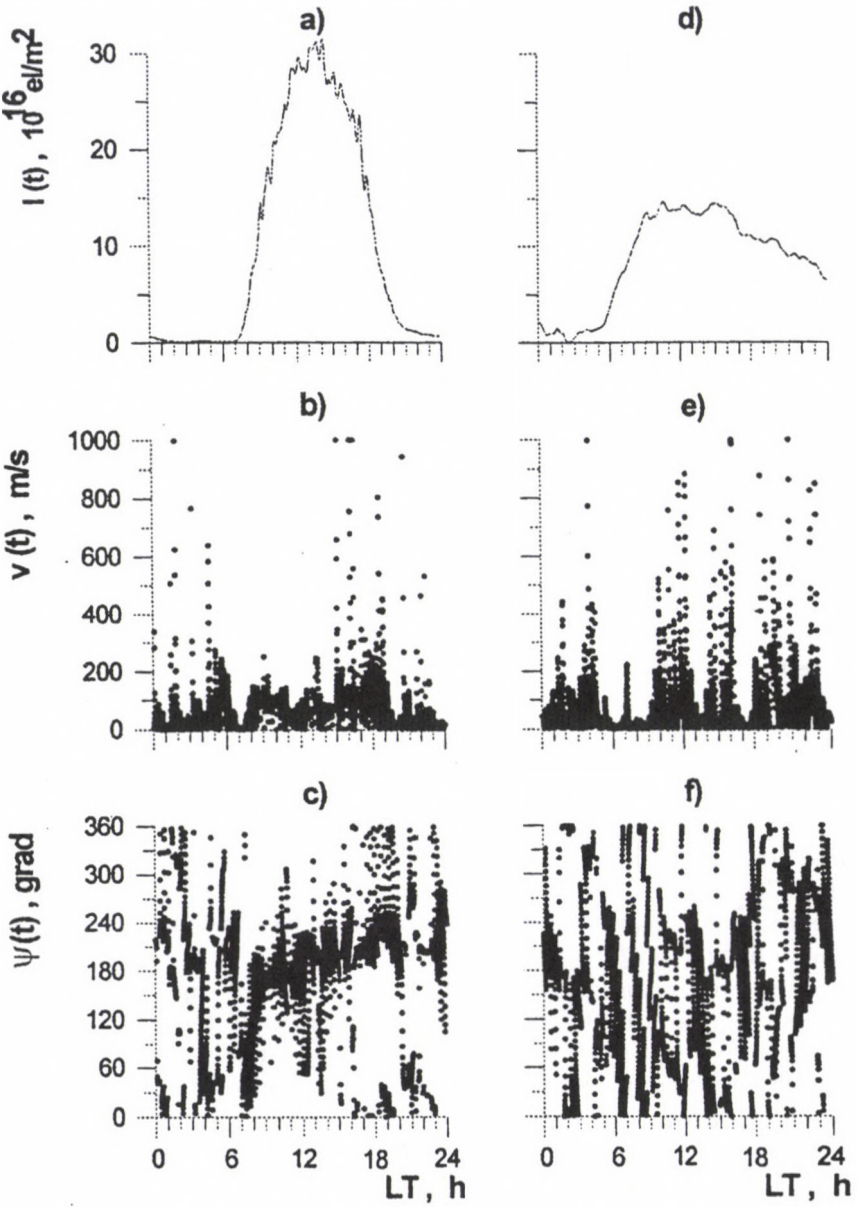


Fig. 1. Diurnal dependence of variations in space-time characteristics of the ETS-2 radio signal phase in the range of 30–60 min periods for magnetically-quiet conditions, November 15, 1990 (at the left), and during a strong magnetic disturbance of April 11, 1990 (at the right): a, d — total electron content (TEC) in a vertical column $I(t)$, 10^{16}el/m^2 ; the zero mark corresponds to a minimum value of TEC at night; b, e — velocity modulus of phase front propagation $v(t)$, m/s; c, f — azimuth of phase front propagation $\psi(t)$, grad, measured from the northward direction clockwise

3. Analysis of the TID dynamics

Figure 2 presents summary distributions of azimuths $P(\psi)$, at the left, and velocities $P(v)$, at the right, constructed a) — for the daytime; b) — for the night-time from data on phase characteristics of the ETS-2 signal for 115 days of 1989–1990; c) — for the night-time from measurements of parameters of motion of the amplitude diffraction pattern obtained by simultaneous recording of scintillations.

The $P(\psi)$-distribution for the daytime (Fig. 2a, at the left) has a symmetric form with the mean value 160° and r. m. s. 68°. It is important to note that the overwhelming number of counts lies in the range of angles 90°–230°. This result agrees well with the existing more-or-less representative statistics of measurements using ionosphere-reflected HF radio waves (Kalikhman 1980, Waldock and Jones 1986, Crowley et al. 1987) and VHF radio waves propagating through the ionosphere (Mercier 1986, 1996, Spoelstra 1992, Jacobson et al. 1995). The same can also be said about the distribution of velocities $P(v)$ — Fig. 2a, at the right, which also has a symmetric form with the mean value 133 m/s.

The $P(\psi)$ for night-time TIDs (Fig. 2b) is a symmetric two-hump distribution with a sharply defined main maximum with the most probable values of the $\psi_{\max}$ directions 30° and 210° and a very small spread of ψ in a narrow interval $\pm 10°$. In this case the direction distribution for TIDs is close to $P(\psi)$ for SSIs (Fig. 2c), whose equal 54° and 234° with the same small spread in ψ. The only difference lies in the fact that the $P(\psi)$ for SSIs is asymmetric (the south-westward direction is more pronounced when $\psi_{\max} = 234°$) and is more than $\psi_{\max}$ for MS TIDs as minimum 20°–25°.

Figure 3 compares seasonally-averaged daytime (0800–1600 LT) values of the propagation direction of the phase front ψ (•) with the neutral wind directions γ (○), calculated using our model (Zhovty and Chernigovskaya 1988): a — winter, 17 days; b — spring, 35 days; c — summer, 20 days; d — autumn, 43 days. Values of K_p for all seasons did not exceed 4.0. Diamonds ◇ denote mean values of hourly differences $\overline{\psi - \gamma}$ of propagation directions ψ of the phase front and the direction γ of the neutral wind using our model.

We see a good agreement of the experimental data and our wind model in the autumn-winter period. In this case, as was to be expected, the difference of mean values of the directions $\bar{\psi} - \bar{\gamma}$ was close to a mean value of the differences $\overline{\psi - \gamma}$. In spring-summer seasons, especially in the afternoon, we detected a noticeable discrepancy between measured and calculated (in our model) values of $\bar{\psi} - \bar{\gamma}$. However, mean values of the hourly differences $\overline{\psi - \gamma}$ were found to be not larger but smaller than values of $\bar{\psi} - \bar{\gamma}$. This is only possible if the model which we are using adequately responds to a change of geophysical parameters for every day of a seasonal sample.

Let us compare our measurements with estimates obtained in the model by Cho and Yeh (1970) in Fig. 3, these values are marked by the symbol □. It was found that unlike our model that includes geomagnetic activity, estimates the model by Cho and Yeh (1970) are closer to experimental mean values of $\bar{\psi}$ in spring and in summer, but they are at significant variance with them in the other seasons. It

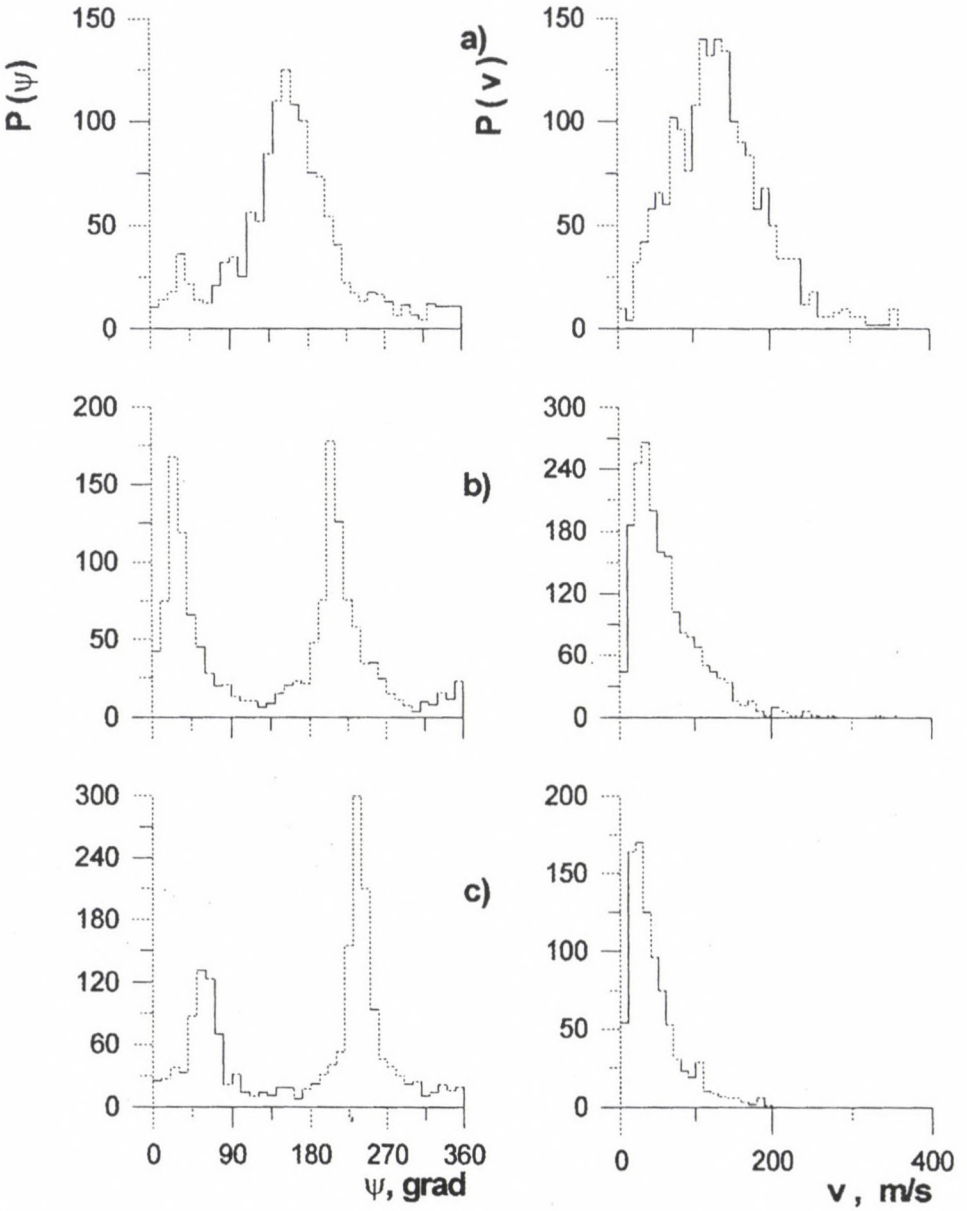


Fig. 2. Summary distributions of azimuths $P(\psi)$ — at the left, and velocities $P(v)$ — at the right, constructed for a) — the daytime (0800–1600 LT); b) — for the night-time (2000–0400 LT) based on measurements of phase characteristics of the ETS-2 signal for 115 days of 1989–1990; c) — for the night-time (2000–0400 LT) based on measurements of parameters of motion of the amplitude diffraction pattern, obtained by simultaneous recording of scintillations

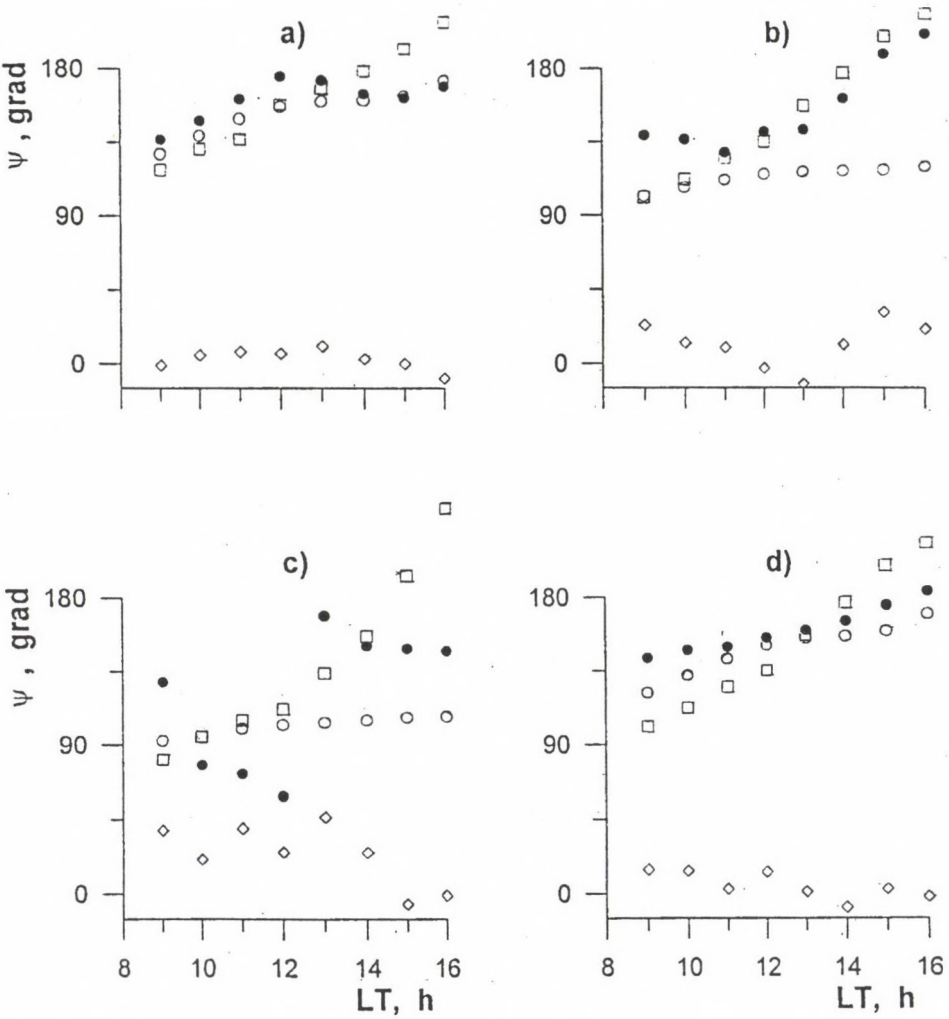


Fig. 3. Results of comparison of seasonally-averaged daytime (0800–1600 LT) values of the propagation direction ψ of the phase front (•) with the neutral wind direction γ, calculated in our model (○) and in the model by Cho and Yeh 1970 (□); a — winter, 17 days; b — spring, 35 days; c — summer, 20 days; d — autumn, 43 days. Diamonds ◇ show mean values of hourly differences $\overline{\psi - \gamma}$

should be noted here that the data examined in our paper refer to conditions of maximum solar activity when in accordance with the model by Cho and Yeh (1970) values of wind velocities are minimal. Consequently, the wind effect on AGW in the thermosphere (TIDs in the ionosphere) can also be minimal (Kalikhman 1980, Waldock and Jones 1986).

For a qualitative comparison of our results with the data reported by Jacobson et al. (1995), Fig. 4 shows a schematic representation of a polar diagram illustrating

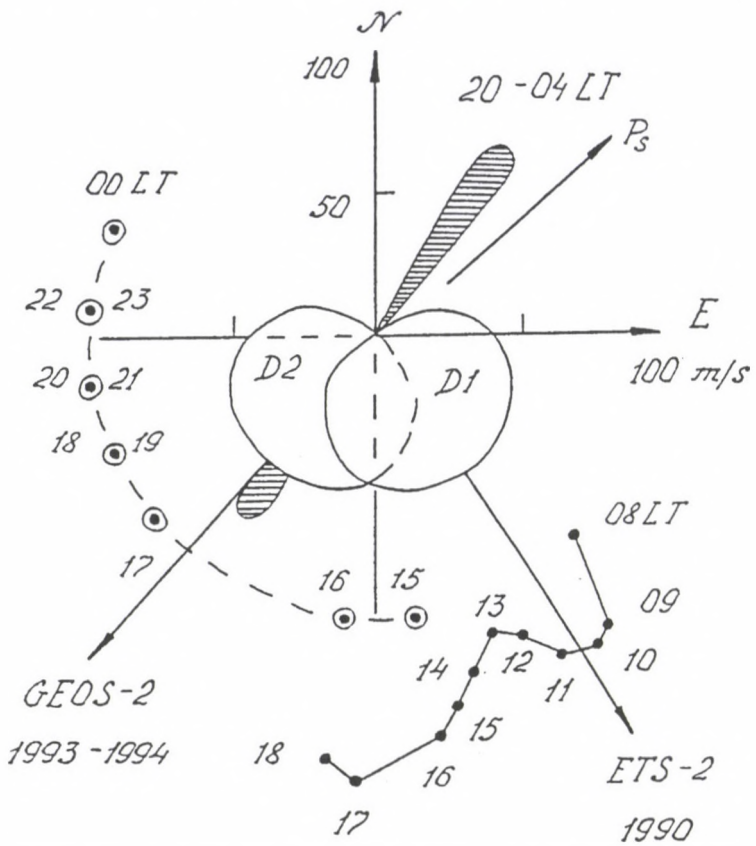


Fig. 4. Polar diagram, illustrating time dependences of mean values of the MS TIDs velocity vector for the daytime (a sequence of heavy dots for the interval 08–18 LT) and for the night-time (hatched regions for the interval 20–04 LT) for 43 days of the autumn of 1990. The direction of the normal $\vec{P}$ to the "radio shadow" of the magnetic field vector for height of disposition of SSIs 300 km in our case averages 50° and exceeds a mean value of ψ for the night-time by 15°–20°. Circles schematically plot time dependences of mean values of the MS TIDs propagation direction ψ from Jacobson *et al.*, (1995). The heavy lines plots the calculated azimuthal diagrams D1 and D2 of the ETS-2 and GOES-2 radio path response to the passage of AGW with the period 30 min

time dependences of our measured mean values of the MS TIDs velocity vector for the daytime of 43 days of the autumn of 1990 (a sequence of heavy dots for the interval of 08–18 LT). The azimuth of the direction toward satellite ETS-2 is $150° + 5°$. The heavy curve shows the calculated azimuthal diagram D1 of the ETS-2 radio path response to the passage of AGW with the period of 30 min, without taking into account the neutral wind influence, from a paper of Bertel et al. (1976).

Circles on this same diagram mark schematically and approximately time dependences of mean values of the MS TIDs propagation directions from data of 500-day measurements in 1993–1994 made by Jacobson et al. (1995). The range of velocities is 80–100 m/s; the azimuth of the direction toward GOES-2 is $223° \pm 9°$.

The heavy curve shows a (similar to D1) calculated azimuthal diagram D2 of the GOES-2 radio path response to the passage of AGW with the period of 30 min. A substantial difference of azimuths toward ETS-2 (up to 70°) enabled us to fill in on the "joint" polar diagram a wide azimuthal sector (from 90° to 300°) and mutually complement the measurements of the diurnal dependence of MS TIDs propagation directions.

4. Conclusions

It was established that during the daytime phase disturbances have the form of a spectrum of travelling waves with closely spaced values of propagation directions ψ (with the spread of 30°–60°) and the phase velocity of 130±50 m/s. The mean value of ψ increases in a clockwise direction in the range from 90° to 230° and opposite to the calculated direction of the neutral wind with the spread not exceeding 10°–30°. This result is in harmony with the hypothesis of neutral wind-induced filtering of acoustic-gravity waves (AGW) and MS TIDs caused by them.

Nighttime observational data reveal two groups of disturbances with oppositely directed phase velocities. They are consistently smaller by 15°–20° than the propagation directions of SSIs obtained through simultaneous measurements of ETS-2 radio signal scintillations and coinciding with a normal to the "shadow" of the magnetic field vector in the antenna system plane.

The effects intimated above are most clearly pronounced in winter and in autumn and are almost independent of the level of magnetic activity (up to $K_p = 5$).

References

Afraimovich E L 1995: Dynamics and anisotropy of travelling ionospheric disturbances as deduced from transionospheric sounding data. I — Statistical angle-of-arrival and Doppler method (SADM), Preprint No. 5–95, Institute of Solar-Terrestrial Physics RAS, Irkutsk

Afraimovich E L, Zvezdin V N, Minko N P, Shapovalov A N 1991: *Radio Sci.*, 26, 1177–1198.

Bertel L, Bertin F, Testud J 1976: *J. Atmos. Terr. Phys.*, 38, 261–270.

Cho H R, Yeh K C 1970: *Radio Sci.*, 5, 881–894.

Crowley G, Jones T B, Dudeney J R 1987: *J. Atmos. Terr. Phys.*, 49, 1155–1162.

Georges T M 1968: *J. Atmos. Terr. Phys.*, 30, 735–746.

Jacobson A R, Carlos R C, Massey R S, Wu G 1995: *J. Geophys. Res.*, 100, 1653–1665.

Kalikhman A D 1980: *J. Atmos. Terr. Phys.*, 42, 697–703.

Mercier C 1986: *J. Atmos. Terr. Phys.*, 48, 605–624.

Mercier C 1996: *Ann. Geophys.*, 14, 42–58.

Spoelstra T A Th 1992: *J. Atmos. Terr. Phys.*, 54, 1185–1195.

Waldock J A, Jones T B 1986: *J. Atmos. Terr. Phys.*, 48, 245–260.

Zhovty E I, Chernigovskaya M A 1988: In: Researching of the dynamics processes in the upper atmosphere, Moscow, Gidrometeoizdat Publ. House, 267–271.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 309–319 (1997)

IONOSPHERIC EFFECTS OF THE SOLAR ECLIPSE OF MARCH 9, 1997, AS DEDUCED FROM DATA FROM THE GPS-RADIO INTERFEROMETER AT IRKUTSK

E L Afraimovich[1], K S Palamartchouk[1], N P Perevalova[1], V V Chernukhov[1], A V Lukhnev[2], V T Zalutsky[3]

This paper presents data from first measurements of total electron content (TEC) and its gradients during the solar eclipse of March 9, 1997, obtained with the GPSINT GPS-radio interferometer at Irkutsk. The interferometer consists of three receivers (one TurboRogue SNR-8000, and two Ashtech Z-12) located at the vertices of a triangle and spaced by about 3–5 km. The measured TEC variations are indicative of profound changes in the ion production process in the ionosphere attendant on the solar eclipse, simultaneously in a large volume of space with a radius of 300 km at 300 km altitude. The delay of a minimum value of TEC with respect to the maximum phase of eclipse at 300 km altitude ranges from 10 to 34 min, and the depression depth of TEC growth varies from 10 to 50%. By analysing the data on TEC gradient variations, one is led to conclude that the depression of TEC growth during the eclipse is essentially independent of the latitude (within the observation range $52 \pm 4°$ N), and longitude ($104 \pm 6°$E).

Keywords: acoustic gravity waves; GPS; ionosphere; SADM; solar eclipse; transionospheric sounding; travelling ionospheric disturbances

1. Introduction

Radio probing of the ionosphere during solar eclipses provides important information about atmospheric processes in a wide height range. The literature on the subject is quite extensive (see, e.g., a thorough review by Cohen 1984).

Regular ionospheric effects of solar eclipses imply a decrease in TEC down to 25%, an increase in F layer minimum height and effective reflection heights, and a decrease in density in the F layer maximum, which is characteristic for the nightside ionosphere (Cohen 1984). The behaviour of the above parameters can be simulated using corresponding models of the ionosphere (see, e.g., Stubbe 1970).

Dynamic processes during the eclipse associated with the appearance of wave motions in the region of local cooling or heating of the atmosphere at the passage of the moon's shadow are more complicated for an investigation. It was shown theoretically (Chimonas and Hines 1970) that such conditions are favourable for generation of atmospheric gravity waves (AGW), whose intensity permits their detection even at some distance from the band of totality. Geophysical conditions

[1]Institute of Solar-Terrestrial Physics RAS SD, POB 4026, 664033 Irkutsk-33, Russia, e-mail: afra@iszf.irk.ru

[2]Institute of the Earth's Crust RAS SD, 664033 Irkutsk-33, 128 Lermontov Str.; e-mail: san@cora.irkutsk.su

[3]East-Siberian Research Institute for Physicotechnical and Radioengineering Measurements; 664056 Irkutsk-56, 57 Borodin Str.; e-mail: eopsr@niiftri.irkutsk.su

1217-8977/97/$ 5.00

during the eclipse have a dramatic effect on observed wave motions; therefore, it is not always possible to attribute observed periodic variations of ionospheric parameters to eclipse effects. Reliable estimates of AGW characteristics under these conditions have not yet been obtained (Cohen 1984).

Earlier investigations into ionospheric effects of eclipses were made using predominantly ionosondes. Of interest are also some examples of phase and Doppler measurements in the HF range (Ishinose and Ogawa 1976). However, results reported in the last reference, along with data from ionosonde measurements, are not always comparable and are difficult to interpret.

Some experimental data on the influence of solar eclipses upon terrestrial ionospheric characteristics using methods of sounding by VHF signals from geostationary satellites were obtained by measuring the Faraday rotation. Those measurements revealed the eclipse-induced effect of deep (as deep as 20–30%) TEC depression in the ionosphere, with a typical time of TEC decrease and recovery of about one hour and with the delay with respect to the maximum phase of occultation ranging from a few to tens of minutes. It is impossible to give here a complete account of these results for reasons of space, and the reader may found it in the cited review by Cohen (1984).

A limitation of methods based on analysing signals from geostationary satellites is that the number of satellites is small, and they are nonuniformly distributed in longitude. For that reason, it was impossible to make measurements in a large number of geophysical interesting regions on the globe, and also during the eclipse under consideration.

The new potentialities of remote diagnostics of the ionosphere are based on using standard measurements of transionospheric radio signal characteristics and coordinate-time measurements through the use of two-frequency multichannel receivers of the global navigation GPS system. These receivers are used to make (almost at any point of the globe and at any time simultaneously at two coherently-coupled frequencies $f_1 = 1575.42$ MHz and $f_2 = 1227.6$ MHz) highly accurate measurements of the group and phase delays between the receiver on the ground and the transmitters on the GPS system satellites in the reception zone (Hofmann-Wellenhof et al. 1992, Melbourne et al. 1994).

In a review by Cohen (1984), mention is made of a private communication of Klobuchar who gives an example of TEC measurements using earlier satellites of the NAVSTAR (GPS) system during the eclipse of February 26, 1979, in California. Unfortunately, publications with a detailed account and analysis of those measurements are unavailable to us.

Until recently, a major effort of researchers went into the study of capabilities and into the development of techniques for measuring and modelling the space-time distribution of TEC on the basis of GPS-measurements. The literature on the use of data from the GPS system to investigate ion density irregularities in the ionosphere is considerably less extensive. At the same time the sensitivity of phase measurements in the GPS system permit the detection of irregularities with an amplitude of up to $10^{-4} - 10^{-5}$ from a diurnal TEC variation.

The above-mentioned merits of using GPS receivers, combined with a high level

of technology and computerisation of the system, awakened ever increasing interest of researchers when studying ionospheric effects of disturbances of a natural or artificial origin, such as earthquakes (Calais and Minster 1995); spacecraft launches (Calais and Minster 1996); and industrial explosions (Fitzgerald 1997).

This paper presents preliminary results derived by investigating ionospheric effects of the March 9, 1997 solar eclipse using data from the "GPSINT" interferometer that was developed and constructed in Irkutsk. Continuous measurements with this new instrument got underway on March 5, 1997. The GPSINT instrument provides information not only on slow regular variations in TEC and its spatial gradients but also on structural and dynamic features of ionospheric irregularities over a wide range of scales (from a few to hundreds of kilometres).

Issues relating to a detailed physical interpretation of the data obtained are deliberately not taken up in this paper. They will be the subject of further studies where we intend to construct a self-consistent model for the dynamics of aeronomic processes in the ionosphere by calculating solar radiation flux variations and appropriate ionospheric models. This applies also to the analysis of ionospheric irregularity characteristics.

2. General characteristics of the March 9, 1997 eclipse, and the geometry of experiment

A solar eclipse was observable on March 9, 1997 in the northern part of the eastern hemisphere (Espenak and Anderson 1995). The band of totality on the terrestrial surface started at the location 49N, 87E, on the Russian Chinese frontier and continued on the territories of Mongolia and Russia. Major urban centres within the path of totality included Chita. The cities of Irkutsk, Ulan-Ude and Ulan-Bator were in the immediate neighbourhood of the path. The largest width of the path of totality was about 370 km. At ionospheric heights the path of totality was much farther to the south.

Figure 1 shows a schematic map of the southern part of East Siberia where our observations were made. Dashed lines represent the equal maximum phase of eclipse $P_{\max}$ at $H = 300$ km corresponding to the phases: 0.14, 0.10, and 0.06; $P_{\max}$ is equal to the ratio of the visible area of the solar disk to its total area. The height of 300 km was taken with some degree of convention in order to be tentatively referenced to the ionospheric F2 region electron density maximum.

The distinctive feature of this eclipse was the supersonic speed in excess of 0.8 km/s, with which the moon's shadow moved on the Earth's surface.

The phase of eclipse at the ionospheric height can differ by more than an order of magnitude compared with the value of $P_{\max}$ on the terrestrial surface. The time $T_{\max}$ for $H = 300$ km over Irkutsk is 1.3 min ahead of that on the ground. The difference in the values of $T_{\max}$ and their onset times is due to the relatively low altitude of the Sun. At the time of the largest phase in Irkutsk (0054 UT) it was as low as 11.5 degrees.

Following are results derived by processing the data of transionospheric sounding by signals from four satellites (PRN01, PRN14, PRN25, and PRN29) observed simultaneously at the three locations of the GPS-interferometer without technical

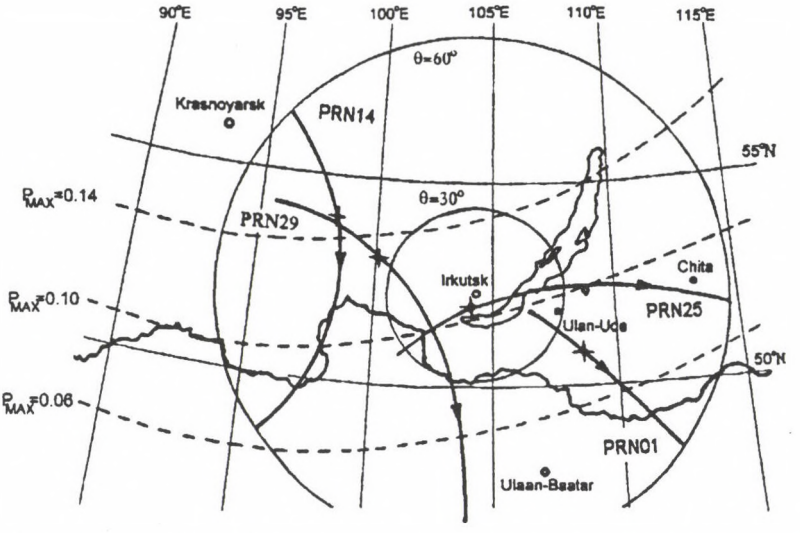


Fig. 1. Schematic map of the southern part of East Siberia where our observations were carried out. Dashed lines are the lines of equal maximum phase P_{max} of eclipse at $H = 300$ km, corresponding to the phases: 0.14, 0.10, and 0.06. Circles enclose areas of subionospheric points at $H = 300$ km for selected satellites visible from Irkutsk in the range of zenith angles from 0 to 30 and 60 degrees. Trajectories of subionopsheric points for 300 km altitudes are shown for each satellite with arrows that show their directions, and asterisks label the location of subionospheric points at the time of the maximum phase of eclipse

failures from 0000 and 0200 UT and visibled at zenith angles Θ less than 60 degrees. Circles in Fig. 1 enclose areas with subionospheric points at $H = 300$ km for selected satellites visible from Irkutsk in the range Θ from 0 to 30 and 60 degrees, respectively. The figure also shows trajectories of subionospheric points for 300 km altitude, corresponding to directions to the GPS satellites used in the analysis; asterisks show the location of subionospheric points at the time of the maximum phase of eclipse at $H = 300$ km. The above-mentioned requirements are not met by measurements from other satellites which were observed either totally or partially at the period of time mentioned above (PRN05, PRN06, PRN15, PRN16, PRN21, PRN22, and PRN30). However, these data will be used subsequently in a more detailed analysis.

Values of T_{max} and P_{max} used in the analysis are presented in Table I for each satellite. The table also lists locations of subionospheric points (X_{max}, km) and (Y_{max}, km) in the Cartesian coordinate system where the values of x and y of the GPS-interferometer are roughly taken to be 0, and the x- and y-axes are directed to the east E and to the north N, respectively.

According to observations at the magnetic observatory Irkutsk, the day of the March 9, 1997 eclipse was quiet, with the diurnal sum of K-indices amounting to 10. March 8 and 10 were also relatively quiet days, with the diurnal sums 16 and 11, respectively. This enabled us to circumvent difficulties arising when analysing ionospheric effects of a solar eclipse under conditions of a disturbed ionosphere.

Table I.

	PRN01	PRN14	PRN25	PRN29
T_{max}, UT	00:50:00	00:51:45	00:50:45	00:51:15
P_{max}	0.079	0.14	0.113	0.133
X_{max}, km	296	−327	−16	−286
Y_{max}, km	−170	192	−10	131
T_{min}, UT	01:24:00	01:01:30	01:00:30	01:00:30
dT, min	34	10	10	9
M, %	10	50	50	50

3. Method to determine TEC variations and gradients

A method of reconstructing TEC variations from measurements of the ionosphere-induced additional increment of the group and phase delay of the satellite radio signal was detailed and validated in a series of publications (Hofmann-Wellenhof et al. 1992, Melbourne et al. 1994; Calais and Minster 1995, 1996, Fitzgerald 1997). We reproduce here only the final formulas

$$\mathrm{TEC} = -\frac{1}{40.308}\frac{f_1^2 f_2^2}{(f_1^2 - f_2^2)}\left[(P_1 - P_2) + nP\right] \tag{1}$$

$$\mathrm{TEC} = -\frac{1}{40.308}\frac{f_1^2 f_2^2}{(f_1^2 - f_2^2)}\left[(L_1\lambda_1 - L_2\lambda_2) + \mathrm{const} nL+\right] \tag{2}$$

where P_1 and P_2 are additional paths of the radio signal caused by the group delay in the ionosphere, m; $L_1\lambda_1$ and $L_2\lambda_2$ are additional paths of the radio signal caused by the phase delay; L_1 and L_2 represent the number of phase rotations at the frequencies f_1 and f_2; λ_1 and λ_2 stand for wavelengths, m; const is the unknown initial value; and nP and nL are errors of determining the group and phase path.

For the sake of convenience when representing TEC data, in current publications a convention uses the unit TECU = 10^{16} m^{-2}. Typically, the range of TEC variation during 24-hour period is 30–60 TECU depending on geophysical conditions.

Phase measurements in the GPS system can be made with a high degree of accuracy corresponding to the error of TEC determination of at least 10^{-3} TECU when averaged on a 30-second time interval, with some uncertainty of the initial value of TEC, however. Group delay measurements furnish an opportunity to determine the absolute value of TEC, but with an error not better than 10^{-1} TECU when averaged for 30 s.

In this paper we therefore limit ourselves only to TEC variations obtained from phase delay measurements by Eq. (2). For definiteness sake, we bring the TEC variations into the region of positive values with the minimum value equal to 0. Subsequently, when solving the above-mentioned aeronomic problem of solar eclipse effect, we take advantage of all tools for reconstructing the absolute value of TEC, including additional measurements of the F2 layer critical frequency with the ionosonde.

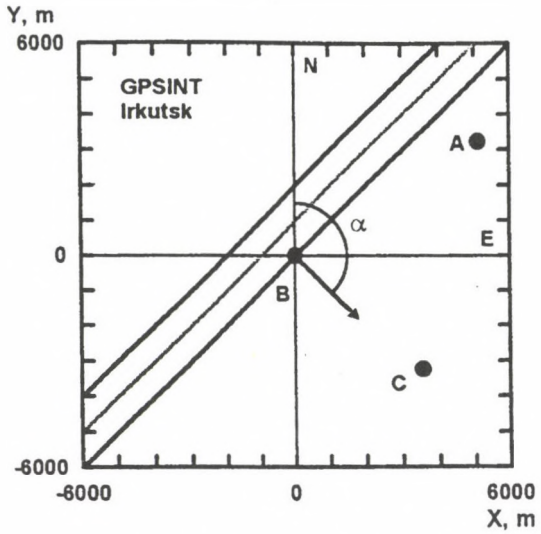


Fig. 2. Fig. 2. Geometry of the GPS-interferometer GPSINT. A,B,C – reception points spaced by about 3–5 km where two-frequency multichannel GPS receivers are installed. Parallel lines are the lines of equal TEC; the arrow shows the direction of a normal α to these lines; and N and E are the directions to the north (axis) and to the east (axis)

The capabilities of reconstructing the space-time distribution of TEC within a wide cone of angles over the site of observation can be considerably extended through the use of the GPS radio interferometer to make additional measurements of TEC gradients for each GPS satellite observed at a given instant of time. This problem can be resolved by using an electron density distribution model that is updated by measurements of TEC and its gradients in many known directions at once.

The geometry of the GPS-interferometer GPSINT at Irkutsk is presented in Fig. 2. Here A, B, and C are the reception sites spaced by about 3–5 km where two-frequency multichannel receivers are installed (one TurboRogue SNR-8000, and two Asgtech Z-12). Parallel lines show arbitrarily the lines of equal TEC; the arrow shows the direction of a normal (to these lines; and N and E are the directions to the north (axis) and to the east (axis).

The longitudinal and latitudinal TEC gradients $G_E = \frac{d(\mathrm{TEC})}{dx}$ and $G_N = \frac{d(\mathrm{TEC})}{dy}$ were determined through linear transformations of TEC differences for the pairs of reception sites (TEC_A – TEC_B) and (TEC_C – TEC_B). For the purposes of this paper, we confine our consideration to gradient variations only. For the sake of definiteness, we bring the variations of TEC gradients into the region of positive values, with the minimum value equal to 0.

All data discussed below were obtained by smoothing out the TEC variations with a 15-min time window, which permitted filtering of rather shallow fast variations caused by the measurement noise and small-scale irregularities. Thus, errors

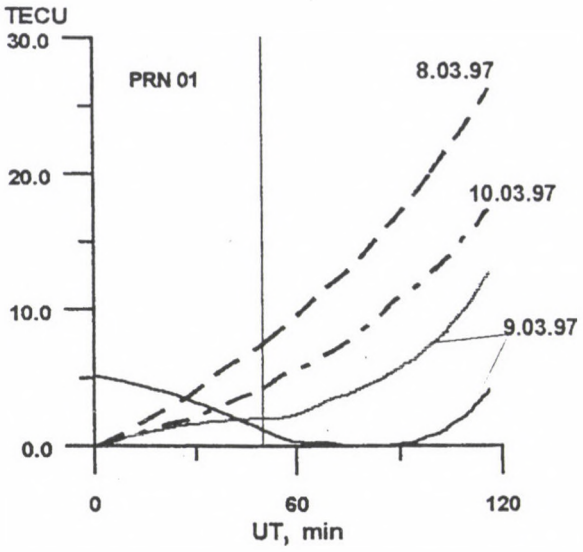


Fig. 3. TEC variations obtained from PRN01 data for March 9, 1997, as well as for the reference days of March 8 and 10, 1997 (dashed curves). The reference days are characterized by a monotonic increase in TEC, whose behaviour is perturbed on March 9. By removing the linear trend, it is possible to identify slow TEC variations caused by the solar eclipse (thick curve). The vertical bar marks the time of the maximum phase T_{max} at 300 km altitude for this satellite

of measurement of TEC gradients with the "GPSINT" interferometer did not exceed 10^{10} m^{-3}. Because eclipse-induced variations of these parameters were 3-4 orders of magnitude as large, they are not shown on all the plots below. All data are plotted for the time interval 0000–0200 UT (in minutes with respect to 0000 UT) where T_{max} corresponds approximately to its middle, and its ends correspond to the time of first and last contact of the solar disk in the optical range at the ground level in Irkutsk. The vertical bar on the plots for each satellite marks the time of maximum phase T_{max} at 300 km altitude.

We also rejected the usually used conversion of measured variations of "oblique" TEC to the value of "vertical" TEC (Calais and Minster 1995, 1996) because such an approach is legitimate only for a horizontally homogeneous ionosphere, but in our situation this is by no means always the case.

4. Variations of TEC and its gradients during the eclipse

Figure 3 presents TEC variations obtained from PRN01 data for March 9, 1997, as well as for the reference days of March 8 and 10, 1997 (dashed curves). The reference days are characterised by a gradual increase in TEC, whose behaviour is perturbed on March 9. By removing the linear trend, it is possible to identify slow TEC variations caused by the solar eclipse (thick curve); a comparative analysis of the effect is made below.

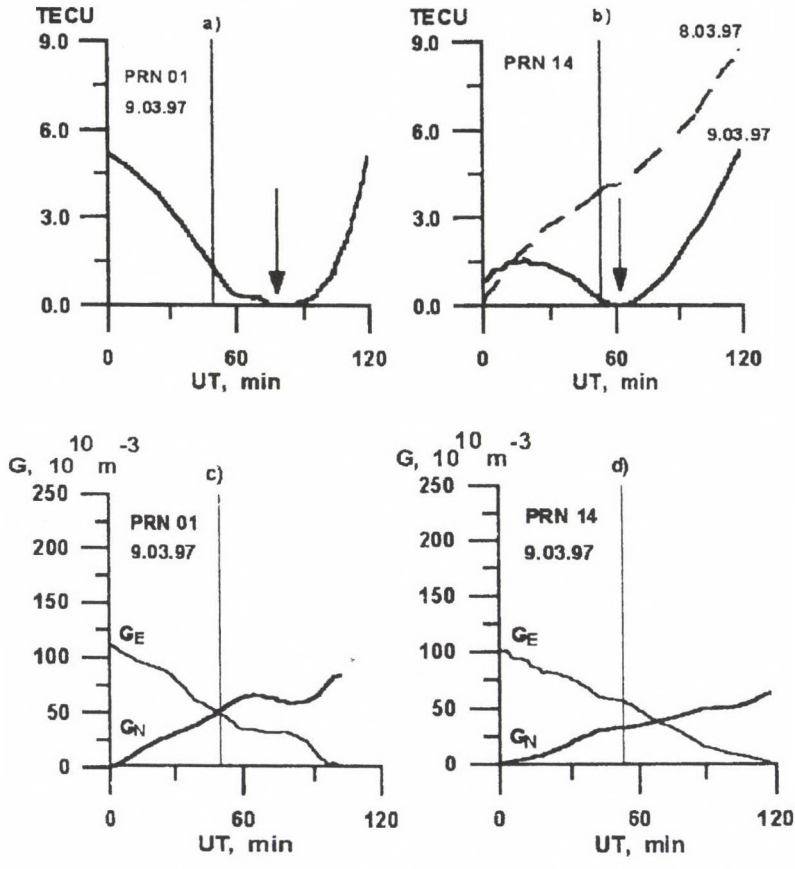


Fig. 4. Variations of TEC (a,b) and gradients G_E and G_N (c,d) as obtained for PRN01 and PRN14 for March 8 and 9, 1997. Vertical bars mark the time of the maximum phase T_{max} at 300 km altitude for these satellites, and the arrows mark the time of the minimum value of TEC

TEC variations obtained for all satellites on March 8 and 9, 1997, are presented in Figs. 4 and 5 (PRN01 – 4a; PRN14 – 4b; PRN25 – 5a; PRN29 – 5b). It is easy to recognise dramatic differences in TEC variations for March 8 and 9, similar to the above-mentioned one for PRN01. All satellites are characterised by almost the same form of monotonic increase in TEC on the reference day of March 8 and by the eclipse-induced depression of TEC growth. During the eclipse the value of oblique TEC starts to decrease gradually immediately after the first contact and reaches a minimum TEC at instants of T_{min} shown by the arrows in Fig. 4 and 5. This is followed by a gradual increase and recovery of TEC.

Main parameters of the TEC(t)-curve that characterise the eclipse effect (T_{min}, the response delay $dT = T_{min} - T_{max}$, and the approximate depth of depression M, as a percentage of background conditions) are presented Table I. Because only TEC variations were used in our analysis, the value of M was estimated in a

very crude way by comparing with averaged values of TEC using data reported by Davies (1980).

A comparison between T_{max} at 300 km and T_{min} of the largest TEC response showed that almost for all satellites, with the beams toward which were directed close to the zenith (PRN25) or north-westward (PRN14 and PRN29), the delay DT and the depression depth M were found to approach 10 min and 20 – 50%, respectively.

Unlike these satellites, the delay dT for PRN01 exceeded 34 min, and the depression depth M was considerably smaller (10%). This difference can be accounted for by the fact that the beam to PRN01 was directed south-eastward, in the region of an established TEC similar to TEC for midday.

Variations of gradients G_E and G_N obtained for all satellites on March 8 and 9, 1997, are presented in Figs 4 and 5 (PRN01 – 4c; PRN14 – 4d; PRN25 – 5c; PRN29 – 5d). Unfortunately, for technical reasons, it was possible to measure gradients on March 8 only after 0042 UT and only for PRN25 and PRN29.

Time-dependencies of TEC gradients for both satellites on the reference day of March 8, are similar in the form and range of variation. The longitudinal gradient G_E decreases monotonically with the time, which corresponds to a decrease in the TEC growth rate in this direction as the midday values of TEC are approached. By contrast, however, the latitudinal gradient G_N also increases monotonically with the time. A comparison of the variations of TEC gradients for March 8 and 9 showed that the form of variation of the gradients G_E and G_N of all satellites is similar to the background form. At the same time, even for these satellites one can identify minor fluctuations of gradients occurring in synchronism with TEC variations, which is more pronounced in case of PRN01.

5. Conclusion

We have presented the first measurements of total electron content (TEC) and its gradients during the solar eclipse on March 9, 1997, made by the GPS-interferometer GPSINT in Irkutsk.

The evidence obtained for TEC variations point to profound alterations in the ion production process in the ionosphere during the solar eclipse simultaneously in a large volume of space with a radius of at least 300 km at 300 km altitude. The delay of the minimum value of TEC with respect to the maximum phase of eclipse ranges from 10 to 34 min, and the depression depth varies from 10 to 50%. Our data are in reasonably good agreement with published evidence (see a review by Cohen 1984). The new element is that such TEC measurements were made simultaneously in four and more directions.

As far as gradients are concerned, such measurements seem to have been made for the first time. The analysis of these data suggests the conclusion that the depression of TEC growth during the eclipse is almost independent of the latitude (within the observed range $52 \pm 4°$N), and longitude ($104 \pm 6°$E). Only an approximate qualitative picture can be gained from the data obtained. We intend to obtain more comprehensive information at a later time by reconstructing, from data on

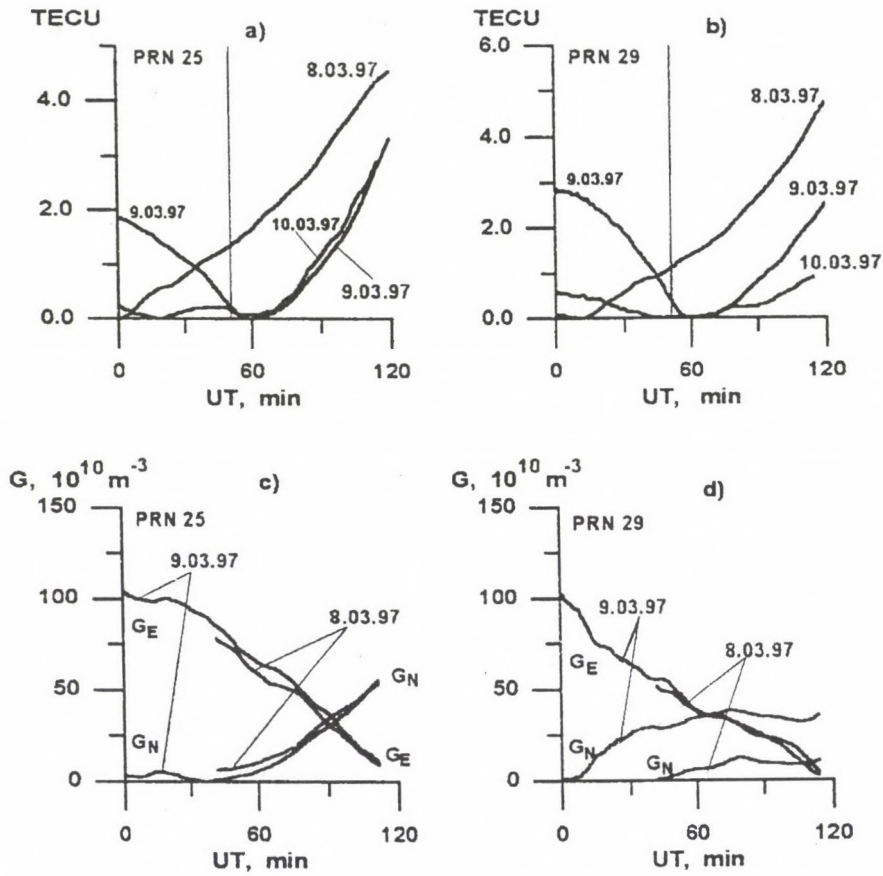


Fig. 5. Variations of TEC (a,b) and gradients G_E and G_N (c,d) as obtained for PRN25 and PRN29 for March 8 and 9, 1997. Vertical bars show the time of the maximum phase T_{max} at 300 km altitude for these satellites

TEC and its gradients, the space-time distribution of TEC in the area of passage of the moon's shadow.

Acknowledgements

We are indebted to V V Koshelev, K G Levi and A D Kalikhman for help in organising the experiment and for helpful discussion of results. Thanks are also due to V G Mikhalkovsky for his assistance in preparing the English version of the manuscript. This work was done with support from the Russian Foundation for Fundamental Research, grants 96-05-64162 and 97-02-96060.

References

Calais E, Minster J B 1995: *Geophys. Res. Lett.*, 22, 1045–1048.
Calais E, Minster J B 1996: *Geophys. Res. Lett.*, 23, 1897–1900.
Chimonas G, Hines C O 1970: *J. Geoph. Res.*, 75, 875.
Cohen E A 1984: *Radio Sci.*, 19, 769–777.
Davies K 1980: *Space Sci. Rev.*, 25, 357–430.
Espenak F, Anderson J 1995: Total solar eclipse of 1997 March 9, NASA Reference Publication 1369, 64 p.
Fitzgerald T J 1997: *J. Atmos. Terr. Phys.*, 59, 829–834.
Hofmann-Wellenhof B, Lichtenegger H, Collins J 1992: Global Positioning System: Theory and Practice. Springer-Verlag, Wien, New York
Ishinose T, Ogawa T 1976: *J. Geoph. Res.*, 81, 2401–2404.
Melbourne W G, Davis E S, Duncan C B et al. 1994: *Jet Propulsion Laboratory Publication*, 94-18.
Stubbe P 1970: *J. Atmos. Terr. Phys.*, 32, 1109–1116.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 321–327 (1997)

TID OBSERVATION USING A SHORT BASELINE NETWORK OF GPS RECEIVERS*

WEIXING WAN[1], BAIQI NING[1], HONG YUAN[1], JINNIAN LI[1], LIBIN LI[1], JUN LIANG[1]

The total electron content (TEC) observed by GPS method is seldom used in the investigation of travelling ionospheric disturbances (TIDs). The main reason is that the motion of the observation point leads to the mixture of both spatial and temporal changes of TEC in the GPS observation. The present work is devoted to demonstrate that such influence can be modified with the observation by a short baseline network of GPS receivers. The experiment results show that the suggested method is useful for estimating wave parameters of TIDs and investigating their evolution with both space and time. The prime advantage of the new experiment method is that the short baseline GPS network may be used to study the ionospheric disturbances over a very large region.

Keywords: GPS network; ionospheric disturbances

1. Introduction

By means of receiving signals of a beacon satellite and measuring their differential Doppler frequency shifts or Faraday rotation angles, the total electron content (TEC) of the ionosphere can be observed and further used to study travelling ionospheric disturbances (TIDs). Traditionally, two kinds of satellites, known as the low orbiting satellites and the geostationary satellites, have been widely used in the investigation of TIDs. The low orbiting satellites such as NNSS (Leitinger et al. 1975, Ogawa et al. 1987) travel very fast with the speed (about several kilometres per second) of observation point at the sub-ionospheric height much higher than the typical horizontal phase velocity of TIDs (about several hundred meters per second). In this case the observed data may represent the spatial changes of TEC and can be used to study the spatial variation of TIDs (Pryse et al. 1995). On the other hand, the geostationary satellites are used to study the temporal variation of TIDs (Davis and da Rosa 1970). It is important to point out that close spaced arrays (short baseline networks) of beacon receivers of geostationary satellites are often used to study the horizontal propagation of TIDs (Jacobson et al. 1995).

Recently, GPS satellites have been widely used in the investigation of ionospheric TEC, and sometimes even in the study of ionospheric disturbances such as TIDs (Calais and Minster 1996, Ho et al. 1996). The GPS satellites belong to the high orbiting satellites. When this kind of satellite is used in the observation of TIDs, the speed of the observation point may reach to 200 m/s, which is close to or sometimes exceed the phase velocity of the observed TIDs. This leads to the

*This research is supported by the National Nature Science Foundation No. 49525407

[1]Wuhan Ionospheric Observatory, Wuhan Institute of Physics and Mathematics, The Chinese Academy of Sciences, Wuhan 430071, China

1217-8977/97/$ 5.00

difficulties in the explanation of the observed results because both their spatial and temporal changes are mixed with each other. Even though, it is desirable for us to tap the potentialities of the GPS observation for the advantage that it may monitor the ionospheric disturbances over a large region. In the present work, we propose a technique to eliminate the influence of the motion of the observation point and estimate the horizontal propagation parameters of the TIDs from the observation of a short baseline network of GPS receivers. In the following we will first describe briefly the suggested method, then give some experimental examples to show the usefulness of the suggested technique, and finally is a brief summary about the advantages of the short baseline GPS network in the observation of TIDs.

2. Method

For beacon signals of a certain GPS satellite received by the j'th receiver of the short baseline network, the observed TEC may be expressed as,

$$TEC_j(t) = TEC\,(x_j(t), y_j(t), t)\ , \tag{1}$$

where TEC is a function of horizontal space $x - y$ and time t; x_j and y_j indicate the horizontal co-ordinates of the sub-ionospheric points in the ray path between the GPS satellite and the j'th receiver.

When the baselines of the network are short enough, i.e., about 10 km as used in the present work, it is a good approximation that the speeds, V_x and V_y , of sub-ionospheric points corresponding to all the receivers are the same. Furthermore, we assume that V_x and V_y are constants during a short period (a short time window) of the observation, i.e.,

$$V_x = \frac{dx_j(t)}{dt}\,, \qquad V_y = \frac{dy_j(t)}{dt}\,. \tag{2}$$

Eq. (1) is then expressed as

$$TEC_j(t) = TEC(x_{0j} + V_x t, y_{0j} + V_y t, t)\,, \tag{3}$$

where x_{0j} and y_{0j} are the horizontal co-ordinates $x_j(t)$ and $y_j(t)$ at $t = 0$.

In our analysis we use the following model of TEC spectrum (Wan et al. 1995, 1996)

$$TEC(k_x, k_y, \omega) = TEC(\omega)\sigma[k_x - k_x(\omega)]\sigma[k_y - k_y(\omega)]\,, \tag{4}$$

where ω and k_y are respectively the frequency, the two orthogonal horizontal components of the wave number vector of TIDs. Equivalently, Eq. (4) may be expressed in space-time domain as,

$$TEC(x, y, t) = \int TEC(\omega)\exp\{i[\omega - k_x(\omega)x - k_y(\omega)y]\}d\omega\,. \tag{5}$$

Substitute Eq. (5) into Eq. (3) we obtain,

$$TEC_j(t) = \int TEC(\omega)\exp\{i[(\omega - k_x(\omega)V_x - k_y(\omega)V_y)t - k_x(\omega)x_{0j} - k_y(\omega)y_{0j}]\}d\omega\,. \tag{6}$$

Take the Fourier transformation of the above equation we get,

$$TEC_j(\omega') = TEC(\omega) \exp\{-i[k_x(\omega)x_{0j} + k_y(\omega)y_{0j}]\}\,, \tag{7}$$

where ω' is the apparent frequency observed in the co-ordinate system moving with the observation point,

$$\omega' = \omega - k_x(\omega)V_x - k_y(\omega)V_y\,. \tag{8}$$

Based on Eq. (7) the parameters ω', k_x and k_y are respectively estimated from auto and cross spectra of $TEC_j(t)$ observed from the different receivers of the short baseline network. Finally, the real frequency and phase velocity is estimated as

$$\omega = \omega' + \Delta\omega, \qquad \Delta\omega = k_x V_x + k_y V_y \tag{9}$$

$$V_{ph} = V'_{ph} + \Delta V_{ph}, \qquad V'_{ph} = \frac{\omega'}{k}, \qquad \Delta V_{ph} = \frac{k_x V_x + k_y V_y}{k}\,. \tag{10}$$

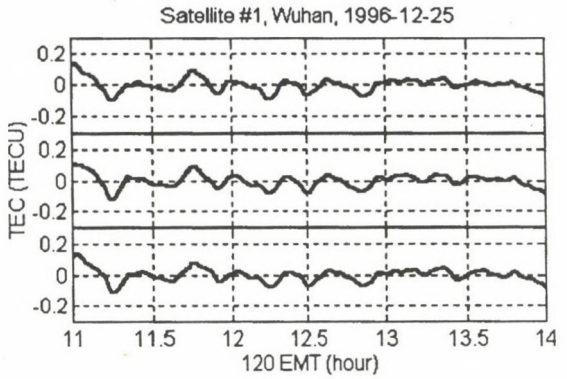
It is clear that the frequency adjustment $\Delta\omega$ is just like the "Doppler shift" caused by the motion of the observation point, and the phase velocity adjustment ΔV_{ph} equals to the projection of the speed of observation point on the direction of the wave propagation. Additionally, the propagation azimuth σ is not affected by the motion of the observation point, i.e.,

$$\sigma = \tan^{-1}\frac{k_x}{k_y}\,. \tag{11}$$

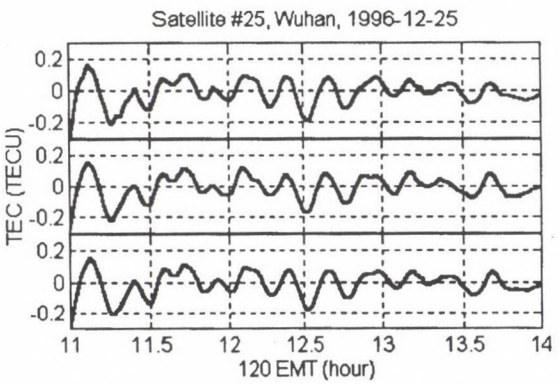
3. Examples and discussion

To demonstrate the usefulness of the above method, three GPS receivers have been running at Wuhan as a short baseline (about 10 km) network. The TEC used in the present work is measured from the observed carry phase delay of dual-frequency (L1 and L2) GPS signals, so that the measured TEC oscillation is precise enough to investigate the ionospheric disturbance. As examples, we consider the observation during the noon time (1100–1400 LT) on September 25, 1996. In this 3-hour period only 3 satellites (Nos 1, 25 and 30) may be "seen" all of the time. To identify the TID oscillation from the observation, the observed TEC data are firstly filtered out with a high pass digital filter so that the constant bias of the phase measurement is eliminated and the large scale variation (such as daily variation) of TEC is restrained. Figure 1 shows the pre-processed TEC curves corresponding to the three satellites and the three receivers.

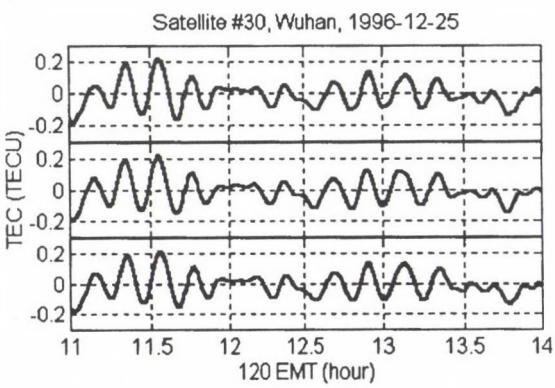
It is obvious from Fig. 1 that there exist wave-like oscillations (TIDs) during the certain observation time. Apparently, the oscillation period is ranged from 10 to 20 minutes. The amplitude of the oscillation is about 0.1 to 0.2 TECU (10^{16} el/m^2). For a certain satellite, the changes of the three TEC curves observed from different receivers are very similar because the baselines between the receivers are much shorter than the typical spatial scale of TIDs. Despite the similarities, some slight time delays among the different TEC curves can still be found if one carefully examines the crests and valleys of the oscillation. The time delays between the TEC



(a)



(b)



(c)

Fig. 1. TEC oscillation observed from a short baseline network of GPS receivers during the period 1100–1400 120EMT. The data are treated by a high pass filter. In each figure, the three curves are corresponding to the three receivers of the network. The three figures indicate the TEC oscillation corresponding to a) satellite No. 1, b) No. 25 and c) No. 30 respectively

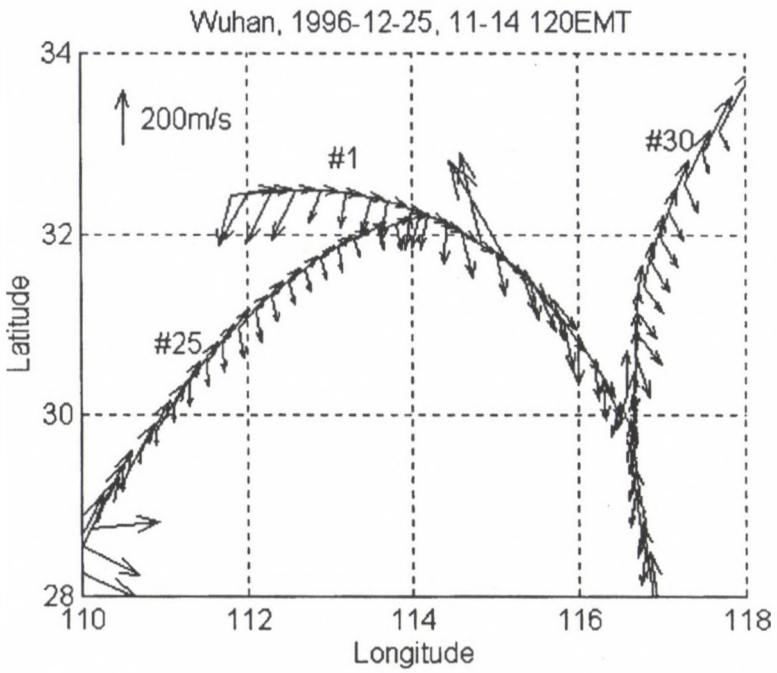
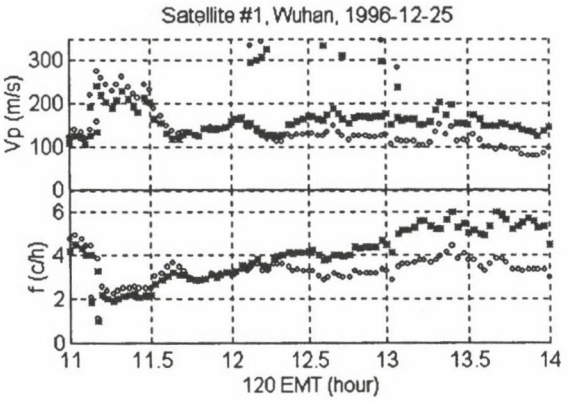


Fig. 2. The quiver plot of TIDs' phase velocity estimated from the TEC data in Fig. 1a-c. For comparison, the velocities of observation point, which are tangent to the orbits, are also plot in the same figure

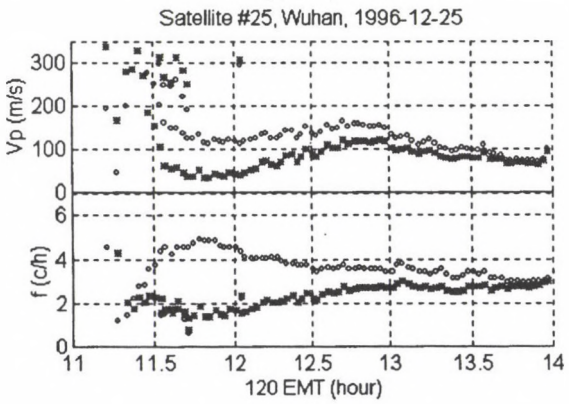
variation observed in different receivers imply that the ionospheric disturbances propagate with certain phase velocities which can be estimated from the time delays themselves.

To estimate the propagation parameters, i.e., the propagation phase velocities V_{ph} and azimuth σ, of TIDs from the observation of the short baseline GPS network, a new analysis technique is used in the present work. This technique is based on the dynamic cross-spectral analysis using the multi-channel maximum entropy method (Shibata 1987, Wan et al. 1995, 1996). But in the present work, we have to use a very narrow time window (1 hour width), which is smaller than those used by Wan et al. (2 hours) and Shibata (3 hours), in order that the speed of the observation point may be assumed invariant within the time window. The procedure estimating the apparent frequency ω' and the wave number vector k_x and k_y by the dynamic cross-spectral analysis will not be presented here (one can refer to Wan et al. 1996), but we point out that the estimated frequency and the phase velocity should be modified with Eqs (9) and (10).

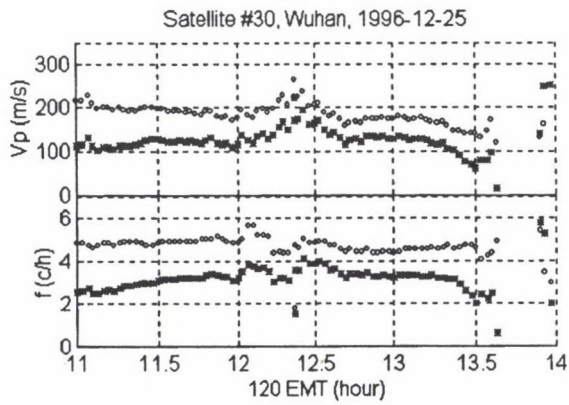
The analysed result, i.e. the vector of the horizontal phase velocity (V_{ph} and σ) of TIDs corresponding to the TEC oscillation in Fig. 1a-c is illustrated as the quiver plot Fig. 2. For comparison, we also plot in the same figure the moving velocities (V_x, V_y) of the observation point, which are of course tangent to the orbits of the



(a)



(b)



(c)

Fig. 3. Comparison of the estimated TIDs' frequency and phase velocity before and after the modification of the motion effect of the observation point. The cycles (o) and the star (*) indicate the parameters before and after the modification respectively

observation points so that one can easily distinguish the two kinds of velocities (V and V_{ph}). As can be seen in this figure, the magnitude of the TID's phase velocities vary from slight less than 100 m/s in the west area to about 150 m/s in the middle and east area, and its propagation direction is approximately from north to south.

As can be seen in Fig. 2 the value of the velocities of the observation point are about the same as those of the TID's phase velocity, which means that the motion of the observation point may seriously affect the estimated propagation parameters, so that the induced error must be modified. To show the requirement of the modification using Eqs (8) and (9), we plot in Fig. 3 both parameters before and after the modification (frequency ω' and ω, phase velocity V'_{ph} and V_{ph}). It is obvious that the modification increases the frequency and phase velocity in Fig. 3a (Satellite No. 1) and decreases in Fig. 3b and Fig. 3c (Satellite No. 25 and No. 30), because the projection of the velocities is along the same direction of TIDs' phase velocities in the former and the opposite direction in the later. The frequency adjustment $\Delta\omega$ and phase velocity adjustment ΔV_{ph} may reach 2 cycle/hour and 100 m/s.

4. Summary

When the TEC measured by GPS method is used in the study of the ionospheric disturbances such as TIDs, the main trouble is that the motion of the observation point makes the mixture of TEC variation both with space and time. The present work provides an experimental technique using a short baseline GPS network to eliminate the effects of the motion of the observation points. Examples show that the present method can be well used to estimate the propagation parameters of TIDs without the influence caused by the motion of the observation points.

It is obvious that the most prime advantage of GPS observation is that it may used to investigate the disturbances of the ionosphere over a large region (about 2000 km). In the observation of TIDs, a short baseline GPS network can be used to not only eliminate influence caused by the motion of the observation points, but also to estimate the wave parameters of the observed TIDs.

References

Calais E, Minster J B 1996: *Geophys. Res. Let.*, 23, 1897.
Davis M J, da Rosa A V 1970: *Nature*, 226, 1123.
Ho C M, Mannucci A J, Lindqwister U J, Pi X, Tsurutani B T 1996: *Geophys. Res. Let.*, 23, 3219.
Jacobson A R, Carlos R C, Massey R S, Wu G 1995: *J. Geophys. Res.*, 100, 1653.
Leitinger R, Schmidt G, Tauriainen A 1975: *J. Geophysics (Zs. Geophysik)*, 41, 201–213.
Ogawa T K, Ifgarashi K, Aikyo K, Maeno H 1987: *J. Geomagn. Geoelectr.*, 39, 709–721.
Pryse S E, Mitchell C N, Heaton J A T, Kersely L 1995: *Ann. Geophys.*, 13, 1311–1319.
Wan W, Yuan H, Ning B, Li J 1995: *Chinese J. Space Sci.*, 15, 301–306.
Wan W, Yuan H, Liang J 1996: *Chinese J. Geophys.*, 39, 17–25.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 329–342 (1997)

EASY TO USE GLOBAL AND REGIONAL IONOSPHERIC MODELS — A REPORT ON APPROACHES USED IN GRAZ

R LEITINGER[1] and G KIRCHENGAST[1]

Out of the necessity to provide quick access to ionospheric model data from various computer programs written in different languages (PASCAL, FORTRAN, IDL, etc.) we developed a "data box" approach in the following way: grid point values are calculated from an existing global or regional model and written into a data file. Good values for grid point distances are 5 degrees in latitude, 10 degrees in longitude and 10 km in height. Third order interpolation is used in all grid variables to find the model value where needed. For ray tracing applications splines can be used over the region of interest instead of simple interpolation.

Two dimensional "data boxes" can enhance computing speed in "classical" models if used for coordinate transforms, e.g., to calculate MODIP and dip latitude within the International Reference Ionosphere (IRI).

By adding a topside F-layer to the existing DGR model which is based on the CCIR parameters we constructed a multi-purpose global model which is continuous in the first derivative of electron density. The topside of this NeUoG model is based on the following 3 parameters, modelled according to the experiences of John Titheridge: O^+ scale height at the F-layer peak, its height gradient, and $O^+ - H^+$ transition height (compare Leitinger et al. 1995, 1996, Titheridge 1976a,b, 1993). John Titheridge has used a semi-theoretical approach to find the parameter values from his interpretation of topside sounder data. The second version of NeUoG uses a magnetic field aligned approach for the plasmasphere. We describe our method of field aligned modelling and the plasmasphere part of NeUoG-plas in some detail. We present graphics on the behaviour of the plasmasphere part.

Keywords: field aligned modelling; global distribution of electron density; magnetic field structure; plasmasphere

1. Introduction

The global ionospheric model NeUoG was constructed as an "easy to use" monthly median model to get reference values of electron density on a global scale and with continuity of values and first derivatives (Leitinger et al. 1995, 1996). For the bottomside including the F2-layer peak, NeUoG is based on the standard ionosonde parameters E and F region critical frequencies (f_0E, f_0F1, f_0F2) and the F2-layer transfer parameter M(3000)F2. For the topside, NeUoG uses models for the O^+ scale height at the peak, its height derivative and for the O^+–H^+ transition height. Input parameters are a solar activity measure (sunspot numbers or solar radio flux), season (month of the year), time (universal or local), (geographic) latitude, longitude and height.

The bottomside of NeUoG is essentially the (modified) Di Giovanni-Radicella model (mDGR — Radicella and Zhang 1995) which uses f_0E, f_0F1, f_0F2, M(3000)F2

[1]Institut für Meteorologie und Geophysik, Universität Graz, Halbärthgasse 1, A–8010 Graz, Austria, e–mail: leitinger@bkfug.kfunigraz.ac.at

1217-8977/97/$ 5.00

as anchors for five semi-Epstein layers ("E-layer" bottom and top, F1-layer bottom and top, F2-layer bottom). With the CCIR parameters mDGR can be used as a global model but does not give a realistic E-layer: the peak height hmE is fixed with 120 km and in lower latitudes the combination of the bottom E and the bottom F1 semi-Epstein layers give unrealistically high electron densities below 90 km. Therefore we used mDGR for an "F region only" model and took measures to decrease equatorial electron density below 90 km to realistic values.

It turned out that for a few input combinations the addition of Epstein layers gave problems just below the F2 peak. We avoid these now by using power law interpolation between the (mDGR) F1 and F2 peaks.

The first version of NeUoG uses an approximation to radial O^+–H^+ diffusive equilibrium for the topside ionosphere with three model parameters: the oxygen scale height at the F2 layer peak, its height gradient, and the O^+–H^+ transition height. Furthermore a ceiling is used for the increase of the scale height with height. Above this ceiling the scale height remains constant (pure exponential height profile). The ceiling height is 1.5 times the transition height or 1500 km whatever is smaller.

The "plasmasphere" version of NeUoG (NeUoG-plas) introduces a magnetic field aligned H^+ diffusive equilibrium above a "plasmasphere foot height". Outside the plasmapause electron density is reduced to 1/100 of the diffusive equilibrium value.

As is the "ionosphere only" version, NeUoG-plas is continuous in all first derivatives, a necessary requirement for ray tracing applications. Furthermore, its FORTRAN formulation is based on double precision variables and constants in order to allow for high-precision applications (e.g., ray-tracing to millimetric accuracy). Therefore calculation time is comparatively large. For regional applications we recommend our "data box" approach. It is based on a three dimensional grid in space with grid point distances of 5 degrees in latitude, 10 degrees in longitude and 10 km in height for the ionosphere from 80 to 800 km. For the uppermost ionosphere and the plasmasphere height spacing can be substantially wider (e.g., 50 km) and for the plasmasphere a height spacing of 100 km is probably sufficient.

A complete "data box" has to take into account solar activity, season and time. For electron density three levels of solar activity are recommended (low, medium, high) and quadratic interpolation. Seasons: the minimum is four boxes: winter (January), spring equinox (April), summer (July), autumn equinox (October). For finer resolution 12 boxes are recommended (one for each month). This approach allows to derive smooth annual variation by means of Fourier interpolation or cyclic third order interpolation (with comparable results). The diurnal variation is gained by means of boxes for each hour in local time. Smooth diurnal curves are obtained with Fourier interpolation or with cyclic third order interpolation. If details of the diurnal curve are not important, one can cut down to a spacing of two hours (12 local time values). A full set of "data boxes" corresponds to a 6 dimensional grid: $3 \times 12 \times 24$ space grid boxes. The amount of grid point values is certainly too high to allow application on a global basis. Space grids only are certainly feasible on a global scale too: the plasmasphere examples of section 4 are based on such grids.

We want to mention that the data box approach is also used inside NeUoG: two

dimensional third order interpolation is used to derive the (CCIR) dip latitudes from a data box avoiding to calculate this quantity by means of a spherical harmonics expansion (the method used by the International Reference Ionosphere). NeUoG-plas uses three data boxes: one for dip latitudes, one for the latitude of equivalent dipole poles, the third for pole longitudes.

2. Foundations of the plasmasphere part

2.1 The geometric/geomagnetic situation

Symbol conventions:
Lower case greek letters are used for geographic coordinates (φ, λ: latitude, longitude; $\vartheta = \pi/2 - \varphi$: pole distance).
Magnetic coordinates are written with upper case greek letters (Φ, Λ, $\Theta = \pi/2 - \Phi$).

Magnetic features and coordinates

To be consistent with the NeUoG peak (and the bottomside formulation of NeUoG) the CCIR spherical harmonics expansion is used as the magnetic field model and all quantities are derived from it in a consistent way, e.g., the Centric Dipole (pole P at (φ_p, λ_p) geographic) and the Dip Equator $\mathcal{E}_d$.

Equivalent (centric) dipole field lines (Fig. 1) are constructed in the following way: they are in great circle planes (= magnetic meridians) through the centric dipole pole but apex above the dip equator. This is ensured by defining "equivalent poles" F in nautical distances of $\pi/2$ from the crossing points E of the magnetic meridians with the dip equator ($\overline{\mathrm{EF}} = \pi/2$).

The Equivalent Magnetic Coordinates (Φ_e, Λ_e) of the point A with geographic coordinates (φ, λ) are found through the nautical triangle F–N–A (N: geographic pole).

The "equivalent pole" F with geographic coordinates (φ_P (ϑ_P), λ_E) is found at the P–F–E great circle path in an angular distance of $\pi/2$ rad from E. It is appropriate to use the Northern equivalent pole even for field points South of the dip equator. (An iterative procedure has to be applied to find E by means of the spherical harmonic expansion of the CCIR geomagnetic field.) The equivalent dipole field line is defined by means of the equivalent dipole latitude Φ_e (or equivalent pole distance $\Theta_e = \pi/2 - \Phi_e$) which is calculated with the nautical triangle F–N–A, N being the geographic North pole (Fig. 1). Equivalent dipole latitude and height defines the field line and its L-value.

Applications: to avoid the lengthy iterative calculations of the equivalent poles a grid point method is used: equivalent pole locations are calculated for a geographic grid with latitude spacing of 5 degrees and longitude spacing of 10 degrees. The equivalent pole positions for geographic locations between the grid points is found by means of a two dimensional third order interpolation procedure. The accuracy obtained is by far sufficient. The conversion of geographic latitude into equivalent dipole latitude could be done by means of a similar interpolation scheme: calculation of the equivalent dipole latitudes for the grid points and third order interpolation.

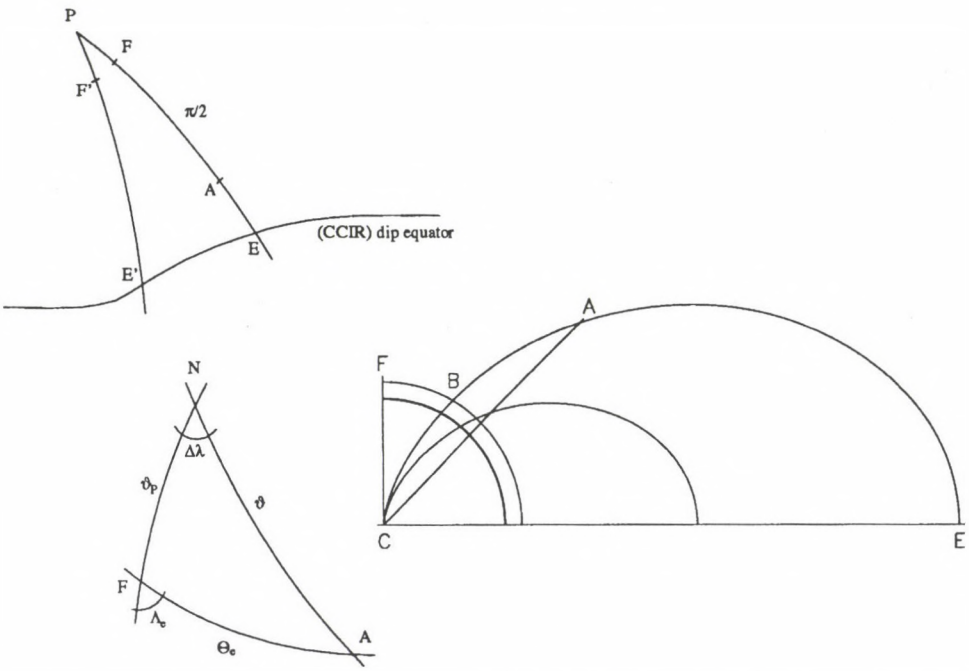


Fig. 1. Geometric/geomagnetic situation. P: dipole pole, F, F': equivalent poles, E, E': crossing points with dip equator, N: geographic pole, A: point in consideration. The nautical distance FE (or F'E') is $\pi/2$. ϑ, λ: geographic coordinates; Θ, Λ: equivalent dipole coordinates. Left: projection on nautical sphere. Right: Equivalent dipole field lines in one magnetic meridian. The field line through A peaks at E over the dip equator and has its foot point at B. C marks the center of the Earth

The equivalent pole locations for the grid points vary in geographic latitude between 75.32°N and 82.76°N, in geographic longitude between 229.39°E and 300.74°E (Fig. 2). The CCIR dipole pole has the geogr. coordinates (78.73°N, 290.32°E). The maximum for the difference between CCIR dipole latitude and equivalent dipole latitude occurs in the geogr. longitude zone around the Greenwich meridian where the difference between dipole and dip latitudes is largest (largest grid point difference: 12.33° for the point (10°N, 0°E)).

Remark: the CCIR mapping procedure uses MODIP calculated from dip latitudes for a height of 300 km. Accordingly the equivalent poles are calculated with the CCIR dip equator for 300 km height.

"Mapping" along equivalent dipole field lines (Figs 1 and 2): the field line equation is $r = \ell \cos^2 \Phi_e = \ell \sin^2 \Theta_e$. The field line is defined by its apex distance ℓ. With given coordinates (r_1, Φ_1) (r_1: distance from center of Earth) of one point, another point on the field line is found by using $r_2/\cos^2 \Phi_2 = r_1/\cos^2 \Phi_1$ etc.

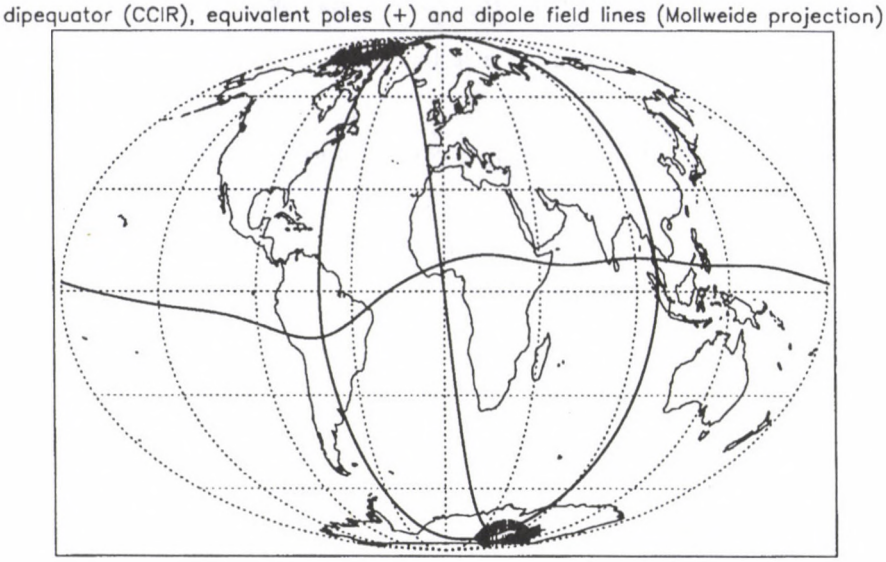


Fig. 2. CCIR dip equator, equivalent poles, and sample dipole field lines on a world map

2.2 Magnetic (dipole) flux tubes for "plasmasphere" modelling

Differential distance between two field lines in one magnetic meridian:

$$da = \sin I\, dr; \quad dr = sin^2\vartheta\, d\ell, \quad \text{tg } I = 2\,\text{ctg}\,\vartheta, \quad \longrightarrow \quad da = \frac{\sin^3\vartheta}{\sqrt{1+3\,\cos^2\vartheta}} d\ell$$

Differential cross section of a dipole flux tube:

$$df = da\,(r\,\sin\vartheta)\,d\lambda = \frac{\sin^6\vartheta}{\sqrt{1+3\,\cos^2\vartheta}}\,\ell\,d\ell\,d\lambda$$

(The widening of the flux tube when going up corresponds to the decrease of the field strength — the cross section is inversely proportional to the field strength.)

Diffusive equilibrium along the field line: $dz = \frac{ds}{H_{||}}$; $H_{||} = H\frac{g}{g_{||}}$;

$$g_{||} = g\,\sin I = \quad = \quad \frac{2\,\cos\vartheta}{\sqrt{1+3\cos^2\vartheta}}; \quad \longrightarrow \quad dz = \frac{dr}{H}, \qquad \text{because} \quad ds = \frac{dr}{\sin I}.$$

At quick look, this could be interpreted that the diffusion "does not care" about the dipole magnetic field (the projection of g onto s is compensated by the projection of ds onto dr).

However, one has to pay attention to the widening of the flux tubes, meaning that with $H \to \infty$ the electron density along the field line is proportional to the field strength, otherwise we have to proceed in two steps: (1) diffusive equilibrium

with $f(s) = \exp(-z) = \exp(-\int ds/H_{||})$; (2) "dilution" by the widening of the flux tubes. Formally this gives along the field line

$$N = N_o \frac{B(s)}{B_o} \exp\left[-\int_{s_o}^{s} \frac{\cos\vartheta(s)}{\cos\vartheta_o} \frac{1+3\cos^2\vartheta_o}{1+3\cos^2\vartheta(s)} \frac{ds}{H(s)}\right]$$

(index $_o$: quantities at $s = s_o$).

The substitution $ds = \sin\vartheta \sqrt{1+3\cos^2\vartheta}\, \ell\, d\vartheta$ gives with *constant* H for the integral

$$\begin{aligned} J &= \int_{s_o}^{s} \frac{\cos\vartheta(s)}{1+3\cos^2\vartheta(s)} \frac{ds}{H} = \frac{\ell}{H} \int_{\vartheta_o}^{\vartheta} \frac{\sin\vartheta\cos\vartheta}{\sqrt{1+3\cos^2\vartheta}}\, d\vartheta = \\ &= -\frac{\ell}{3H} \sqrt{1+3\cos^2\vartheta}\,\Big|_{\vartheta_o}^{\vartheta} \end{aligned}$$

therefore

$$N = N_o \frac{B(\vartheta)}{B_o} \exp\left[-\frac{\ell}{3H} \frac{1+3\cos^2\vartheta_o}{\cos\vartheta_o} \left(\sqrt{1+3\cos^2\vartheta_o} - \sqrt{1+3\cos^2\vartheta}\right)\right].$$

The coordinates of the start point (r_o, ϑ_o) give $\ell = \dfrac{r_o}{\sin^2\vartheta_o}$.

At the magnetic equator ($\vartheta = \pi/2$) we get

$$N_{\text{equ}} = N_o \frac{\sin^6\vartheta_o}{\sqrt{1+3\cos^2\vartheta_o}} \exp\left[-\frac{\ell}{3H} \frac{1+3\cos^2\vartheta_o}{\cos\vartheta_o} \left(\sqrt{1+3\cos^2\vartheta_o} - 1\right)\right].$$

Example: $\vartheta_o = 45°$ ($\ell = 2\,r_o$) : $\rightarrow$ $N_{\text{equ}} = 0.079057\, N_o \exp\left[-2.05464 \dfrac{2\,r_o}{3\,H}\right] =$

$= (0.079057\, N_o)\, 0.13533$ with $\dfrac{r_o}{H} = 1.4601.$

A constant scale height H is not realistic because the decrease of g is not taken into account. Even with constant composition (H^+ — plasmasphere) and constant plasma temperature (a realistic assumption), we have to use $H = H_o\, g_o/g = H_o\, r^2/r_o^2 = H_o \sin^4\vartheta/\sin^4\vartheta_o$.

The square root expansion

$$\begin{aligned} \frac{1}{\sqrt{1+3\cos^2\vartheta}} &= \frac{1}{2\sqrt{1-(3/4)\sin^2\vartheta}} = \\ &= \frac{1}{2}\left(1 + \frac{1}{2}\frac{3}{4}\sin^2\vartheta + \frac{3}{8}\frac{9}{16}\sin^4\vartheta + \frac{5}{16}\frac{27}{64}\sin^2\vartheta\cdots\right) \end{aligned}$$

and the integrals

$$\mathcal{I}_k = \int \frac{\sin\vartheta\cos\vartheta}{\sin^4\vartheta} \sin^{2k}\vartheta\, d\vartheta;$$

$$\mathcal{I}_0 = \int \frac{\cos\vartheta}{\sin^3\vartheta}\,d\vartheta = \frac{-1}{2\sin^2\vartheta}; \qquad \mathcal{I}_1 = \int \frac{\cos\vartheta}{\sin\vartheta}\,d\vartheta = \ln(\sin\vartheta)$$

$$\mathcal{I}_k = \int \sin^{2k-3}\vartheta\cos\vartheta\,d\vartheta = \frac{1}{2k-2}\sin^{2k-2}\vartheta \qquad (k>1).$$

lead to the expansion

$$\begin{aligned}\mathcal{I}(\vartheta) &= \int \frac{\sin\vartheta\cos\vartheta}{\sqrt{1+3\cos^2\vartheta}\,\sin^4\vartheta}\,d\vartheta = \\ &= \frac{1}{2}\left(\mathcal{I}_0 + \frac{3}{8}\mathcal{I}_1 + \frac{3}{8}\frac{9}{16}\mathcal{I}_2 + \frac{5}{16}\frac{27}{64}\mathcal{I}_3\cdots\right) = \\ &= \frac{1}{2}\left[-\frac{1}{2\sin^2\vartheta} + \frac{3}{8}\ln(\sin\vartheta) + \frac{3\cdot 9}{8\cdot 16\cdot 2}\sin^2\vartheta + \frac{5\cdot 27}{16\cdot 64\cdot 4}\sin^4\vartheta\cdots + \right. \\ &\quad \left. + \binom{-\frac{1}{2}}{k}\left(-\frac{3}{4}\right)^k \frac{1}{2k-2}\sin^{2k-2}\vartheta\cdots\right]\end{aligned}$$

For $\vartheta = \pi/2$ we get $\mathcal{I}(\pi/2) = -0.1648097$, for small ϑ we can cut off after a few terms ($\vartheta = 45°$ requires 4 or 5 terms, depending on the accuracy wanted).

With $\mathcal{I}(\vartheta_o)$ (and the expansion for this integral) we get at the magnetic equator

$$N_{\text{equ}} = N_o \frac{\sin^6\vartheta_o}{\sqrt{1+3\cos^2\vartheta_o}} \exp\left[-\frac{\ell}{H_o}\frac{1+3\cos^2\vartheta_o}{|\cos\vartheta_o|}\left(-0.1648097 - \mathcal{I}(\vartheta_o)\right)\right].$$

An additional height dependence of plasma temperature would require numerical integration to solve the integral, except when the scale height is chosen to be proportional to $(\sin\vartheta/\sin\vartheta_o)^\gamma$.

Example: $H = H_o(\sin\vartheta/\sin\vartheta_o)^2$ leads to

$$\mathcal{I}(\vartheta) = \int \frac{\sin\vartheta\cos\vartheta}{\sqrt{1+3\cos^2\vartheta}\,\sin^6\vartheta}\,d\vartheta = \frac{1}{2}\left(\mathcal{I}_{-1} + \frac{3}{8}\mathcal{I}_0 + \frac{3}{8}\frac{9}{16}\mathcal{I}_1 + \frac{5}{16}\frac{27}{64}\mathcal{I}_2\cdots\right)$$

with $\mathcal{I}_{-1} = \int \frac{\cos\vartheta}{\sin^5\vartheta}\,d\vartheta = -\frac{1}{4\sin^4\vartheta}$ ($\mathcal{I}_0$, $\mathcal{I}_1$ $\cdots$ as above).

2.3 Adding a plasmasphere part to an existing ionospheric model

There are several possibilities to couple a height oriented ionospheric model with a magnetic field oriented plasmasphere model. One is to use the ionospheric model for the bottomside only and to start a magnetic field oriented modified O^+–H^+ diffusive equilibrium at the F-layer peak. The maximum at the peak could be forced by means of Chapman layer formulation. An other possibility is to use the ionospheric model up to a "foot height" for the plasmasphere.

For NeUoG-plas we have chosen this approach because it saves the carefully modelled topside parameters of the original NeUoG. Furthermore, with a plasmasphere foot height h_B well above the O^+–H^+ transition height h_{tr} it is justified to use a scale height increase with $(\sin\vartheta/\sin\vartheta_o)^\gamma$ (see above).

The plasmasphere part is matched to the ionosphere part by using the electron density $(N_e)_B$ and the scale height H_B found at the footpoint B (Fig. 1) at the foot height h_B. Starting with a point of interest in the plasmasphere, A (Fig. 1) we have to proceed as follows:

(1) Calculate the equivalent dipole coordinates for point A and find the footpoint B by mapping downwards along the equivalent dipole field line through A using the field line equation.

(2) Calculate the geographic coordinates (φ_B, λ_B) of B from its equivalent dipole ones and compute $(N_e)_B = (N_e)_B(\varphi_B, \lambda_B, h_B)$ and $H_B = H_B(\varphi_B, \lambda_B, h_B)$ using NeUoG.

(3) Calculate plasmaspheric electron content at A by field aligned mapping.

This field aligned approach is sufficient for "hemispheric" modelling if plasmasphere electron density is calculated by means of the foot point values in the same (equivalent) magnetic hemisphere in which we find the point of interest, A. If A is above the (CCIR) dip equator we have to decide whether to take the footpoint in the magnetic Northern or Southern hemisphere. For single point values or for height profiles this is still acceptable. For slant ray path profiles it is certainly not when the ray path crosses the dip equator: substantial differences in apex electron density (N_e above the dip equator) can occur when we cross from one hemisphere into the other. Large differences which can reach two orders of magnitude and more arise if the scale heights at the two footpoints are strongly different (summer-winter situation).

Consistent global use of the plasmasphere part needs a formulation which balances the hemispheric differences along the equivalent field line. The following weighted interpolation does this and gives a consistent solution for plasmaspheric electron density:

$$N_e(\Theta_e) = \frac{(N_e)_N f_N + (N_e)_S f_S}{f_N + f_S}, \quad f_N = (\Theta_S - \Theta_e)^\delta, \quad f_S = (\Theta_e - \Theta_N)^\delta;$$

$(N_e)_N$: electron density calculated with (magnetic) Northern hemisphere footpoint values, $(N_e)_S$: electron density calculated with (magnetic) Southern hemisphere footpoint values, Θ_e: equivalent magnetic pole distance for point of interest, Θ_N: e. m. pole distance for Northern hemisphere footpoint, Θ_S: e. m. pole distance for Southern hemisphere footpoint.

3. Parameter choices for NeUoG-plas

For the plasmasphere part of NeUoG-plas we have chosen a plasmasphere foot height of 2000 km and a plasmapause at L=5 (equivalent dipole apex). $h_B = 2000$ km ensures that the "ionospheric" scale height at the foot points is the H^+ scale height (h_B is sufficiently larger than the largest O^+–H^+ transition height).

The coefficient γ (additional increase of scale height along field line) has been chosen to be 0 (no additional increase).

The coefficient δ (hemispheric balance) has been chosen to be 1 (weighted linear interpolation) for the following reason. $\delta = 1$ gives the smoothest transition between

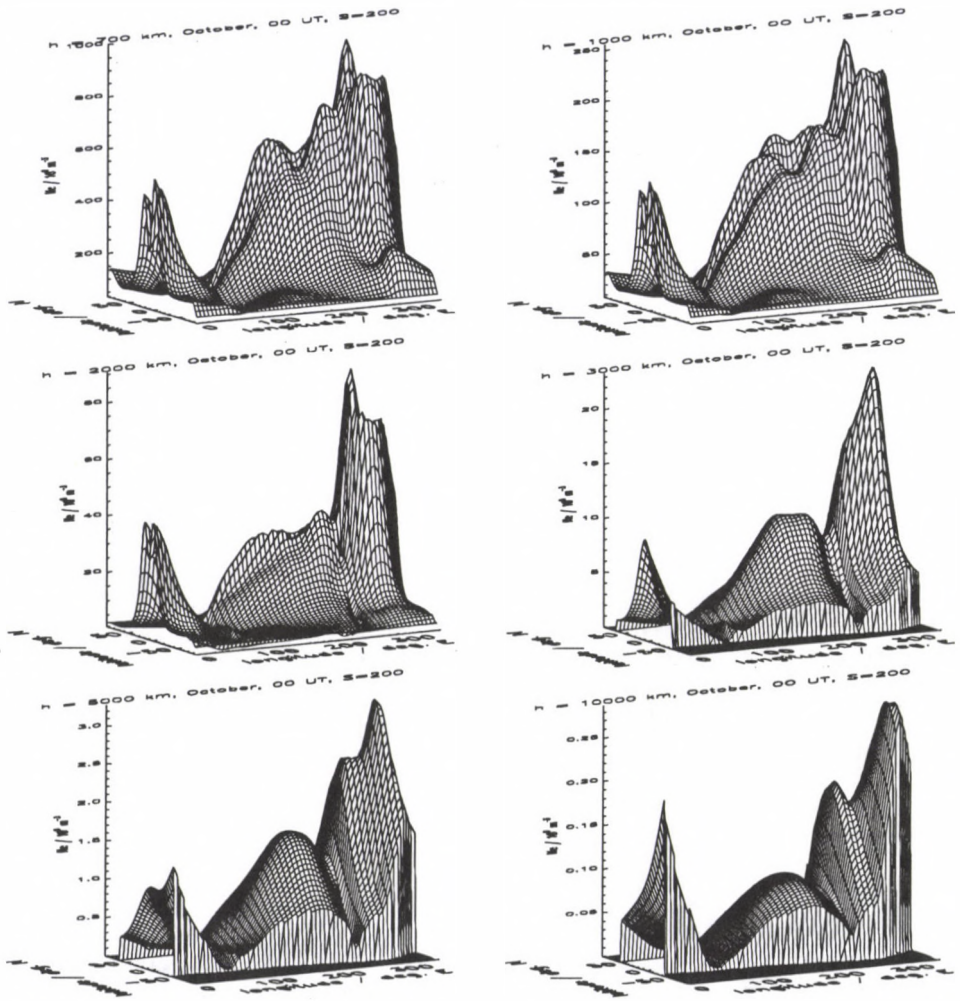


Fig. 3. NeUoG-plas example: electron density surfaces at heights 700, 1000, 2000, 3000, 5000 and 10000 km (top left to bottom right). Electron density in units of $10^9 m^{-3}$ over geographic coordinates (x-axis: longitude 0 to 360°E; y-axis: latitude -90 to 90°N). Data for October, 0 UT, high solar activity ($S_F = 200$). View from the Southern hemisphere

the Northern hemisphere field aligned profile and the Southern hemisphere one. $\delta > 1$ gives shorter transition regions but a strange form of transition. $\delta = 1$ still ensures that the "local" diffusive equilibrium is preserved near footpoints.

For reasons of continuity electron density values are needed outside the plasmapause. We have taken 1/100 of the "plasmaspheric" value (electron density is calculated using an equivalent dipole field line with $L > 5$ and the N_e-values obtained is divided by 100).

The "ionosphere only" version of NeUoG uses a ceiling height h_c to end the

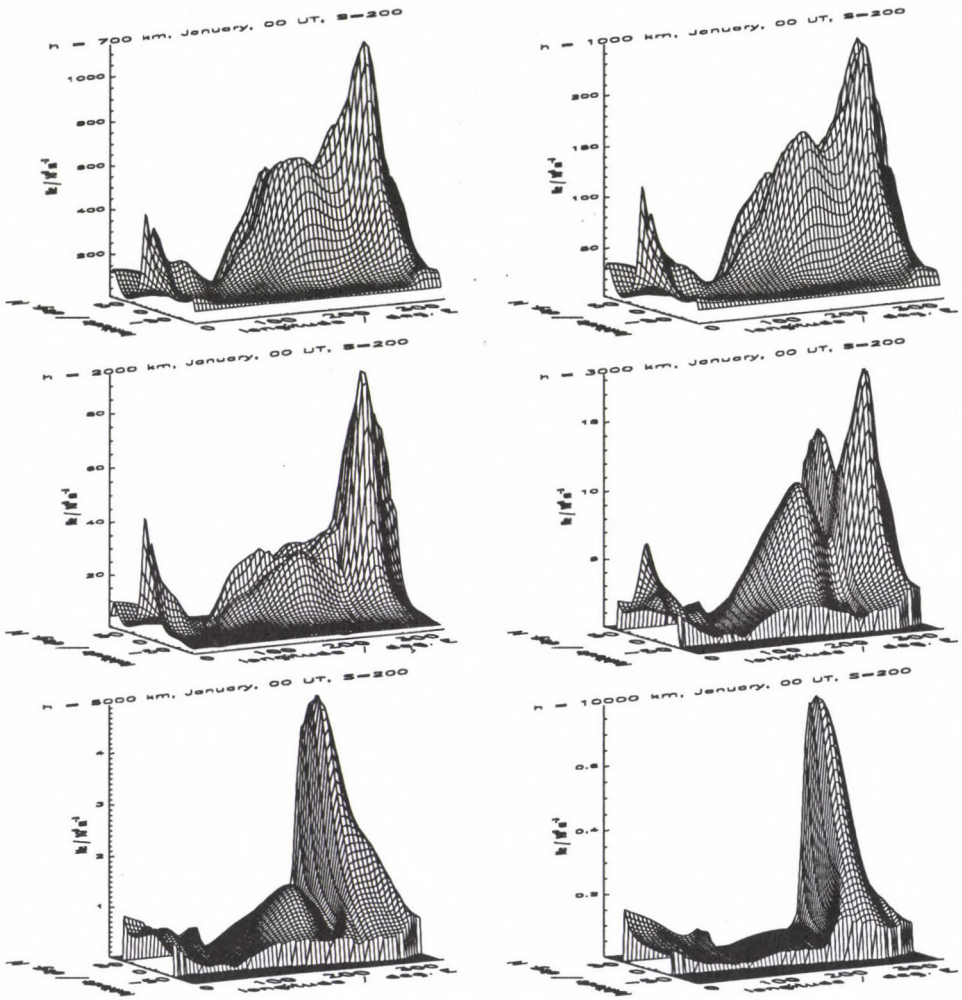


Fig. 4. NeUoG-plas example: electron density surfaces at heights 700, 1000, 2000, 3000, 5000 and 10000 km (top left to bottom right). Electron density in units of $10^9 m^{-3}$ over geographic coordinates (x-axis: longitude 0 to 360°E; y-axis: latitude -90 to 90°N). Data for January, 0 UT, high solar activity ($S_F = 200$). View from the Southern hemisphere

linear increase of the oxygen scale height with height ($h_c = 1.5\, h_{tr}$ or $h_c = 1500$ km whatever is smaller). NeUoG-plas uses no ceiling height. In the plasmasphere a quadratic increase with height occurs with the decrease of acceleration of gravity (see section 2).

The plasmasphere part too preserves continuity of first derivatives. Therefore it is necessary to use transition regions instead of sharp boundaries. One transition region is needed around the foot height, another one around the plasmapause.

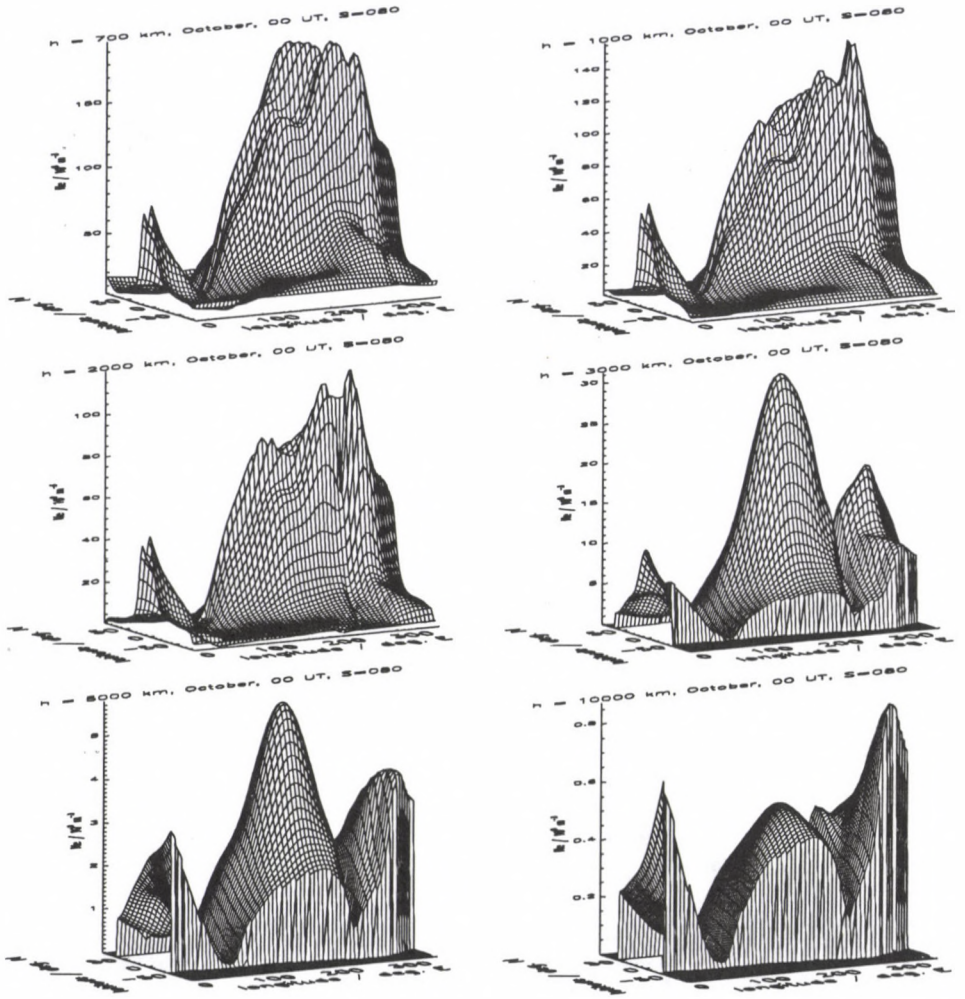


Fig. 5. NeUoG-plas example: electron density surfaces at heights 700, 1000, 2000, 3000, 5000 and 10000 km (top left to bottom right). Electron density in units of $10^9 \mathrm{m}^{-3}$ over geographic coordinates (x-axis: longitude 0 to 360°E; y-axis: latitude -90 to 90°N). Data for October, 0 UT, low solar activity ($S_F = 80$). View from the Southern hemisphere

The transitions are exponential:

$$N_e(h) = \frac{(N_e)_1(h)\,\alpha(h_o - h)] + (N_e)_2\,\alpha(h - h_o)]}{\exp[\alpha(h_o - h)] + \exp[\alpha(h - h_o)]}, \quad h_o : \text{center of transition region,}$$

$(N_e)_1$ from formula for $h < h_o$, $(N_e)_2$ from formula for $h > h_o$, α regulates the steepness of the transition. For the ionosphere-plasmasphere transition ($h_o = h_B$) we have chosen $\alpha = 0.05$ which leaves a 1% influence of $(N_e)_1$ at $h = h_o$+46 km. For the plasmapause we have chosen $\alpha = 0.035$ which leaves a 1% influence of $(N_e)_1$ at $h = h_o + 66$ km.

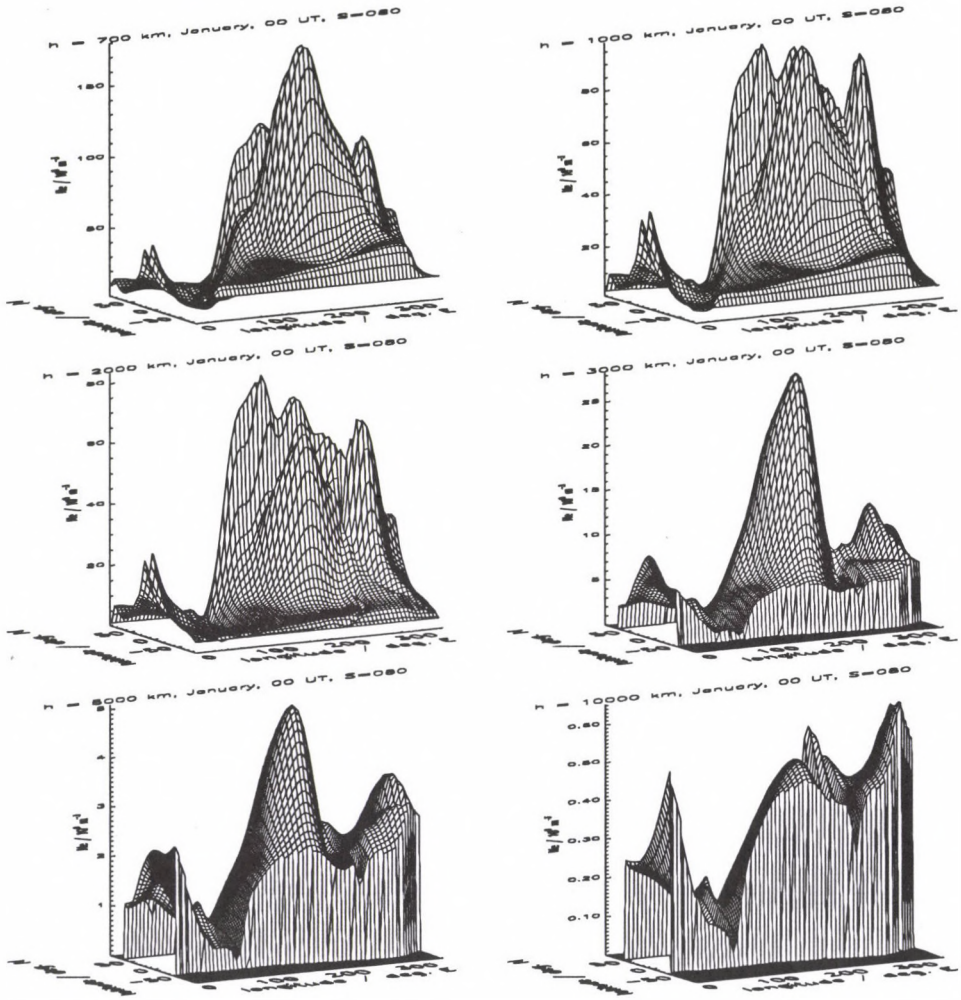


Fig. 6. NeUoG-plas example: electron density surfaces at heights 700, 1000, 2000, 3000, 5000 and 10000 km (top left to bottom right). Electron density in units of $10^9 m^{-3}$ over geographic coordinates (x-axis: longitude 0 to 360°E; y-axis: latitude -90 to 90°N). Data for January, 0 UT, low solar activity ($S_F = 80$). View from the Southern hemisphere

4. Some properties of NeUoG-plas

Some of the properties of the plasmasphere part of NeUoG-plas are demonstrated with the Figs 3 through 7. They show the global distribution of electron density in fixed heights (uppermost ionosphere: 700 km, 1000 km; plasmasphere foot height: 2000 km; plasmasphere: 3000 km, 5000 km, 10000 km). The "surface plots" give a view from the Southern hemisphere. As can be seen from Fig. 7, some of the structures in the Northern hemisphere are masked.

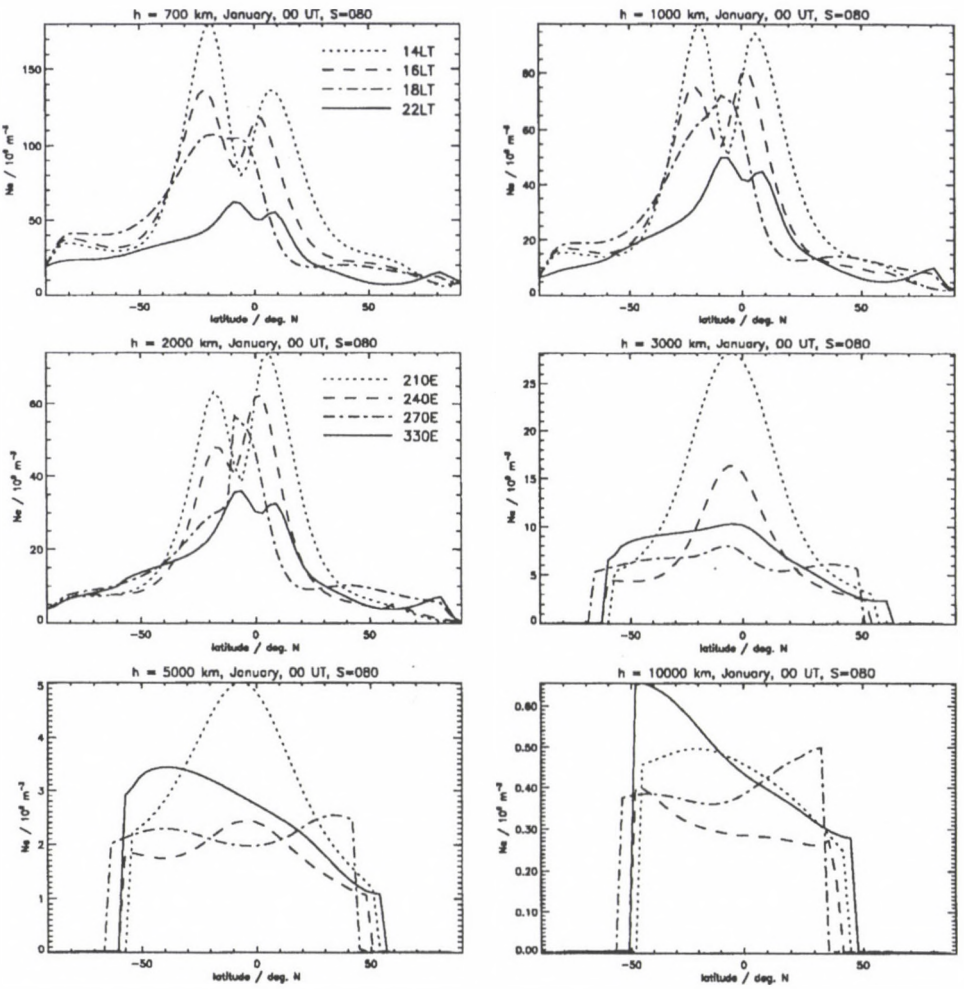


Fig. 7. NeUoG-plas example: cuts through the surfaces of Fig. 6 = latitude dependence of electron density at heights 700, 1000, 2000, 3000, 5000 and 10000 km (top left to bottom right). Longitudes: 210°E (dotted), 240°E (dashed), 270°E (dot-dashed), 330°E (full lines). Electron density in units of $10^9 \mathrm{m}^{-3}$ for January, 0 UT, low solar activity ($S_F = 80$)

The N_e surfaces at $h = 5000$ and $h = 10000$ km demonstrate the strong influence of the footpoint scale height which is much larger during daytime than during nighttime. Consequently, at large constant heights the daytime values of N_e show a strong dependence on (equivalent dipole) latitude. The "night"-"day" transitions are very sharp and the local time dependence of N_e for constant geographic coordinates (height, latitude, longitude) is very large too.

Figure 7 clearly shows that the equatorial anomaly "fades out" with height. Already at 3000 km there is only one peak in the vicinity of the dip equator instead

of the two anomaly crests (latitude dependence of electron density for the longitudes 210°E and 240°E).

Remark: Substantial differences between the "old" NeUoG (Version 1) and NeUoG-plas will occur only in the case of small scale height and/or large transition height. A large scale height at the foot of the plasmasphere means comparatively small decrease of electron density with height both for the "old" NeUoG and NeUoG-plas.

References

Leitinger R, Titheridge J E, Kirchengast G, Rothleitner W 1995: A "simple" global empirical model for the F layer of the ionosphere. Wiss. Bericht 1/1995, IMG Uni Graz

Leitinger R, Titheridge J E, Kirchengast G, Rothleitner W 1996: *Kleinheubacher Ber.* 39, 697–704.

Radicella S M, Zhang M-L 1995: The improved DGR analytical model of electron density height profile and total electron content in the ionosphere. *Annali di Geofisica*, 38, 35–41, 1995

Rawer K 1990: *Adv. Space Res.*, 10, (11)75–(11)78.

Titheridge J E 1976a: *Planet. Space Sci.*, 24, 229–245.

Titheridge J E 1976b: *Planet. Space Sci.*, 24, 247–259.

Titheridge J E 1993: Atmospheric winds calculated from diurnal changes in the mid-latitude ionosphere. *J. atmos. terr. Phys.*, 55, 1637–1659.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 343–354 (1997)

THE REGIONAL TEC MODEL DEVELOPED IN GRAZ — A PROGRESS REPORT

E FEICHTER [1] and R LEITINGER[1]

In the frame of PRIME (Prediction and Retrospective Ionospheric Modeling over Europe) a regional TEC model has been developed from data gained by the Differential Doppler effect on the 150/400 MHz signals of the polar orbiting NNSS satellites. The model uses 4 sets of 5 × 5 Fourier coefficients, 2 sets for high solar activity (HSA), 2 for low solar activity (LSA). In each case one set is valid for 52.5°N, the other gives the latitudinal gradients. The Fourier terms represent the seasonal and local time mean (time independent term), the annual, the semi-annual, the diurnal, the semi-diurnal components and the combinations of annual and diurnal variation. The model was adopted by the PRIME project.

Various combinations of Differential Doppler with Faraday data were tested too, the Faraday data were used to improve the diurnal variation in the models.

Since differences in the TEC behaviour were found between odd and even solar cycles, two different regional models, one for odd, the other for even cycles, have been constructed. The differences concern the vernal-autumnal asymmetry around noon time: in the even cycles TEC was higher in spring than in autumn, the odd cycles showed the opposite (autumnal maxima).

Keywords: longitudinal variation; regional model; total electron content, TEC; vernal-autumnal asymmetry

1. Introduction

A regional TEC model has been developed in the frame of PRIME (Prediction and Retrospective Ionospheric Modeling over Europe). The range of validity in latitude is 35°N to 60°N, larger than the "PRIME area" (35°N to 55°N). In longitude it coincides with the "PRIME area" (10°W to 30°E). The model is based on data gained by means of the Differential Doppler effect on the 150/400 MHz signals of the polar orbiting NNSS satellites.

Several "hybride" models have been developed too combining data from Faraday effect observations on the VHF signal of SIRIO made at Firenze/Italy, Graz/Austria and Neustrelitz/Germany (1978–1982). Electron content from the Faraday observations has very good temporal resolution therefore the "hybride" models (acronym DF for "Differential Doppler-Faraday") can have more fine structure in the diurnal variation. The SIRIO observations were made for conditions of intermediate to high solar activity which means restriction of the "hybride" models to these conditions.

PRIME, however, had strict modelling guidelines. The models had to be valid for the whole range of solar activity. Furthermore it was decided to aim at a more or less uniform distribution of the differences between measured data and model values. Details of the diurnal variation (e.g., around sunrise) matter a lot when

[1]Institut für Meteorologie und Geophysik, Universität Graz, Halbärthgasse 1, A–8010 Graz, Austria, e-mail: feichter@bkfug.kfunigraz.ac.at

1217-8977/97/$ 5.00

relative differences are of interest. They play a minor role in minimizing absolute differences.

"Hybride" models are available for special applications. Here we concentrate on progress made with the pure "Differential Doppler" (DD) models.

Ionospheric electron content (TEC) from the Differential Doppler effect on signals from the US Navy Navigation Satellites (NNSS, formerly TRANSIT, with polar orbits, almost circular with a height around 1100 km) (Leitinger et al. 1975, Leitinger and Putz 1978) was the main data source for investigations of solar cycle to solar cycle differences. It is the best indicator for the solar cycle differences but the effect shows up in peak electron density (N_{max}) too.

Two European receiving stations have been in continuous coordinated operation since the beginning of 1975: Lindau/Harz in Germany (51.6°N, 10.1°E) and Graz in Austria (47.1°N, 15.5°E). The evaluation results are latitudinal profiles of ionospheric electron content, which means sequences of data equidistant in latitude (data distance 0.5° in geographic latitude of the 400 km ionospheric points). Up to the mid-80s, the observations from Lindau gave more material and were used for the statistical investigation for the interval 1975–1986 (solar activity ranging from low to high). The Graz data were applied indirectly, namely to calibrate electron content by means of the "two stations method" (Leitinger et al. 1975). For a high solar activity interval, 1988–1992, the observations from Graz were used because for a given latitude there is no significant difference in the monthly medians and quartiles calculated from Graz or Lindau data. From the latitudinal profiles the data for the geographical latitudes 60°N, 55°N, 50°N and 45°N were selected.

The TEC data used are monthly medians for two hours intervals. The data were divided into two classes: "Low Sunspot number" (LSA) ($\overline{R} \leq 40$, nominal monthly mean sunspot number for modelling purposes: $\overline{R} = 20$) and "High Sunspot number" (HSA) ($130 \leq \overline{R} \leq 170$ [cycle 21], $120 \leq \overline{R} \leq 180$ [cycle 22], resp., nominal value $\overline{R} = 150$). The traditional break between solar cycles occurs near the middle of solar minimum. For modelling purposes however, solar minimum is attributed to the new cycle. An appropriate condition must be defined for "solar minimum": e.g., $R_{12} < 40$. (The widening of the $\overline{R}$ interval for HSA/cycle 22 was necessary in order to ensure data from at least two years for each month.)

These selection criteria meant all data from 1975 and 1976 and data from July 1984 through December 1986 were included in the LSA class and selected months from 1978 to 1982 and from 1988 through 1992, resp. in the HSA class (see Feichter and Leitinger 1993).

The first statistical studies (see Feichter et al. 1988, 1990, 1991, Feichter and Leitinger 1993) were based on two LSA intervals (1975–1976/cycle 21 and 1984–1986/cycle 22) and on the cycle 21 HSA interval (1978–1982). In 1989 data from 1984–86 were compared with those from 1975–76 (Fig. 1, Feichter et al. 1990). Since 1995 we have enough data to compare the HSA period of cycle 22 (selected months from 1988–1992) with that of cycle 21 too. Both LSA and HSA data show a switch from an autumn maximum for cycle 21 to a spring one for cycle 22.

To extend the investigation to a wider range of solar cycles we included the annual variation of N_{max} from ionosonde scalings (see Feichter and Leitinger 1997).

Table I. Regional TEC models for cycle 21 (top half) and cycle 22 (bottom half): Regression line parameters (central values and slopes) for the 5 × 5 Fourier-coefficients. Horizontally: diurnal coefficients, vertically: annual coefficients. Left hand tables: low sunspot numbers (LSA), right hand tables: high sunspot numbers (HSA)

Cycle 21, LSA: Values for 52.5°N					
	a0	a1	a2	b1	b2
A0	59.8	-31.7	5.0	-13.6	-0.8
A1	-19.4	1.7	6.2	5.1	5.8
A2	-2.0	7.9	-2.0	2.2	-0.8
B1	5.6	2.0	-5.2	-2.7	-0.0
B2	-1.2	-3.5	-0.4	-1.1	0.1

Cycle 21, HSA: Values for 52.5°N					
	a0	a1	a2	b1	b2
A0	254.8	-142.1	32.1	-61.6	9.7
A1	-5.0	-63.3	30.0	-19.5	19.2
A2	-56.3	38.5	-2.7	23.6	2.9
B1	10.8	35.8	-20.6	-5.2	-6.3
B2	4.9	-2.4	-5.4	-17.4	-1.2

Slopes of regr. lines ($\Delta\phi = -5°$)					
	a0	a1	a2	b1	b2
A0	5.7	-0.2	0.6	-1.4	-0.5
A1	1.0	0.9	-0.0	1.8	0.4
A2	-0.4	-0.2	-0.8	-0.0	-0.8
B1	-0.4	-0.3	-0.4	-1.1	0.1
B2	0.4	0.7	0.2	0.7	0.3

Slopes of regr. lines ($\Delta\phi = -5°$)					
	a0	a1	a2	b1	b2
A0	22.9	-10.2	1.7	-9.0	-1.9
A1	-0.8	-1.0	-1.3	2.7	-3.1
A2	-10.8	2.8	-1.9	5.0	-2.7
B1	2.2	0.3	2.6	-3.1	0.3
B2	3.3	-0.5	1.5	-2.4	1.9

Cycle 22, LSA: Values for 52.5°N					
	a0	a1	a2	b1	b2
A0	58.8	-27.7	5.8	-14.6	-1.2
A1	-21.8	-1.6	6.2	9.1	6.9
A2	-1.2	9.5	-2.0	-1.0	-0.5
B1	11.0	0.3	-2.0	-4.7	-2.8
B2	-2.0	-5.1	-1.8	-0.7	2.1

Cycle 22, HSA: Values for 52.5°N					
	a0	a1	a2	b1	b2
A0	238.5	-147.2	33.4	-69.0	11.5
A1	-64.5	-39.6	27.9	-5.0	17.0
A2	-67.8	63.7	-3.3	41.3	-4.2
B1	20.5	-21.3	12.3	-11.8	4.8
B2	-18.7	16.9	-0.8	11.1	0.3

Slopes of regr. lines ($\Delta\phi = -5°$)					
	a0	a1	a2	b1	b2
A0	5.5	0.3	0.8	-1.4	-0.4
A1	1.2	-0.4	0.3	2.8	0.3
A2	-0.6	0.8	-0.1	-1.2	-0.5
B1	-0.2	-0.2	-0.1	-0.5	-0.4
B2	0.2	0.7	0.4	0.9	0.6

Slopes of regr. lines ($\Delta\phi = -5°$)					
	a0	a1	a2	b1	b2
A0	25.0	-11.0	2.5	-8.4	-0.3
A1	-9.7	0.5	-2.3	3.6	-2.2
A2	-9.8	6.6	-2.1	3.5	-1.0
B1	5.8	-5.5	1.6	-2.9	-2.6
B2	-4.6	2.4	-1.5	2.3	-1.8

The solar cycle to solar cycle differences nearly disappear when an "ionospheric" index is used instead of $\overline{R}$ or R_{12}. An alternative is to use two models, one for odd and one for even cycles (Fig. 4).

The 22 years periodicity aspect of our investigations is dealt with in another publication (Feichter and Leitinger 1997), here we stress modelling consequences.

The basis of our modelling approach is Fourier analysis in local time and in season (double Fourier analysis; Table I)(see Feichter et al. 1988).

2. The vernal-autumnal asymmetry in electron content over Europe

In the following we demonstrate the vernal-autumnal asymmetry in electron content over Europe by means of a few examples. For simplicity, we include all of the sunspot minimum 1975/1976 into cycle 21, all of the sunspot minimum 1984/86 into cycle 22.

Figure 1 contains the annual variation for a latitude of 50°N and for 12 LT. It was reconstructed from the first 3 Fourier components (mean, annual, semi-annual) for cycle 21 (top) and cycle 22 (bottom), LSA (left) and HSA (right). Figure 2 contains the diurnal variation of the first 3 Fourier components (amplitudes and phases) for HSA of cycle 21 and cycle 22 and for the latitudes 45°N (left), 50°N (middle), 55°N (right).

A more complete explanation is found in Feichter and Leitinger (1997). Here we note that both for LSA and HSA the main reason for the asymmetry is a shift of the phase of the annual component: during LSA the maximum of the one year component occurs in even cycles *one or two months earlier* than in odd cycles (Fig. 1), during HSA *one or two months later* (Fig. 1).

We want to remark that in the Southern hemisphere the vernal-autumnal asymmetry does not appear (Titheridge et al. 1996) but there is some indication that it appears (perhaps in modified form) in North American West Coast data (Feichter and Leitinger 1997).

A double sunspot-cycle variation also occurs in terrestrial magnetic activity. In even-numbered cycles the last half of the sunspot-number-cycle is more active than the first half and the converse is true for the odd-numbered cycles (Chernosky, 1966). The curve of the 22-years cycle of magnetic activity is not symmetric relative to the minimum between two cycles.

In 1988 the vernal-autumnal asymmetry in the seasonal variation of geomagnetic activity was investigated by Třisková using the *aa* index. Analysing the variation of geomagnetic activity, not from the viewpoint of sunspot cycles, but with respect to the polarity of the main solar dipole, she found an annual wave in the *aa* index with maxima alternatively around the vernal and the autumnal equinoxes (Triskova 1989). This *aa* index behaviour might help explain the double sunspot cycle in total electron content of the ionosphere (Fig. 3, Feichter and Leitinger 1997).

3. Modelling consequences

An easy way to construct local models for electron content (or for f_oF2) is to restrict the Fourier terms to a lower order. For many purposes restriction to 5×5 terms each for LSA and HSA and linear interpolation to account for varying solar activity is sufficient for TEC. (For f_oF2 one could use three levels of solar activity and quadratic interpolation.) The 5×5 terms comprise the average, the one day, the half day, the one year, the half year periods and combinations thereof. This approach was used for the "PRIME TEC map".

In modelling efforts for f_oF2 and other ionosonde parameters (e.g., M3000(F2)) no true longitudinal variation appeared over Europe in the sector 10°W to 30°E. Therefore at least for this sector regional models need only a local time variation

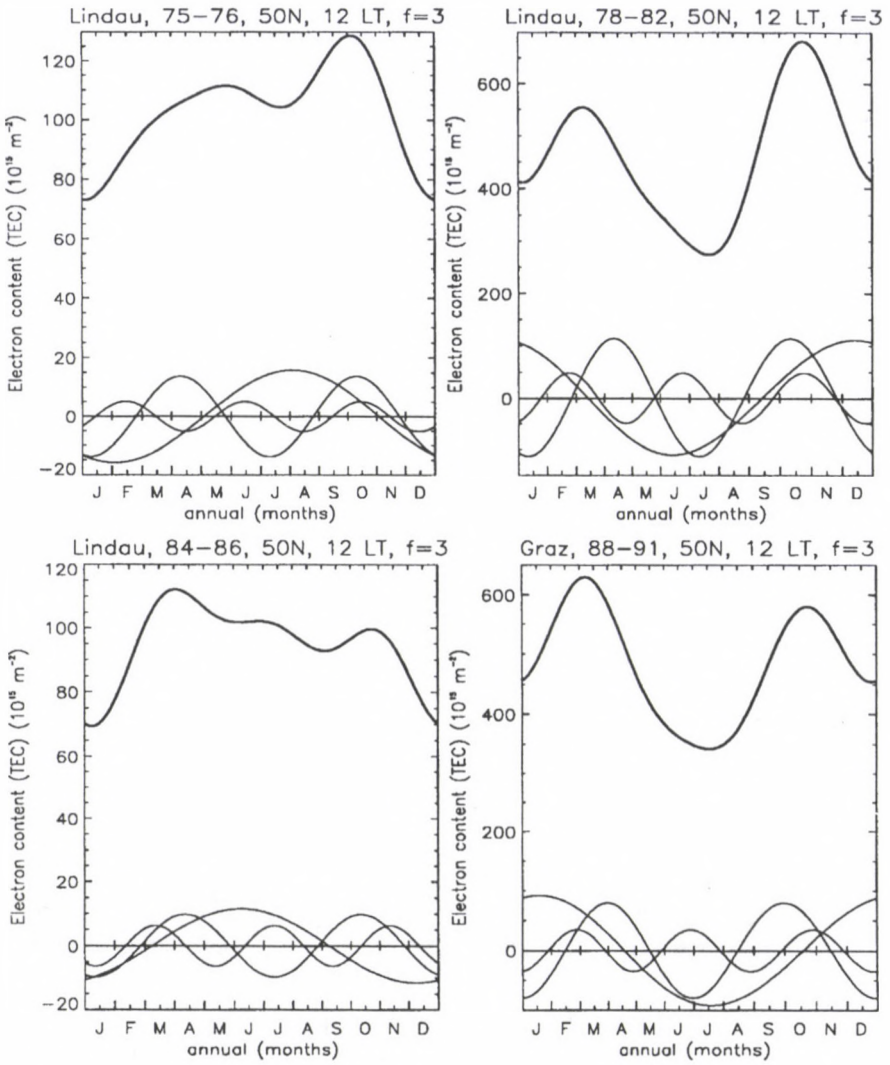


Fig. 1. Annual variation of ionospheric electron content (TEC) from four Fourier components (mean, one year, half year, four months). Low sunspot numbers ($\overline{R} = 20$) (left hand side), high sunspot numbers ($\overline{R} = 150$) (right hand side), solar cycles 21 (top) and 22 (bottom). The one year, half year, four months components are shown separately (thin lines). Based on bi-hourly monthly medians from months selected according to sunspot numbers criterium

and a latitude variation. To cover the latitude range 40°N to 60°N it is sufficiently accurate to adapt a linear latitude dependence for all Fourier coefficients by means of linear regressions. The region of latitudes < 40°N is already influenced by the equatorial anomaly. For PRIME a simple solution was adopted to account for this influence: for latitudes between 35°N and 40°N the latitudinal gradient was

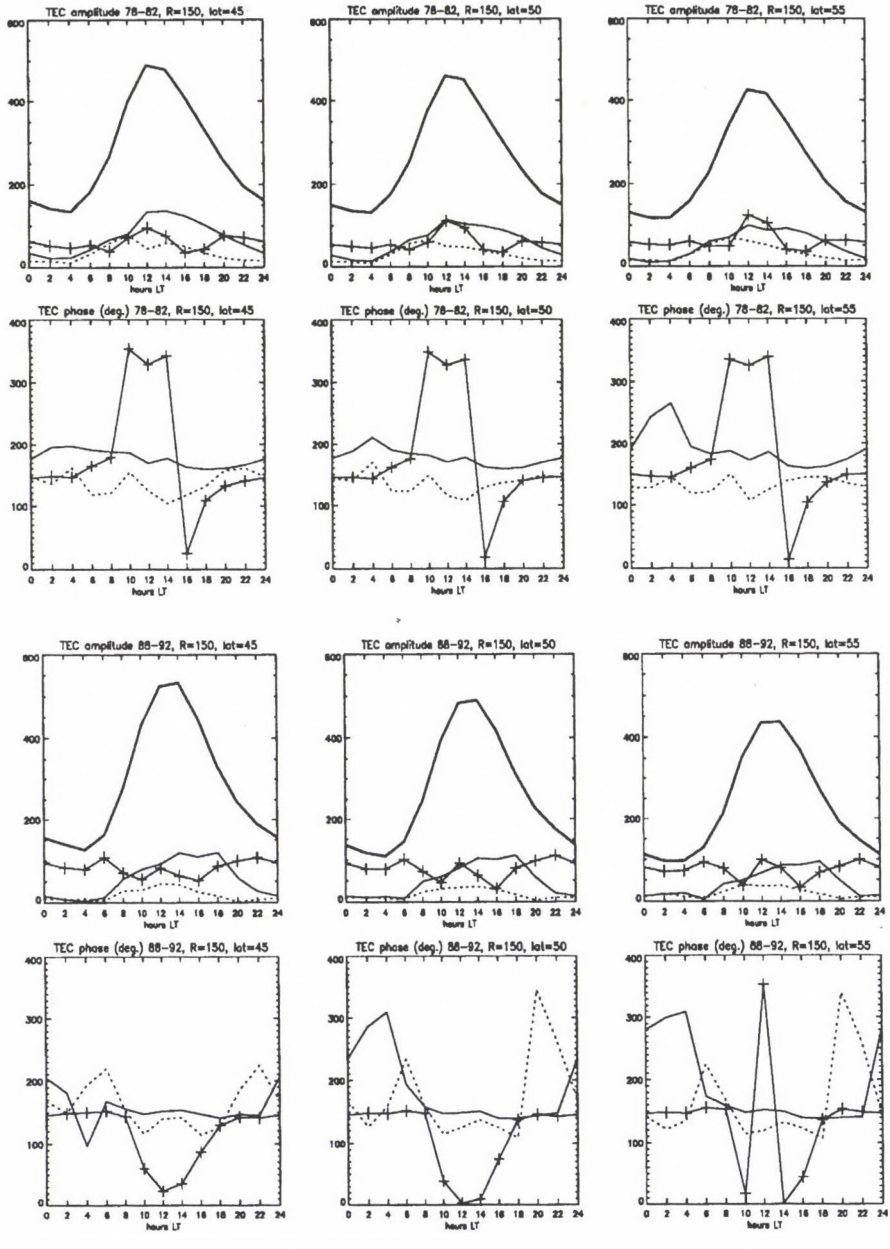


Fig. 2. Diurnal variation of amplitudes (top and third row) and phases (second and bottom row) for 4 (3) Fourier components of the annual variation of ionospheric electron content (TEC): mean (heavy lines), one year (marked with crosses), half year (thin line), four months (dashed line). High sunspot numbers ($\overline{R}$ = 150). Solar cycles 21 (top) and 22 (bottom). LT from 00 to 24 hours. Geographic latitude 45°N (left hand panels), 50°N (middle panels) and 55°N (right hand panels). Based on bi-hourly monthly medians from months selected according to sunspot numbers criterium

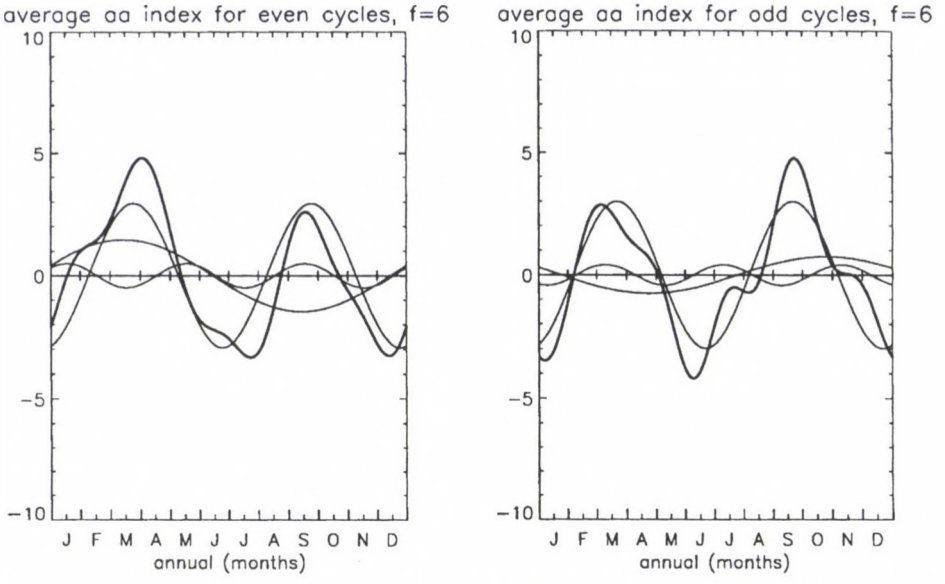


Fig. 3. Annual variation of the geomagnetic *aa* index (6 Fourier coefficients) (heavy lines). Left: average of 3 even cycles, right: average of 3 odd cycles. Decomposition into the first 3 Fourier coefficients (thin lines)

doubled. (Example: when the model electron content increases by 10 units from 45°N to 40°N it increases by 20 units from 40°N to 35°N.)

Finally the modelling procedure leads to 4 sets of 5 × 5 coefficients, namely mid point (52.5°N) values and latitudinal slopes for LSA and for HSA. An appropriate retrieval algorithm constructs the output values (linear interpolations in R12 and in latitude and Fourier reconstruction).

Table I gives the results for cycle 21 and cycle 22. In this case too comparison of the coefficients shows that the main cycle to cycle difference is found in the phases of the one year components (A1:B1; see lines 2 and 4 of the sub-tables for the central values).

Of course it is an open question whether the solar cycle to solar cycle change in the vernal-autumnal asymmetry of the F layer ionization is considered to be strong enough to justify the complication of cycle dependent modelling. If not it is recommended to base models on data from two or from four complete cycles. The Fourier analysis approach is a purely linear process and therefore allows easy construction of an average model from the separate even cycle/odd cycle models (Table I) simply by averaging for each coefficient. Examples for three DD models, one for odd cycles, one for even cycles, the third constructed as the average, are shown in Fig. 4. Conditions: low solar activity (LSA, $R_{12} = 20$, Fig. 4a and 4b), high solar activity (HSA, $R_{12} = 150$, Fig. 4c and 4d), geographic latitude 40°N (Fig. 4a and 4c), 60°N (Fig. 4b and 4d). The odd cycle model is identical to the model adopted by PRIME (COST 238).

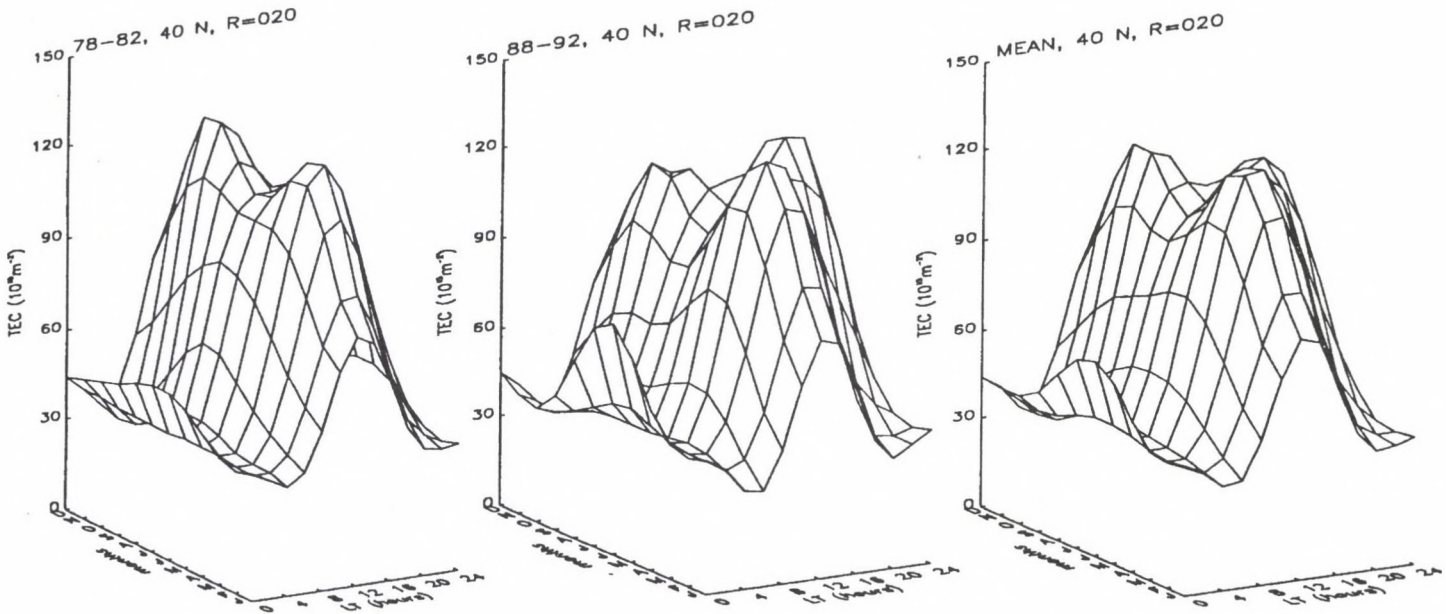


Fig. 4a. Sample results for three DD models, one for odd cycles, one for even cycles, the third constructed as the average. 3D-displays, ionospheric electron content over Local Time-Season (months) grid. Low solar activity (LSA, $R_{12} = 20$, geographic latitude 40°N)

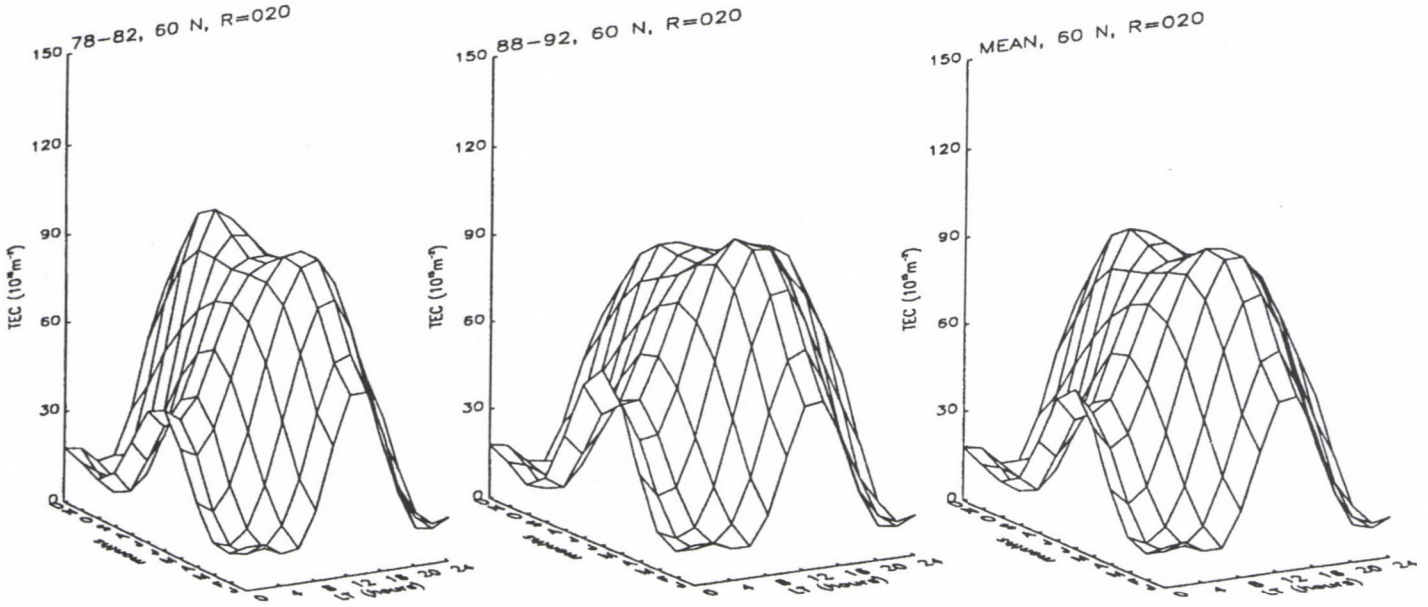


Fig. 4b. Sample results for three DD models, one for odd cycles, one for even cycles, the third constructed as the average. 3D-displays, ionospheric electron content over Local Time-Season (months) grid. Low solar activity (LSA, $R_{12} = 20$, geographic latitude 60°N)

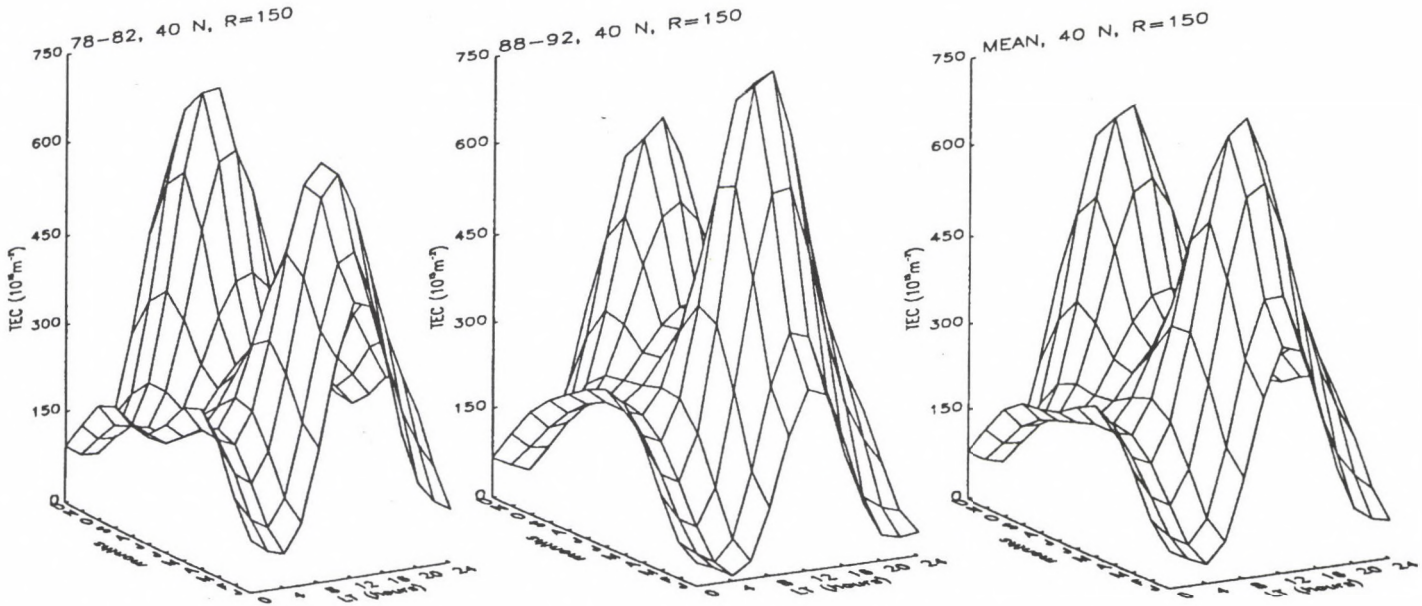


Fig. 4c. Sample results for three DD models, one for odd cycles, one for even cycles, the third constructed as the average. 3D-displays, ionospheric electron content over Local Time-Season (months) grid. High solar activity (HSA, $R_{12} = 150$, geographic latitude 40°N)

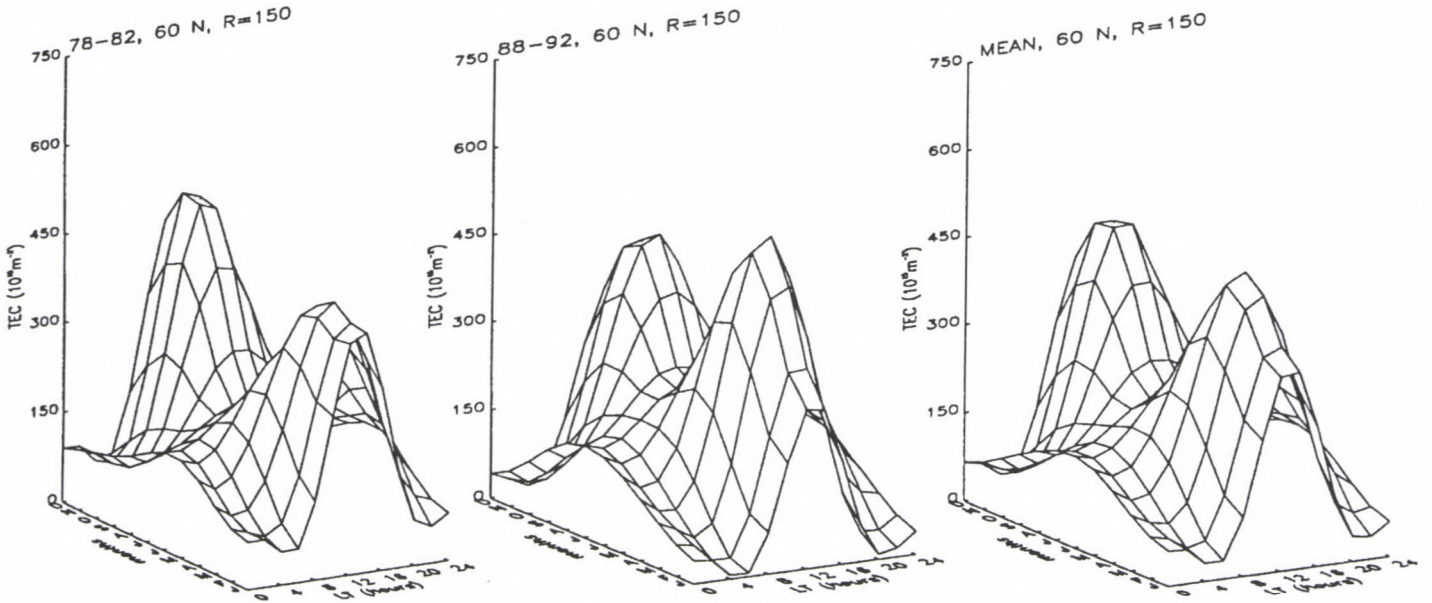


Fig. 4d. Sample results for three DD models, one for odd cycles, one for even cycles, the third constructed as the average. 3D-displays, ionospheric electron content over Local Time-Season (months) grid. high solar activity (HSA, $R_{12} = 150$, geographic latitude 60°N)

References

Chernosky E J: *J. Geophys. Res.* 71, 965–974, 1966
Feichter E, Leitinger R, Hartmann G K: *Kleinheubacher Ber.* 31, 249–258, 1988
Feichter E, Leitinger R, HartmannG K: *Kleinheubacher Ber.* 33, 93–102, 1990
Feichter E, Leitinger R, Hartmann G K: *Kleinheubacher Ber.* 34, 207–214, 1991
Feichter E, Leitinger R: Longterm studies of ionospheric electron content. Wiss. Ber. 1/93, Institut für Meteorologie und Geophysik, Universität Graz, 1993
Feichter E, Leitinger R: *Ann. Geophys.*, 15, 1015–1027, 1997
Leitinger R, Schmidt G, Tauriainen A: *J. Geophysics (Zs. Geophysik)* 41, 201–213, 1975
Leitinger R, Putz E: Die Auswertung von Differenz-Doppler-Messungen an den Signalen von Navigationssatelliten. Technischer Bericht, Universität Graz, 1978
Titheridge J E, RLeitinger R, Feichter E: *Kleinheubacher Ber.* 39, 749–755, 1996
Triskova L: *J. atm. terr. Phys.* 51, 111–118, 1989

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 355–363 (1997)

TEC RESIDUAL SLAB-THICKNESS BETWEEN BOTTOMSIDE AND TOPSIDE IONOSPHERE

T L GULYAEVA[1]

Observed TEC values during the disturbances cannot be adequately reproduced by integrating available model electron density profile along the ionospheric heights. It is shown that TEC and N_mF2 disturbances do not coincide for $\sim$ 50% of days (enhanced TEC with normal N_mF2 or reduced N_mF2 with normal TEC). In such cases presence of an additional layer between the bottomside and topside ionosphere is assumed; its thickness may range from zero to 300 km. Available ionospheric models should be revised inserting broad peak of ionization.

Keywords: electron density-height profile; ionosphere disturbance; slab-thickness; total electron content, TEC

1. Occasions of TEC enhancement

Integral electron content in the ionosphere affects operation of systems which rely on the transionospheric radiowave propagation. It serves as indicator of reliability of different ionospheric models both when TEC is obtained by integrating the vertical distribution of ionization, and by inverting the observed TEC into model electron density profile (Klobuchar and Leitinger 1993). When observed F2 layer peak parameters are used for updating the model Ne(h) profile, there still remain occasions of discrepancies between the observed and modelled TEC.

Latitudinal TEC variability during the global geomagnetic storm is shown in Fig. 1 hour-by-hour for nine days starting on 23 July 1981 for 6 locations (Table I). Equatorial electrojet D_{st} index had dropped below -200 nT on 25 July as it is seen in the bottom panel of Fig. 1. The TEC storm onset (enhancement followed by depletion) occurred at high latitudes prior to D_{st} decrease. At low latitudes (Ramey and Lunping) the TEC enhancement delayed by one to three days as compared with higher latitudes. However our subject of study requires to combine TEC and ionosonde observations.

Days of enhancement or depletion of the ionospheric ionization with parameters TEC, f_0F2, N_mF2 can be detected by analysis of 3 daily indices: M – mean daily value, P – the daily peak value, and R – the daily mean rate of variability:

$$\begin{aligned} M &= (\sum_{i=1}^{n} Y_i)/n\,, \qquad n \le 24 \\ P &= \max(Y_i)\,, \qquad i = 1,\ldots,n \\ R &= \left(\sum_{i=1}^{n} (Y_i - Y_{i-1}/(t_i - t_{i-1})\right)/n\,, \end{aligned} \tag{1}$$

[1]IZMIRAN, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia

1217-8977/97/$ 5.00

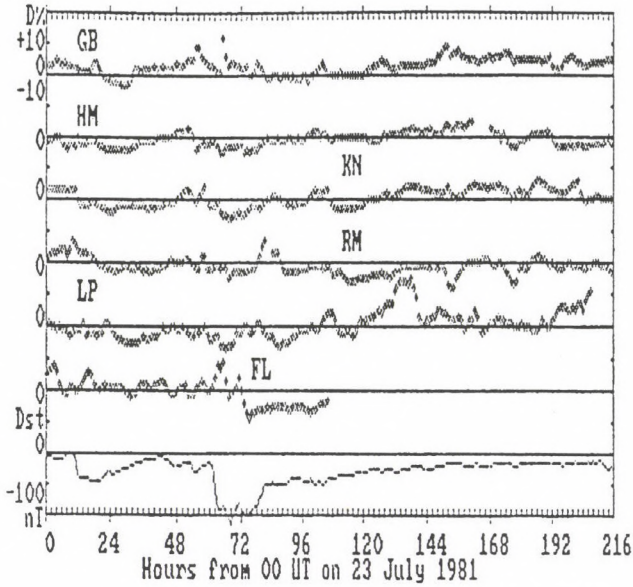


Fig. 1. Percentage departures of TEC from their monthly mean hour-by-hour starting from 00 UT on 23 July 1981 at 6 locations: GB – Goose bay, HM – Hamilton, KN – Kennedy Sapce Center, RM — Ramey, LP – Lunping, FL – Florence. Equatorial $D_s t$ index variation – bottom section

where Y_i means hourly values of TEC, f_0F2 or N_mF2, t – hour, LT. For each of indices M, P, R (1) the montly means and standard deviation is determined. When daily percentage deviation of any of the said indices from their monthly mean exceeds the percentage standard deviation, we will define it as the positive disturbed day (enhanced ionization) or the negative disturbed day (depletion of ionization).

Selection of the disturbed days by above criteria is illustrated in Fig. 2 for TEC observations in Florence (solid line) and f_0F2 in Rome (dashed line) for December, 1989. Standard deviation (dotted line for TEC and dashed line for f_0F2) as measure of dispersion shows greater variability of TEC than f_0F2. Results for all 3 indices (1) are in good agreement for TEC and f_0F2 though they are different in amplitude for many days and variability of two ionospheric quantities is opposite on some days. We will consider further effect of such differences on TEC modelling results.

Results of evaluation of the ionophseric disturbed days during a year of 1981 based on TEC and N_mF2 ionosonde observations at Goose bay (See Table I for coordinates) are presented in Table II where the positive disturbed day is designated as D♣ (or D+) and the negative disturbed day as D• (or D–) for TEC (or N_mF2). While N_mF2 is more variable than TEC for solar minimum and vice versa for solar maximum at middle and low latitudes (Gulyaeva and Spalla 1995, Jayachandran et al. 1995), the results presented in Table II for sub-auroral zone at high solar activity show greater number of days with the ionosphere peak disturbances (86 D+ days and 92 D– days) than with TEC (61 D♣ days and 51 D• days). Note that

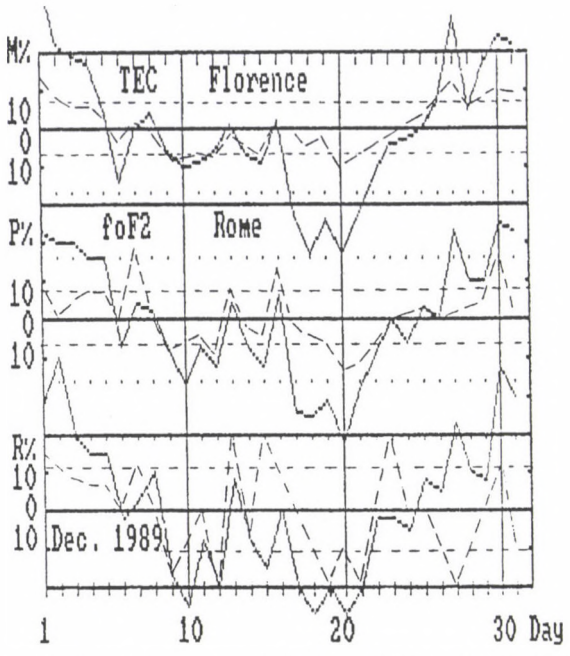


Fig. 2. Day-by-day variation of percentage deviation of TEC and f_0F2 daily indices from their monthly mean in Florence and Rome. Top section: daily mean value; middle section: daily peak value; bottom section: daily rate of mean hourly increment of TEC (solid line) and f_0F2 (dashed line)

the same criteria (1) are applied to the both parameters which allow to compare differences in their variability.

In particular, when enhancement of TEC occurs with normal (quiet) or reduced N_mF2 — on 25 days — this means that updating of the model electron density profile with observed N_mF2 would underestimate the TEC value. Similar failure to reach the observed TEC by updating Ne(h) profile with ionosonde N_mF2 observation happens when the peak electron density is depressed while TEC is normal (quiet) — on 48 days in Table II. Thus we have found considerable amount of about 50% of the ionospheric disturbed days when available procedures for model Ne(h) profile updating with ionosonde peak observations (Kishcha 1995) would fail to fit TEC observations.

2. Comparison of observations with model predictions

Example of comparison of measured TEC with IRI (Bilitza 1990) and PRIME (Bradley 1995) models is given in Fig. 3 for positive TEC and N_mF2 disturbance occurred in Florence (see Fig. 2) for winter high solar activity on 30 December, 1989. Observed TEC$_o$ are well represented by IRI and PRIME integrating electron density height profile at night. Daytime PRIME prediction is close to monthly median (as it is justified by the quiet model) but failed to correctly reproduce enhanced

Table I. Geomagnetic and geographic sub-ionospheric coordinates of the TEC observing sites and the ionospheric stations

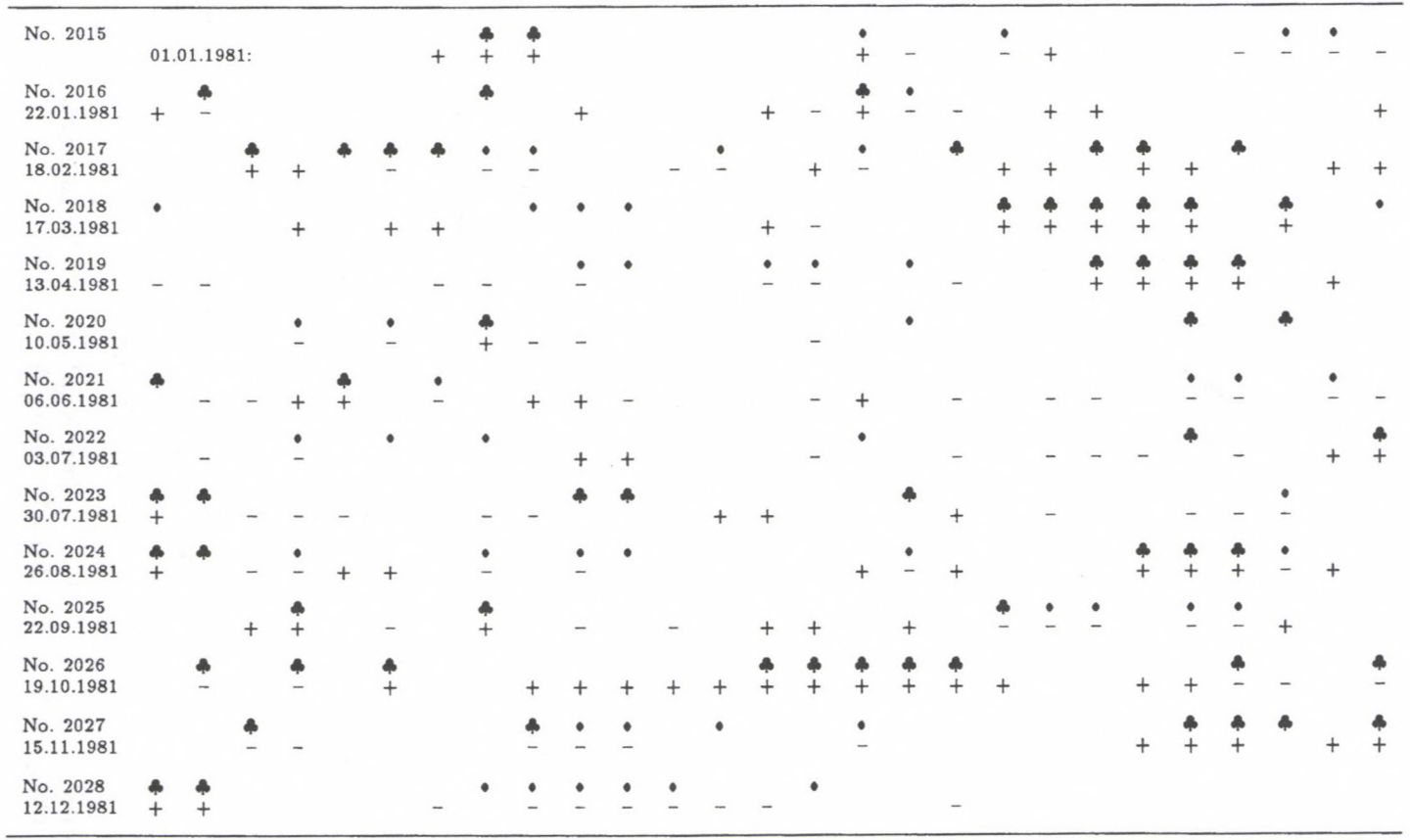
N	TEC location	Mlati/lati	Mlong/longi
1	Goose bay	59/48°N	10/298°E
2	Hamilton	50/39	358/289
3	Florence	41/40	90/11
4	Kennedy Space Center	37/26	348/280
5	Ramey Puerto Rico	28/17	3/293
6	Lunping	12/23	192/123
N	Ionosonde	Mlati/lati	Mlong/longi
7	Goose bay	64/53°N	14/300°E
8	Wallaps Is	49/38	353/285
9	Rome	43/42	92/12
10	Gibilmanna	38/38	93/14
11	Chung-Li	14/25	190/121

TEC_o variation. When f_0F2 data observed at Rome are used for IRI updating (IRI = NeOBS), they yield TEC_m at Florence for sunrise hours underestimating the observed TEC_o but exceeding TEC_o for the rest of day. Thus while fitting $Ne(h)$ model to observed f_0F2 data one should be careful if such procedure would improve TEC_m results or make them worse.

Another option for IRI updating is provided by using the F2 layer peak height. It is illustrated in Fig. 4 (a,b) for winter December, 1981, at high solar activity for low latitude Ramey, Puerto Rico, and (c,d) at Florence on the solar eclipse summer day of 11.07.1991. IRI-CCIR prediction is shown by dashed curve. Integrating $Ne(h)$ profile by IRI procedure, different peak heights h_mF2 at the same peak electron density N_mF2 (solid curve) yield almost the same results of TEC. The nighttime TEC_m for 22 h at Ramey (b) are equal to observed TEC_o. However TEC_m at daytime (a,c,d) underestimates TEC_o by 20 to 30%.

To overcome ambiguity in selection of the right $Ne(h)$ profile, satellite HINOTORY average Ne values at the height 600 km for northern winter monts 1981–1982 are relevant source in our case for Ramey obervations (Watanabe et al. 1995). They are shown by squares in Fig. 4 (a,b). The initial IRI Ne profile (dashed curve) fits HINOTORY data by night but differs by daytime. If we choose the IRI $Ne(h)$ profile (at h_mF2 between 370 to 420km) which fits well the satellite observations, the bottomside profile become too high for the daytime F1-F2 region altitudes. So it is not enough to vary N_mF2 and h_mF2 to correctly reproduce TEC observed data.

Table II. Bartels diagram ofthe ionospheric positive D♣ (D+) and negative D♦ (D−) disturbed days at Goose Bay during 1981 based on TEC (the upper symbols) and N_mF2 (the lower symbols) ionosonde observations. No. of Bartels Rotation and staring day for each line is indicated



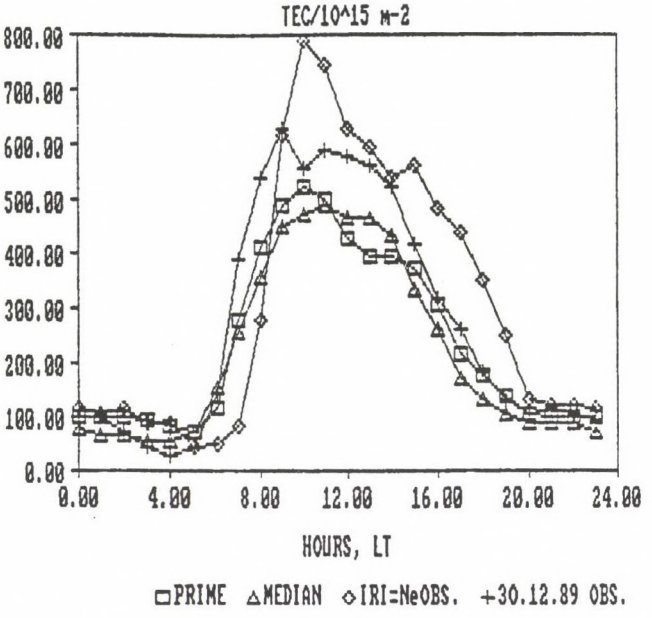


Fig. 3. Comparison of diurnal TEC observations with IRI and PRIME models and monthly median for the positive ionosphere disturbance over Florence on 30 December, 1989

TEC-based ionosphere peak structure

Taking into account enhanced TEC_o at the ionospheric disturbances — see Figs 1, 2, we assume that the broad peak electron density is observed in the ionosphere. Existence of such kind of thick peak was shown (Huang 1974) due to upward electromagnetic plasma drift observed by incoherent scatter technique at Jicamarca. It has been validated later on with TEC and slab-thickness data over Indian equatorial anomaly region owing the results to fountain effect over the peak of equatorial anomaly (Das Gupta et al. 1975). Lift of the F2 layer peak height upto 600 km at the crest of equatorial anomaly was modelled to fit TEC observations (Su et al. 1995).

The difference between observed and modelled TEC allows to calculate residual electron content (REC) between bottomside and topside ionosphere. Similar to definition of the equivalent slab-thickness TAU which is widely used as a shape factor of proportionality of total electron content in a column from the ground to the top of the ionosphere to the peak electron density N_mF2 (Davies and Liu 1991), the additional slab-thickness producing altitude step h_s is determined from intermediate REC value from the difference between TEC observed (o) and model (m) and N_mF2:

$$h_s = \mathrm{REC}/N_m\mathrm{F2}\,, \tag{2}$$

where REC = TEC_o–TEC_m. Final Ne(h) profile fitting observed TEC_o data including interlayer dashed area is shown in Fig. 4 by circles. For day-time we obtain

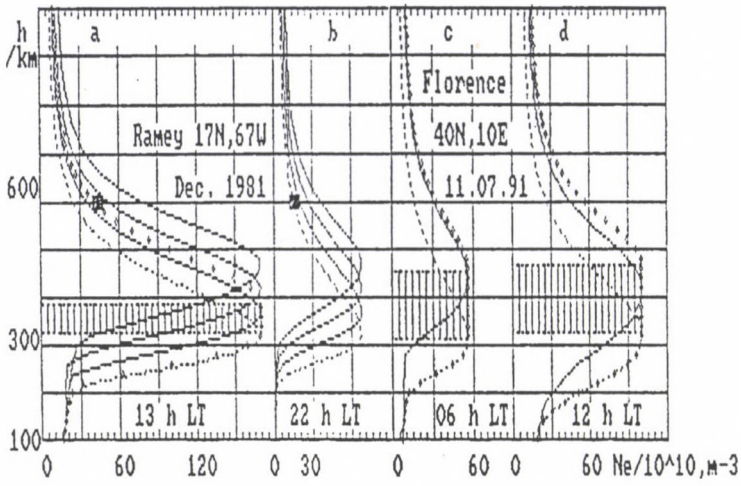


Fig. 4. Set of electron density profiles used for TEC integrating. IRI predictions – dashed line. Squares – HINOTORY Ne measurements. Dashed segment – interlayer peak area. Circles – final profile

thickness of the peak density "ionospheric cloud" at the altitudes: a) 330 to 390 km, c) 310 to 455 km, d) 325 to 465 km.

Three examples of the broad peak structure inserted between the bottomside and topside ionosphere are presented in Fig. 5 for Goose bay, Florence and Lunping at the times of the ionospheric-geomagnetic storm on 25–28 July 1981 (see Fig. 1). Here the initial model electron density profiles have been first updated with the relevant ionosonde observations at Goose bay, Gibilmanna and Chung-Li (solid curve). However to reach consistency of observed TEC_o and model TEC_m after the profile updating requires broad peak (dashed area) specified by Eq. (2).

Figure 6 shows diurnal variation of monthly median Ne(h) profiles fitting TEC median values measured at Lunping for December, 1996, based on IRI predictions with broad peak of ionization (shown between two solid lines) defined by Eq. (2). We obtain peak thickness h_s varying from zero (no additional REC) to 300 km between the bottomside h_mF2 and the topside ionosphere.

3. Conclusion

The above calculations are applied in the present paper to cases of daily-hourly TEC and monthly median TEC enhancements which exceed model TEC values obtained by integration along the IRI and PRIME electron density profiles. Clear difference between the disturbances manifested by TEC and ionosonde peak density N_mF2 is obtained in this study. The TEC and N_mF2 deviations from their normal (mean or median) values does not coincide in about 50% of disturbed times. Enhanced N_mF2 (or f_0F2) means more dense electron population but does not specify any variation of the F2 region peak height. However enhanced TEC often requires presence of the thick ionization cloud occurring with the observed peak density. In

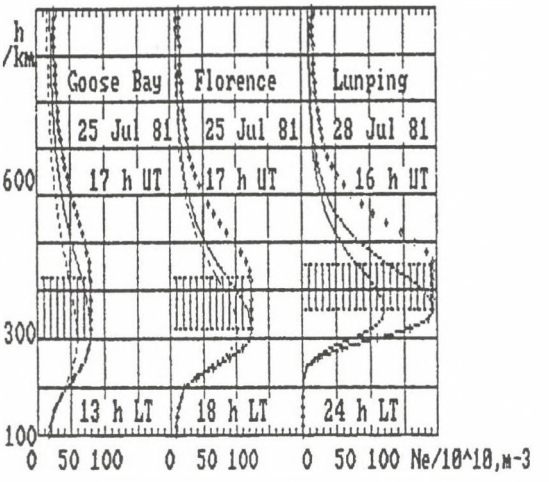


Fig. 5. Set of electron density profiles used for TEC integrating at Goose bay, Florence and Lunping for hours of extremely enhanced TEC during the geomagnetic storm at the end of July 1981. Dashed curve – IRI predictions. Solid curve – results of IRI updating with ionosonde critical frequency observations. Circles – final profile including extra layer between the bottom and top profiles

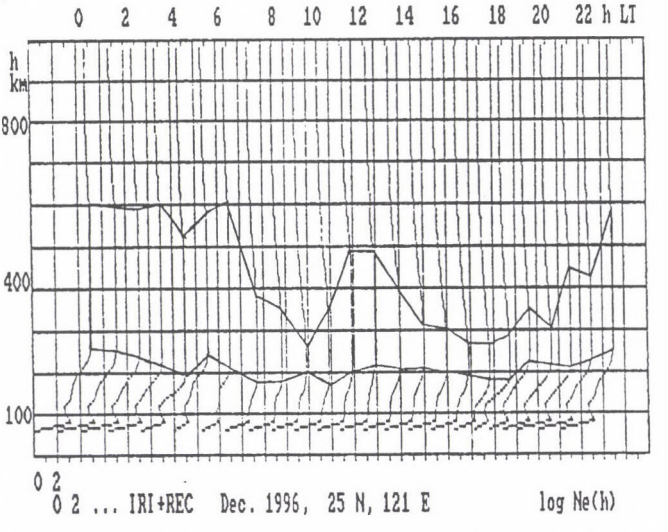


Fig. 6. Diurnal variation of IRI monthly median Ne(h) profiles fitting TEC measurements at Lunping for December 1996, with thick additional peak layer between the bottomside and topside ionosphere

some cases both effects could occur simultaneously. Detection of relevant occasions in the ionosphere relies on day-to-day and hour-to-hour analysis for the ionospheric quiet or disturbed state using TEC and ionosonde observations.

Possibility of presence of an additional peak density layer in ionosphere would considerably change the shape of total electron density profile. Strategy or producing TEC_m by integrating model electron density profile fitted to current ionosonde peak parameters should be significantly changed to include profile peak thickness updating with current TEC observations.

Acknowledgements

The author is grateful to J Klobuchar, P Spalla, K I Oyama, R O Conkright, Y N Huang and K Cheng for cooperation providing TEC, HINOTORY, and ionosonde data used for this study, and to Mrs A S Besprozvannaja and G F Deminova for helpful discussion. The financial support provided by Prof. R Leitinger and Dr. P Bencze for presenting this paper is gratefully acknowledged.

References

Bilitza D 1990: International Reference Ionosphere, IRI-90. Rep. NSSDC/WDC-A R and S, 90–22, Greenbelt, USA

Bradley P A 1995: PRIME (Prediction and Retrospective Ionospheric Modelling over Europe). COST238, Final Report, ECSC-EEC-EAEC, Brussels

Das Gupta A., Basu S, Bhar J N, Bhattacharya J C 1975: *J. Geoph. Res.*, 80, 699.

Davies K, Liu X M 1991: *Radio Sci.*, 26, 997.

Decker DT, Anderson D N, Preble A J 1997: *Radio Sci.*, 32, 2003.

Huang Chun-Ming 1974: *Radio Sci.*, 9, 519.

Gulyaeva T., Spalla P 1995: *Annali di Geophysica*, 38, 429.

Jayachandran B, Balachandran Nair R., Balan N, Rao P B 1995: *J. Atm. Terr. Phys.*, 57, 1599.

Kishcha P V 1995: *Adv. Space Res.*, 15, 55.

Klobuchar J A, Leitinger R 1993: In: Proc. of COST238/PRIME Workshop, Part 1, Karl-Franzens University, Graz, 13.

Su Y Z., Bailey G J., Balan N 1995: *J. Atm. Terr. Phys.*, 54, 433.

Watanabe Sh., Takahashi T., Oya H., Oyama K I 1995: Table of electron density at 600 km altitude during 1981–1982 in the equatorial region. ISAS RN 579, Yoshinodai, Sagamihara, Tokyo

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 365–377 (1997)

TAILORING OF EOFs FOR IONOSPHERIC TOMOGRAPHY AND THE WEIGHTED, PARTITIONED, LEAST-SQUARES ALGORITHM

E J FREMOUW[1], J A SECAN[1], CHUCAI ZHOU[2]

Among the linear-algebraic class of inversion techniques used for ionospheric tomography is the Weighted, Damped, Least-Squares (WDLS) algorithm. One application of it uses empirical orthogonal functions (EOFs) as a vertical basis set and Fourier harmonics horizontally. To date, the EOFs employed have been generated from a large database of electron-density profiles with the intent of spanning all ionospheric states. In this work, we explore the utility of tailoring the EOFs for specific ionospheric conditions by limiting their span in latitude, longitude, local time, season, solar epoch, and geomagnetic disturbance level. In this paper, we demonstrate early results, showing some improvement by tailoring in latitude. We also describe a successor to the WDLS method, the Weighted, Partitioned, Least-Squares (WPLS) algorithm, which combines weighting with the partitioning capability of singular-value decomposition (SVD). In particular, we introduce a simple criterion for truncating the SVD series of singular values.

Keywords: empirical orthogonal functions; ionospheric tomography; total electron content, TEC

1. Introduction

One of the important classes of inversion algorithms for performing ionospheric tomography is that based on linear-algebraic matrix inversion. Previously, we have reported on the Weighted, *Damped*, Least-Squares (WDLS) algorithm, using a basis set consisting of Fourier terms horizontally and Empirical Orthogonal Functions (EOFs) vertically (Fremouw et al. 1994). The EOFs were generic, attempting to span all ionospheric states. Here, we describe initial development of a successor — the Weighted, *Partitioned*, Least-Squares (WPLS) algorithm — and EOFs tailored to more specific observing conditions.

We present the WPLS algorithm in Section 2, pointing out its advantage over its WDLS predecessor. In Section 3, we describe our approach to tailoring of EOFs and present examples. Our conclusions are stated in Section 4.

2. WPLS method

The algorithm

As with all algebraic inversion algorithms, we envision a "forward" problem as the product of a geometry matix, **G'**, and a vector, **m'**, made up of parameters and a basis set that characterize — or "model" — the ionosphere. As indicated in Eq.

[1]NorthWest Research Associates, Inc. Bellevue, WA, USA
[2]University of Washington, Seattle, WA, USA

1217-8977/97/$ 5.00

(1), that product equals a vector, **d'**, constructed from the TEC values measured along many raypaths.

$$\mathbf{G'm'} = \mathbf{d'} . \tag{1}$$

In the WDLS algorithm, we employed a generalized inverse that permits weighting the basis functions according to *a priori* information about their statistical variances and the data inversely according to their relative uncertainty. In the WPLS algorithm, we absorb those weights, which respectively are the standard deviations σ_a and σ_d, into the forward geometry matrix. That weighted matrix then undergoes singular-value decomposition (SVD), as follows:

$$\mathbf{G'} = \sigma_d^{-1}\mathbf{G}\sigma_a = \mathbf{U\Lambda V}^T . \tag{2}$$

The advantage is that, using SVD, we now can truncate the series of singular values to partition the problem into its over-determined and under-determined parts. The former then is solved by means of the ordinary SVD inverse,

$$\mathbf{G_S'^{-g}} = \mathbf{V_S\Lambda_S^{-1}U_S^T} , \tag{3}$$

yielding those "model" parameters spanned by the data, and the latter is left as a minimum-length solution, m_0, relative to the *a priori* background.

That is, we obtain the following weighted model parameters:

$$\mathbf{m'} = \mathbf{G_S'^{-g}d'} + \mathbf{m_0} = \sigma_a^{-1}\mathbf{m} \tag{4}$$

from the weighted version,

$$\mathbf{d'} = \sigma_d^{-1}\mathbf{d} \tag{5}$$

of the original data vector, **d**. The solution is obtained finally by "de-weighting" the weighted estimates of the model parameters, as follows:

$$\mathbf{m} = \sigma_a\mathbf{m'} . \tag{6}$$

Truncation

We are dealing with ill-conditioned matrices, and the singular values typically do not show a clear point of dropoff to use for truncating the series. In searching for a simple truncation criterion, we have been led through a rather lengthy matrix-algebra development (to be reported separately) to the following result:

$$\sigma_p^2 = \sigma_a\mathbf{V}\begin{bmatrix} \mathbf{\Lambda}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\mathbf{V}^T\sigma_a . \tag{7}$$

Equation (7) relates the post-inversion model variances on the left, to their *a priori* values (the squares of the σ_a), in terms of the SVD solution. The least-squares part of the weighted solution is contained in the upper left-hand portion of this matrix, quantified by the inverse singular values. Singular values greater than unity decrease the variances relative to their *a priori* values; smaller ones do not. Thus

we truncate the series of weighted singular values, Λ, at unity. The remaining part of the solution is the minimum-length one, and it resides in the identity matrix in the lower right- hand corner of Eq. (7).

Since $\mathbf{V}$ times its transpose is an identity matrix, we see that Eq. (7) would reduce to the expected statement if all singular values were excluded. That is, the entire problem then would reduce to the minimum-length one, and the post-inversion variance would equal the *a priori* model variance. Weighted singular values greater than unity, which we keep in the least-squares solution, resolve some of the *a priori* variance.

3. Tailored EOFs

Subjective comparisons

Seeking to learn what, if any, advantage might be gained by using EOFs tailored to the conditions of a given observation, we have generated several such sets and employed them with the WDLS algorithm. Specifically, we have used limited ranges of geomagnetic latitude and local time, season, sunspot number, and planetary magnetic activity (parameterized by means of K_p).

Before generating the tailored EOFs, we produced a generic set from the Parameterized Ionospheric Model (PIM) developed by Daniell et al. (1995). To do so, we performed 108 PIM runs, spanning an equinox and both solstices; four GMTs; low, medium, and high sunspot number; and low, medium, and high K_p. Each run produced a file containing 2,136 electron-density profiles on a 2°-by-1° grid of geomagnetic latitude and longitude (or local time). Thus, a total of 230,688 profiles were calculated. From this total came a mean background profile and a set of eight generic EOFs — that is a vertical basis set spanning all observing conditions and the whole range of PIM-based ionospheric states.

The generic EOFs are illustrated in Fig. 1. The upper left plot contains the average plasma-density profile, dominated by a very smooth F layer and containing just a hint of an E layer. Progressing to the right and downward, EOFs #01 through #08 characterize increasingly fine structure, especially in the bottomside. A tomographic image produced on the basis of these generic EOFs will be referred to as Case 'All' in comparisons contained in the remainder of this section.

Figure 2 contrasts Case 'All' with the best one could do with eight EOFs and 60 horizontal harmonics. The latter, termed Case 'Best', employed EOFs generated, not from PIM runs, but from the "Initial Image" illustrated at the top of the figure. We produced this field of electron densities some time ago (Fremouw et al. 1992) by hand digitizing a display of incoherent-scatter returns from Chatanika, Alaska, published by Muldrew and Vickrey (1982). It contains a mix of auroral and polar structures in the F layer, and we've left out the E layer for these tests because of limited resolution in our hand-digitizing procedure. The unnaturally sharp contour features stem from that crude procedure. We integrated through the Initial Image to obtain simulated TEC data for testing a variety of EOFs (including Case 'Best' and all others illustrated in this paper).

The main difference between Case 'Best' and the Initial Image is a generally

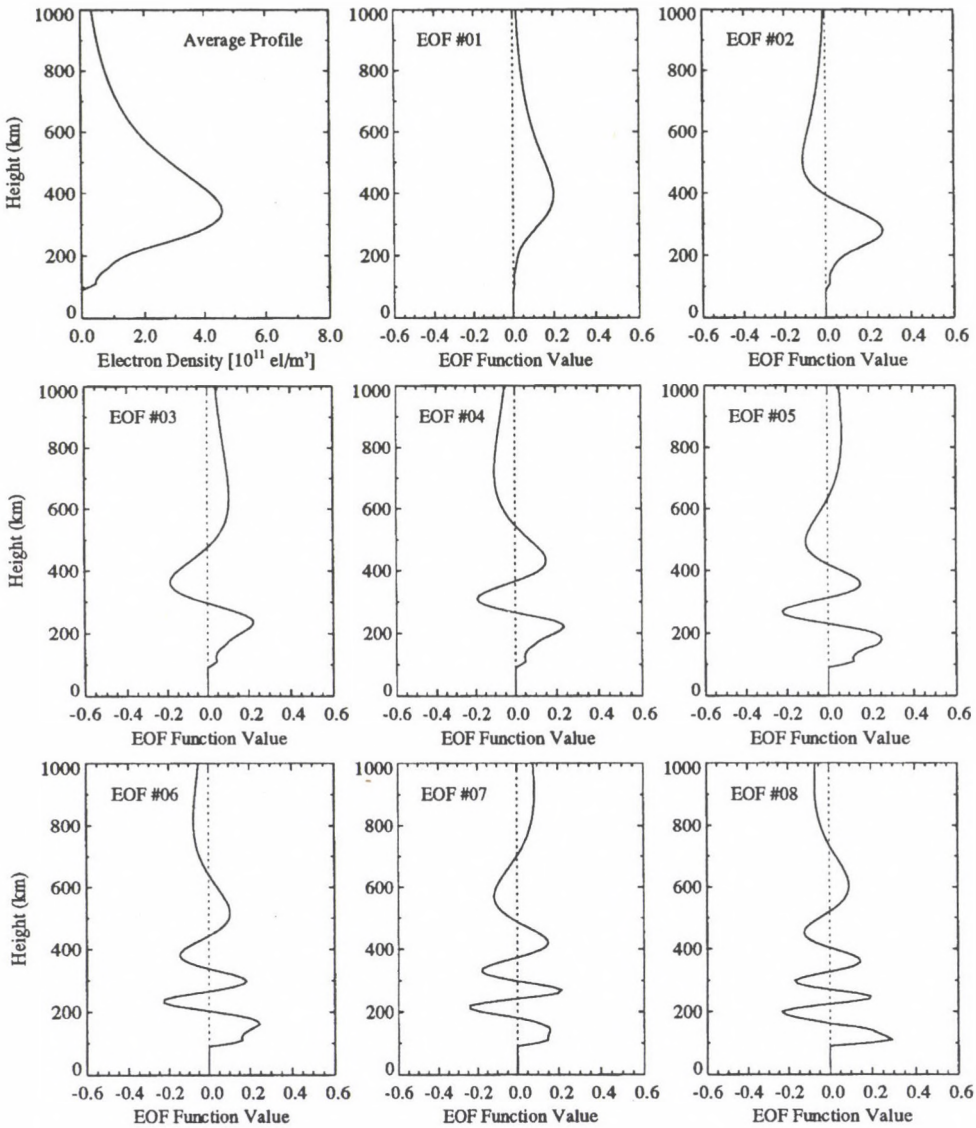


Fig. 1. Average profile and generic set of EOFs, generated from full span of PIM profiles. The *a priori* variance associated with the EOFs decreases with increasing mode number

smoother appearance in the former, resulting from limited resolution and facilitated by the k^{-2} *a priori* weighting that we employ for the horizontal spectrum. The bottom panel shows the image produced with the generic set of EOFs spanning all observing conditions, including middle and equatorial latitudes. Thus, it represents the result of the least-tailored EOF set employed in our tests.

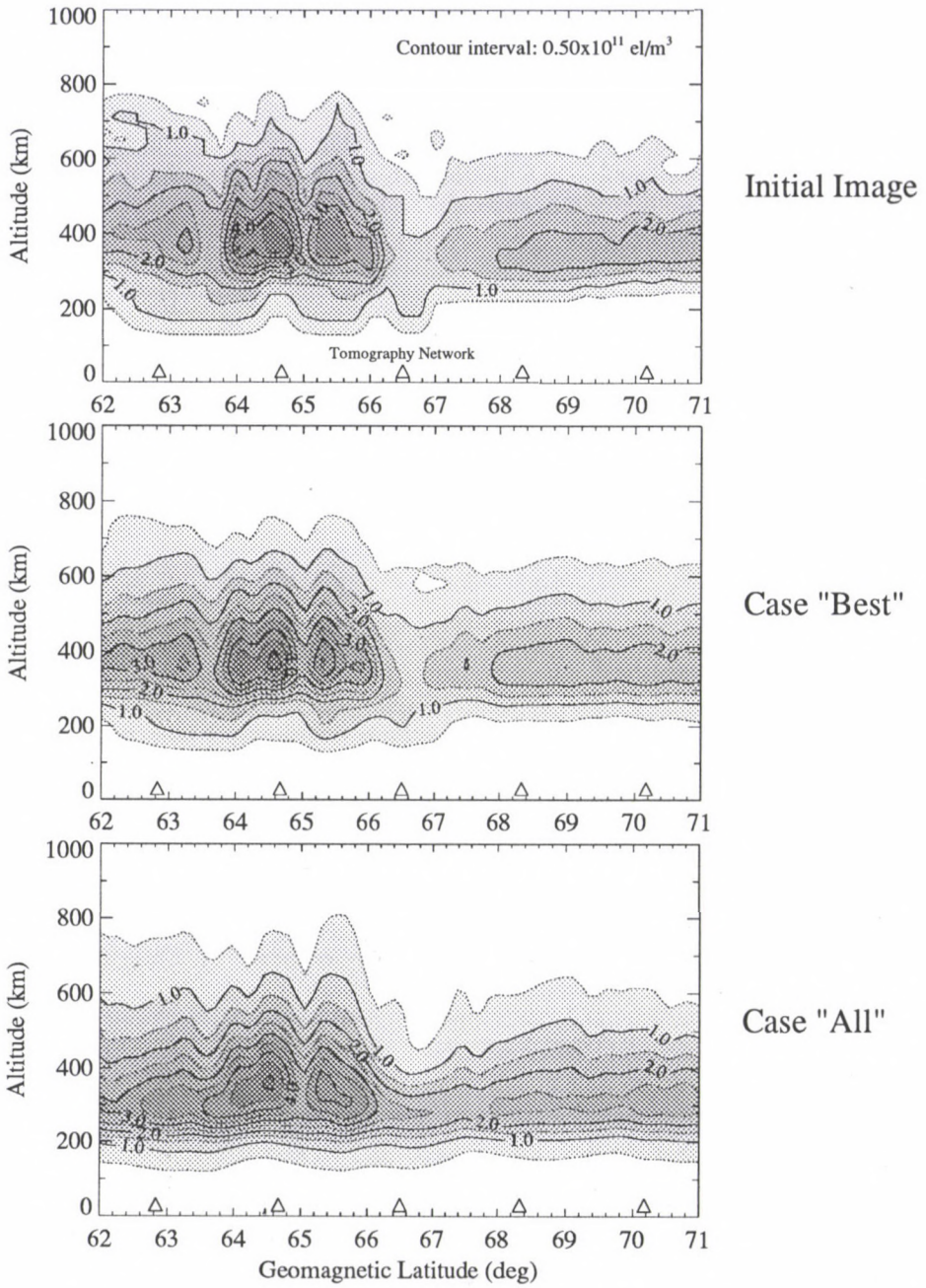


Fig. 2. Contrasting the tomographic result obtained using generic EOFs (Case 'All') with the best one could do (Case 'Best') employing eight EOFs and 60 horizontal harmonics. EOFs for Case 'Best' generated directly from "Initial Image" (top panel), which was integrated to obtain simulated TEC records for all reconstructions contained in this paper

Figure 3 illustrates an advantage gained from limiting the latitude span of the EOFs. The top panel again contains Case 'Best', and the bottom one illustrates the least-tailored result. The middle panel presents an image generated using the 'Tailored' EOFs that span PIM profiles limited to geomagnetic latitudes greater than 55° — that is, sub-auroral, auroral, and polar locations. Subjectively, the rendering is improved compared with that obtained with the generic EOFs, an improvement which we'll quantify shortly.

In Fig. 4, we find a caution against excessive, or ill-chosen, tailoring. At the top, we have again the 'Best' image that can be produced with eight EOF's, and the set 'Tailored' to latitudes above 55 is repeated in the middle panel. At the bottom is an image produced from the same (high-latitude) TEC data employing EOFs generated from the "Wrong" range of latitudes, namely equatorial ones between ±25° geomagnetic. It is decidedly worse than that obtained from proper latitude "tailoring" and somewhat worse than that obtained with the generic EOFs (Case 'All' in Figs 2 and 3).

Quantitative comparisons

To provide a more objective basis for assessing the efficacy of various reconstructions, Fig. 5 compares the horizontally averaged vertical profile in Case 'All' (thick-line solid curve) with that from the Original Image (thin-line solid curve). It also shows the horizontally averaged difference (dotted curve) between vertical profiles in the two electron-density fields and the horizontal root- mean-square (RMS) difference (dashed curve) between them, the latter ranging from less than $1 \cdot 10^{10}$ to a little over $4 \cdot 10^{10}$ el/m^3. Departures of Case 'All' from the Original Image include a peak density that is somewhat too large and a peak height that is a little too low. In addition, it appears that the integral through the average reconstructed profile (i.e., the vertical TEC) is too large, owing at least in part to a steeper topside decay in the original than that achieved in the average reconstruction.

We'll use the foregoing three characteristics — i.e., peak density translated to f_0F2; peak height (h_mF2); and vertical TEC — as measures of performance in the rest of this section. Figure 6 provides comparisons of these three vertical-profile measures between Case 'All' and the Original Image, as functions of latitude across the two-dimensionally mapped region. At the top is a comparison of the reconstructed f_0F2 across the image with that of the true f_0F2. For the most part, the reconstruction is too large by 1/2 to 1 MHz or so, and its peak height is too low. As expected from inspection of the average profiles, vertical TEC in the reconstruction is too large by about a TEC unit (10^{16} el/m^2).

Figure 7 presents horizontally the vertical-profile measures for Case 'Best' (employing EOFs generated from the original data themselves) compared with those for the Original Image. It shows that eight EOFs can reproduce our three measures of performance rather well, in principle.

In practice, of course, one never could achieve Case 'Best' since estimating the original electron-density field, which is unknown, is the purpose of ionospheric tomography. Figure 8 shows, however, that one can improve upon Case 'All' by

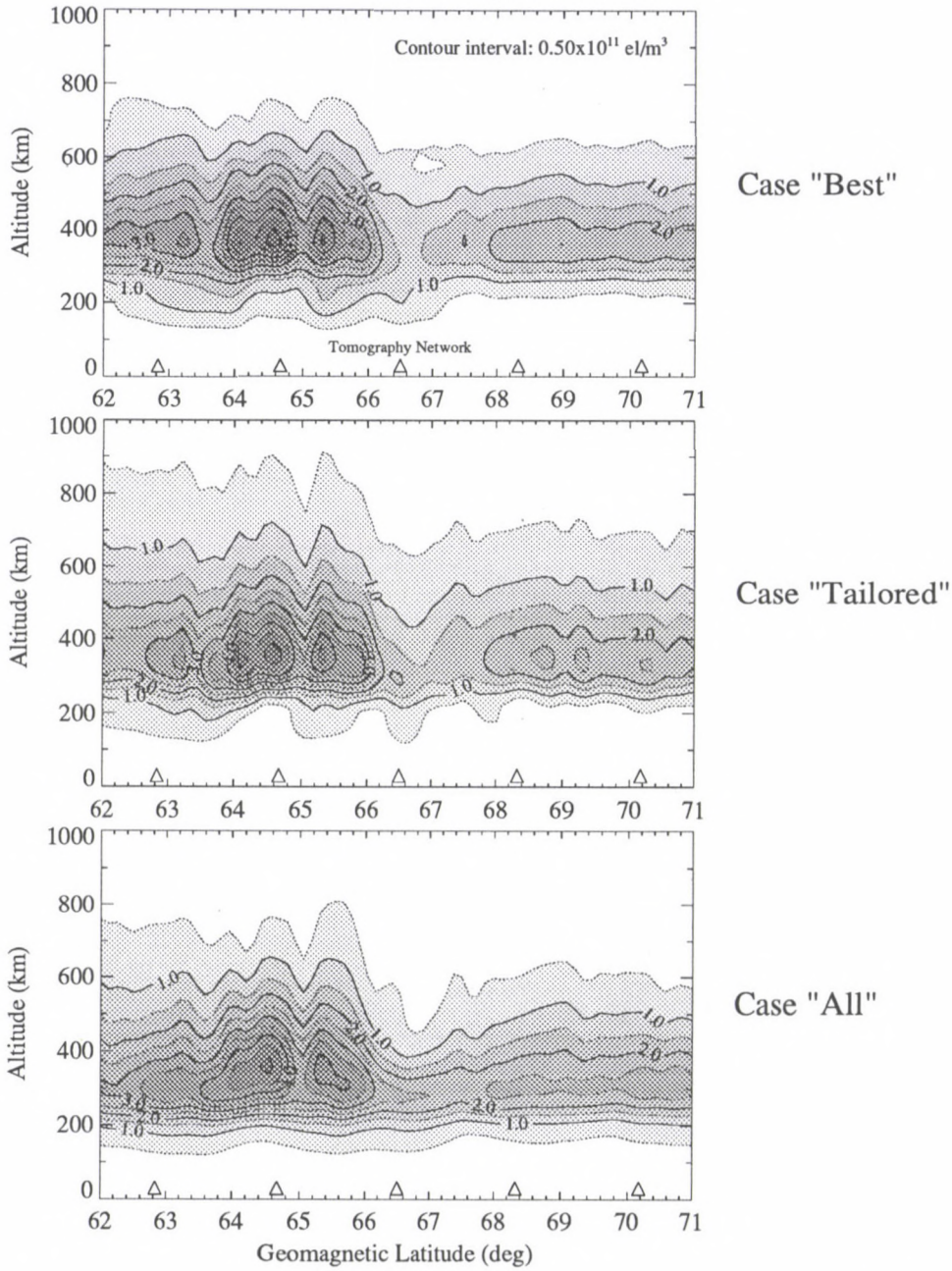


Fig. 3. Illustrating improvement (middle panel) gained by generating EOFs from PIM runs limited to latitudes greater than 55°. Case 'Tailored' resembles Case 'Best' more closely than does Case 'All'

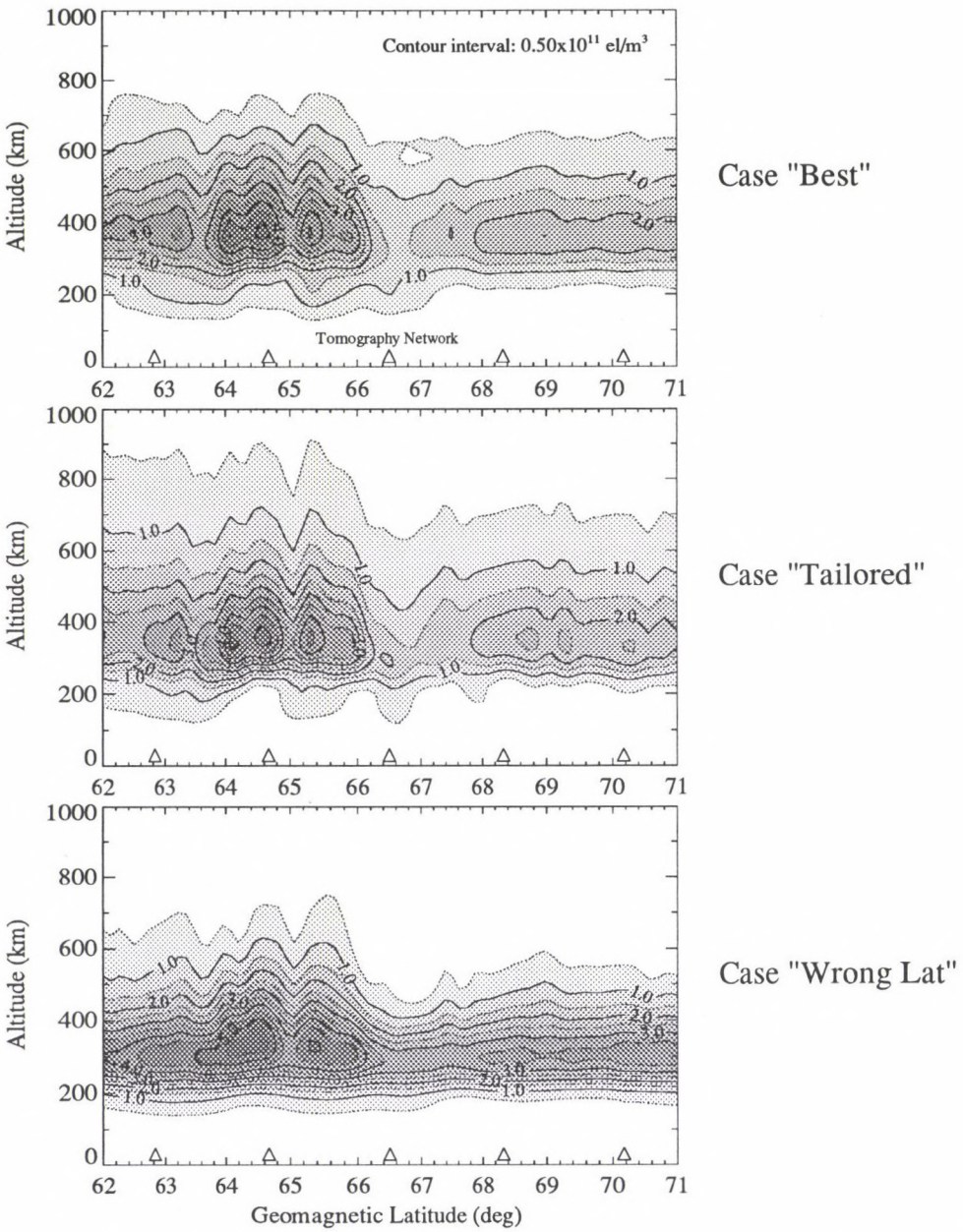


Fig. 4. Illustrating deterioration (bottom panel) experienced when tailoring incorrectly. Case 'Wrong Lat' is tomographic reconstruction of high-latitude TEC data using EOFs generated from equatorial runs of PIM

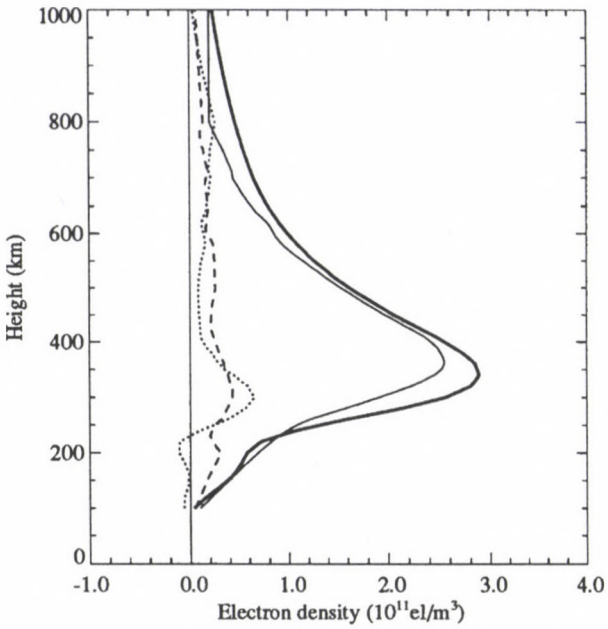


Fig. 5. Quantitative comparison of vertical profiles in Case 'All' with those in Original Image. Solid curves: horizontally averaged vertical profile (thick line for Case 'All'; thin line for Original Image). Dotted curve: horizontally averaged difference. Dashed curve: horizontal RMS difference

employing latitudinally tailored EOFs. That is, Case 'Tailored' reproduces f_0F2 within less than 1/2 MHz, with few exceptions, across the image. Reproduction of h_mF2 also has improved, especially in the polar F layer, although it is still somewhat too low in the auroral region.

Surprisingly, vertical TEC has not improved much in Case 'Tailored' compared with that in the generic Case 'All'. We are employing relative TEC data, so we were concerned about estimating the ambiguity properly. A different test case, however, produced very close agreement in vertical TEC. We believe that the TEC disagreement in the test case employed in this paper is due to the unusually sharp dropoff in topside density in our original image, which may actually result from lack of topside sensitivity by the Chantanika incoherent-scatter radar. If so, one would not expect our PIM-based EOFs to span such a behaviour.

4. Conclusion

We conclude that limited tailoring in latitude is useful, but the EOFs must span all regimes encountered at the receiving sites (e.g., not only auroral in Alaska). We've also investigated tailoring the case illustrated in this paper by limiting the EOF span in local time, season, sunspot number, and K_p. Some improvement was achieved by employing only high sunspot-number runs of PIM, the incoherent-scatter data having been obtained near solar maximum. The obser-

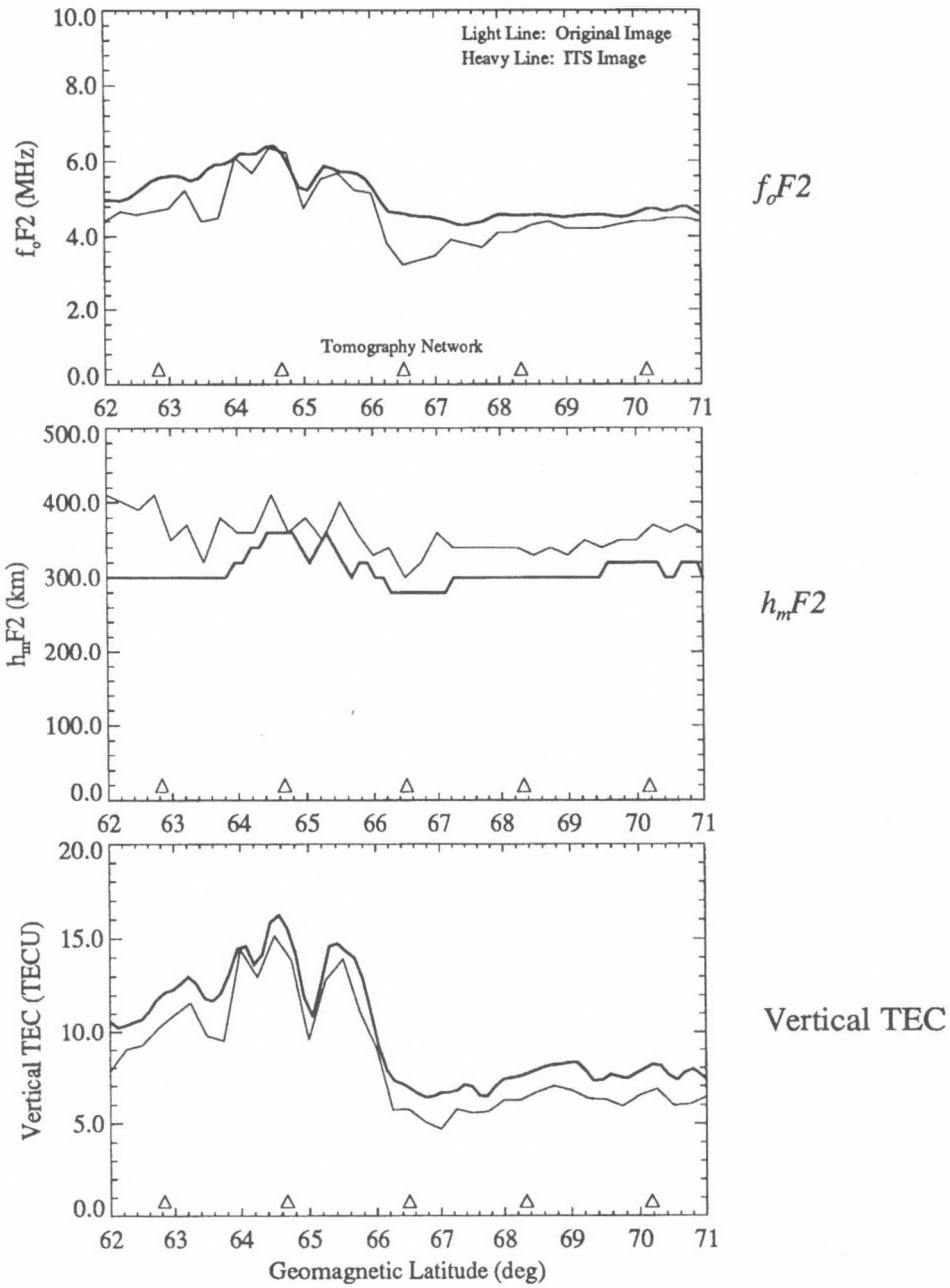


Fig. 6. Horizontal comparison of vertical-profile characteristics between Case 'All' and Original Image. Thick lines: Case 'All'. Thin lines: Original Image

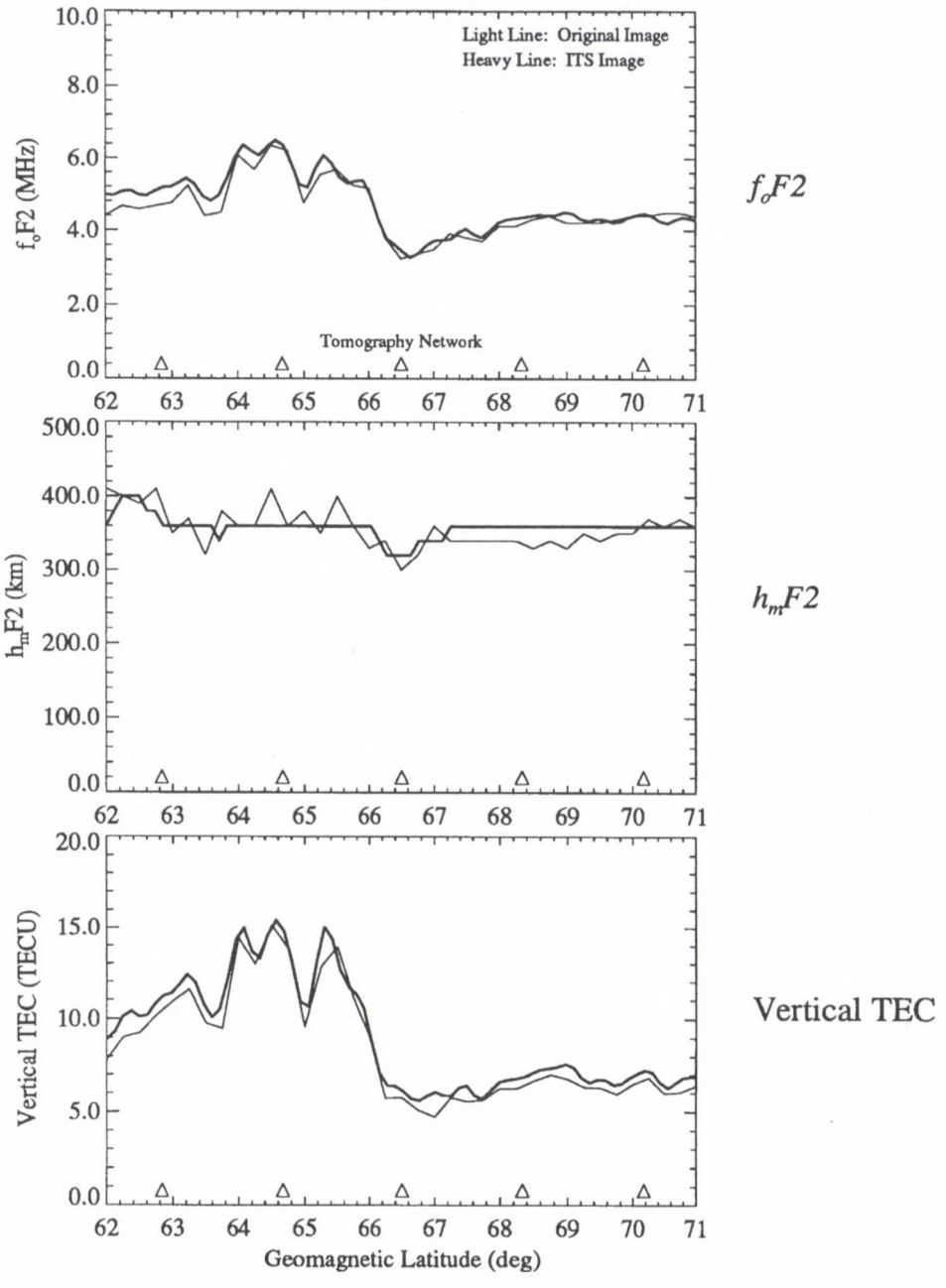


Fig. 7. Horizontal comparison of vertical-profile characteristics between Case 'Best' (which cannot be achieved in practice) and Original Image. Thick lines: Case 'Best'. Thin lines: Original Image

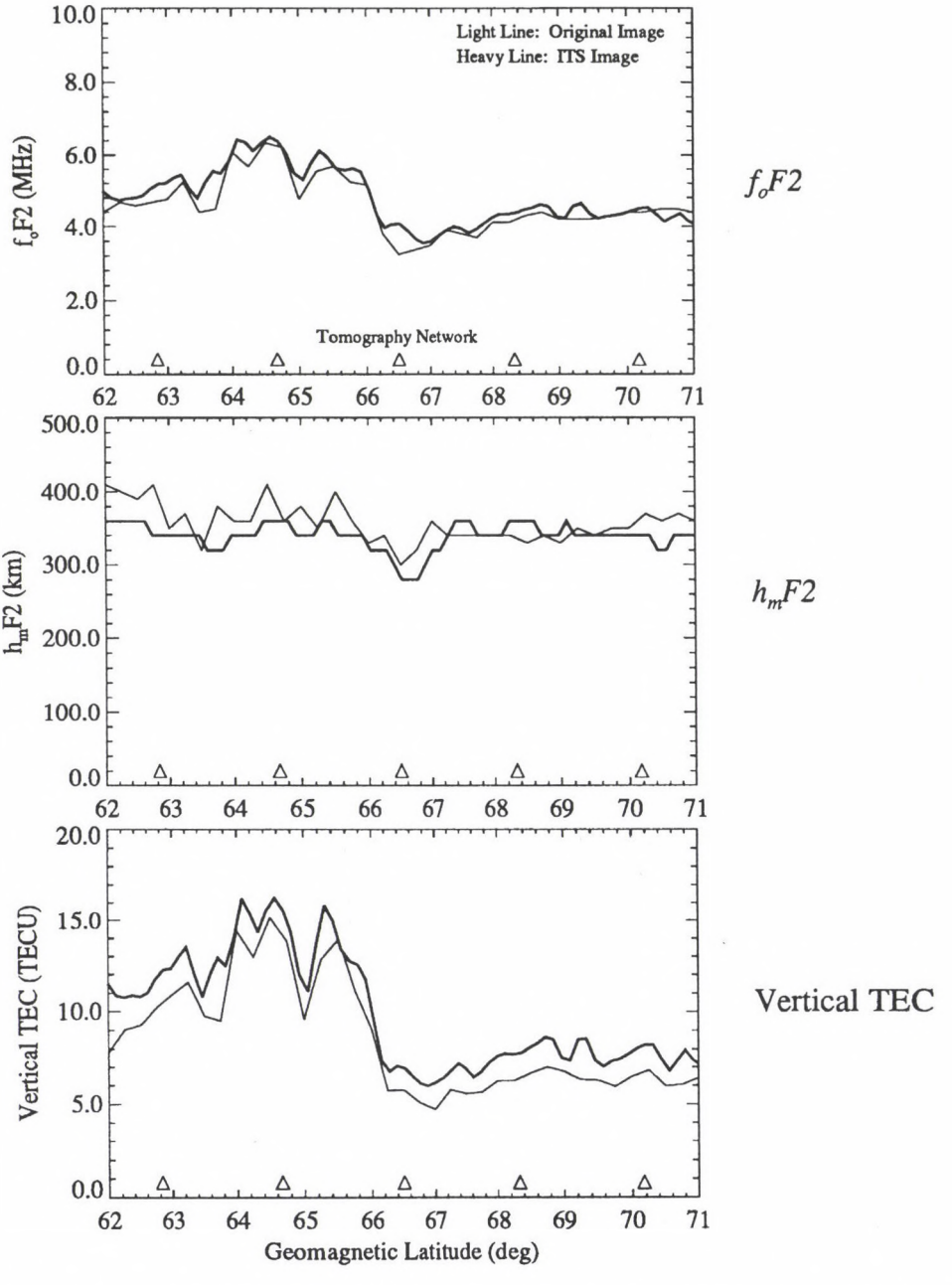


Fig. 8. Horizontal comparison of vertical-profile characteristics between Case 'Tailored' and Original Image. Thick lines: Case 'Tailored'. Thin lines: Original Image. Note improvement in f_0F2 and h_mF2 compared with Case 'All' (Fig. 6)

vation was performed at a moderate K_p level, but tailoring EOFs accordingly did not provide significant improvement. Limiting in season actually made the reproduction a little worse, although this may be because we omitted the E layer in our input test data.

We plan additional tests of EOF tailoring with an expanded database of incoherent-scatter measurements. Thereafter, we intend to employ tailored sets with the WPLS algorithm.

Acknowledgment

The research presented in this paper was supported by the U.S. National Science Foundation under Grant ATM-9528146.

References

Daniel R E, Brown L D, Anderson D N, Fox M W, Doherty P H, Decker D T, Sojka J J, Schunk R W 1995: *Rad. Sci.*, 30, 1499–1510.

Fremouw E J, Secan J A, Howe B M 1992: *Rad. Sci.*, 27, 721–732.

Fremouw E J, Secan J A, Bussey R M, Howe B M 1994: *Int. J. of Imaging Systems and Technology*, 5, 97–105.

Muldrew D B, Vickrey J F 1982: *J. Geophys. Res.*, 87, 8263–9272.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 379–394 (1997)

INVERSION OF THE PLASMA SIGNAL IN GNSS OCCULTATIONS — SIMULATION STUDIES AND SAMPLE RESULTS

R LEITINGER[1] and G KIRCHENGAST[1]

In the near future the reception of GNSS (Global Navigation Satellite Systems, presently GPS and GLONASS) signals onboard low orbiting satellites will play an important role. If GNSS occultation can be observed, one can "measure" satellite to satellite electron content (hTEC) by means of "Differential Doppler" (carrier phase difference). In case of occultations with suitable ray geometry, inversion of the hTEC data gives horizontally averaged height profiles of electron density.

We demonstrate by means of model calculations that at mid latitudes the inversion results are representative electron density profiles with high accuracy of F layer peak height.

One receiver — GPS/Met onboard the small satellite MicroLab 1 — has been launched in April of 1995 and provided the data for sample results demonstrating the practical applicability of the inversion technique.

Keywords: occultation; inversion; simulation; total electron content, TEC

1. Introduction

Global Navigation Satellite System (GNSS) receivers on board of Low Earth Orbit (LEO) satellites allow to gain "horizontal" electron content (hTEC) measured along the ray path from the transmitter to the receiver.

Classical inversion of hTEC gives height profiles of electron density approximately averaged over the horizontal ray sections between the intersection points of the ray with the LEO orbit height shell. Two dimensional ionospheric imaging and (in favorable cases) true tomographic reconstruction is possible when the GNSS occultation observations are combined with ground observations of ionospheric electron content.

A GNSS receiver on board a LEO satellite with an orbit inclination between about 50° and 140° provides around 1100 useful occultations per day if the operational system of 48 GNSS satellites (24 GPS plus 24 GLONASS) is considered. Restriction to setting occultations and to GPS still leaves near 300 events per day. Of course not all are suitable for inversion. A good criterion can be derived from the ground projection of the point on the occultation ray which is tangent to the Earth's surface ("tangent point" or "ray perigee") or from the duration of the occultation event (Table I). We define the latter to be the transit time from the highest possible perigee (= height of the LEO satellite) to the ground-grazing occultation (setting occultation) or from ground-grazing occultation to the highest possible perigee (rising occultation). Typical "good" occultation event durations range from about 6 to 9 minutes for a LEO height of 800 km.

[1]Institut für Meteorologie und Geophysik, Universität Graz, Halbärthgasse 1, A–8010 Graz, Austria, e–mail: leitinger@bkfug.kfunigraz.ac.at

1217-8977/97/$ 5.00

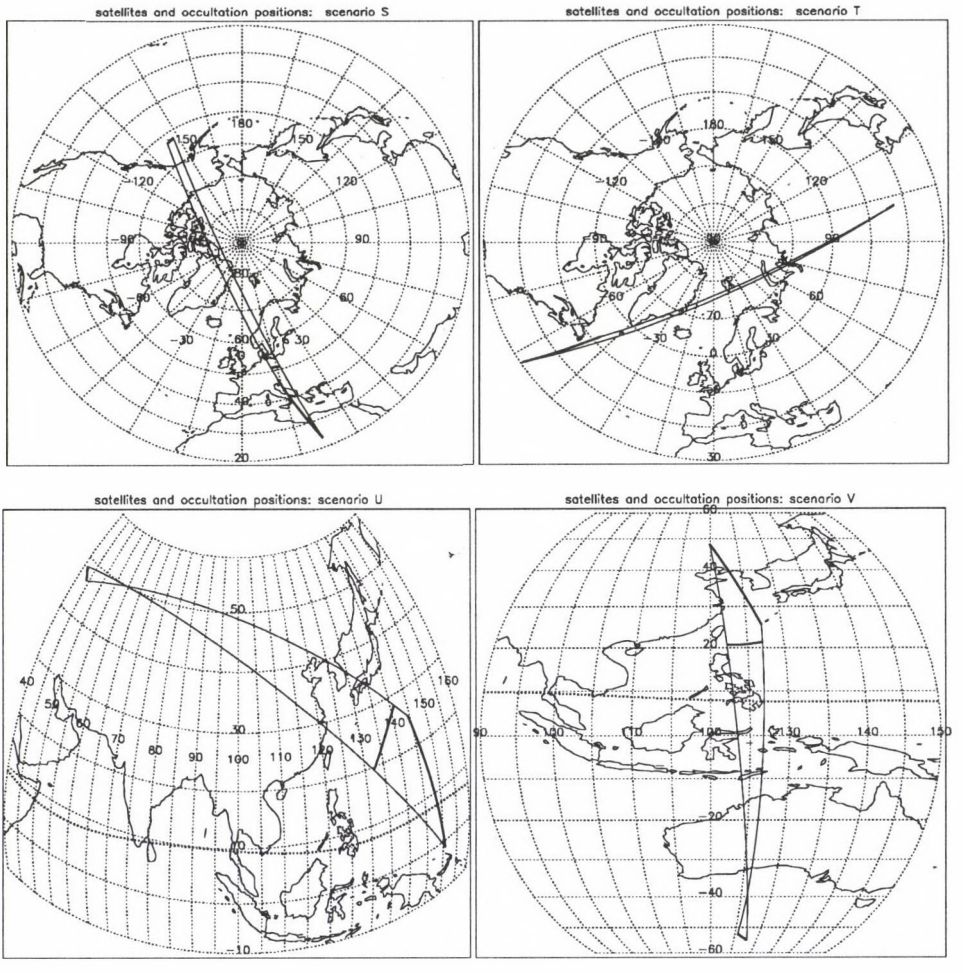


Fig. 1. Maps with ray and position projections for the scenarios S (top left), T (top right), U (bottom left), V (bottom right). Lambert's polar projections at the top, orthographic projections at the bottom. Thin lines: ground grazing occultation ray and ray with $h_t = 800$ km; heavy lines: positions of the LEO satellite (long trace at end of rays), of the GNSS satellite (short trace at end of rays), and of the ray perigees (near the LEO traces). The heavy dotted line marks the CCIR dip equator (scenario U and V maps only)

The length of the ground projection of the tangent point trajectory during an occultation event ("smearing length") can vary from a few km to thousands of km. (Theoretically the lower limit is 0 but in practice this case is not realistic.) For inversion purposes a maximum smearing length of 550 km (or 5 degrees) is certainly acceptable. If a ray with a perigee near the center of the ionosphere (e.g., at 400 km height) is chosen to define the inversion plane (which contains this ray and is perpendicular to the surface of the Earth), a smearing length of 5 degrees

Table I. Radii and heights

Description	Radius	Height	Typical values for radius	for height
Earth radius	$R = R_o$		(6367 ± 11) km	0 km
GNSS orbit	r_G	h_G		20000 km
LEO orbit	r_L	h_L		800 km
ray perigee	r_t	h_t		(0···800) km
upper limit of "ionosphere" (footpoint of plasmasphere)	r_u	h_u		2000 km

(R_o: distance of the occultation point from the center of the Earth; $h = r - R$)

Table II. Ray path parts and electron content quantities ($I = \int N_e ds$)

Description	Acronym	Symbol	From – to
"horizontal"	hTEC	I_h	LEO – GNSS (L – G)
"occultation"	oTEC	I_o	LEO – point at LEO height (L – P)
"upper ionosphere"	iTEC	I_i	point at LEO height-limit of ionosphere (P – U)
"plasmasphere"	pTEC	I_p	limit of ionosphere – GNSS (U – G)

hTEC consists of three parts: the "occultation" part from the LEO position (L) via the tangent point to the ray intersection with the LEO height at the "GNSS side" (P), the "upper ionosphere" part to the foot height of the plasmasphere (U), the "plasmasphere" part to the GNSS location (G) $[I_h = I_o + I_i + I_p]$

gives an extension of the inversion layers perpendicular to this plane of ≤ 7 degrees (including ray divergence). With LEO inclinations between 50° and 140° about a quarter of the GNSS occultations have smearing lengths $\leq 5°$.

We can safely assume that all rays can be considered as straight lines from a transmitter to a receiver. Since transmitter frequencies are well above 100 MHz the refraction effect of ray bending is negligible.

The GPS receiver GPS/MET onboard the small research satellite MicroLab 1 (nearly circular orbit in a height around 750 km, inclination: about 70°, launch date: 3 April 1995) produced occultation data useful for inversion. A relevant example is found by Leitinger et al. (1997) (occultation 95/195/0063 in the list of the GPS/MET data center at UCAR, Boulder, 22 October, 1995, 7:34:51 UT, mid-latitudes).

Here we deal with simulations only. In general, simulations are the only way to assess the relevance and the usefulness of inverted remote sensing data on a larger scale. In the case of the inversion of GNSS occultation (hTEC) data, full comparison of electron density height profiles for validation purposes is practically restricted to those from Incoherent Scatter observations as the only other full profile method available. This means restriction to a few locations and restriction to scarce spatial and temporal co-location.

Table III. Properties of occultation scenarios

Scenario S									
	LEO			GNSS			P		
r_L	ϕ_L	λ_L	r_G	ϕ_G	λ_G	r_P	ϕ_P	λ_P	Δt
7195.0	23.442	21.376	26582.6	52.174	211.958	6371.208	50.866	16.479	1.10
7192.1	46.753	14.144	26580.5	52.849	215.693	7171.202	50.995	12.577	0.83
Scenario T									
	LEO			GNSS			P		
r_L	ϕ_L	λ_L	r_G	ϕ_G	λ_G	r_P	ϕ_P	λ_P	Δt
7193.2	30.495	303.010	26865.5	42.908	101.963	6363.192	56.685	316.323	-2.91
7190.9	51.918	310.644	26850.5	45.215	100.944	7163.243	56.580	313.883	-3.07
Scenario U									
	LEO			GNSS			P		
r_L	ϕ_L	λ_L	r_G	ϕ_G	λ_G	r_P	ϕ_P	λ_P	Δt
7190.6	-1.351	150.726	26630.3	48.485	41.318	6372.119	19.489	132.284	8.82
7189.6	25.437	144.436	26636.8	46.418	44.131	7171.204	28.356	141.216	9.41
Scenario V									
	LEO			GNSS			P		
r_L	ϕ_L	λ_L	r_G	ϕ_G	λ_G	r_P	ϕ_P	λ_P	Δt
7191.7	48.160	119.884	25459.0	-54.679	127.861	6375.582	20.663	122.218	8.15
7194.4	25.517	127.162	25455.4	-52.274	125.536	7175.304	21.341	127.083	8.47

r: distance from center of Earth (km); ϕ: geocentric latitude (°N); λ: geocentric longitude (°E);
Δt: LT-UT (hours);
indices: L: LEO, G: GNSS, P: ray perigee

2. Simulation

2.1 Geometry, definitions, nomenclature

The simulation is based on satellite to satellite electron content ("horizontal TEC" — hTEC — because of the nearly horizontal ray paths with respect to the occultation ray) and parts of it (Table II). We use calculated satellite positions (true GNSS orbits and the approximate orbit of the planned METOP satellite for the LEO) and occultation by an ellipsoidal Earth. The heights refer to the ground-grazing occultation tangent point and for the electron densities we replace the ellipsoidal Earth by a spherical one using the distance of the ground-grazing occultation tangent point from the center of the Earth as the Earth radius.

2.2 Scenarios

We use 4 different occultation situations ("scenarios" S, T, U, V). Figure 1 shows map projections of the occultation rays, the rays with perigee heights $h_t = 800$ km and the ground-grazing rays (thin lines), the orbit projections for the GNSS and the LEO satellites during the occultation events (heavy lines at the ends of the ray projections — the long ones for the LEO satellite, the short ones for the GNSS satellite), and the ray perigees (heavy lines between the orbit projections but in the vicinity of the LEO orbit). It can easily be seen that substantial differences exist: In scenarios S and T the rays remain at mid- and high latitudes. In scenario U the rays remain in the (magnetic) Northern hemisphere, do not reach higher latitudes but the rays with low h_t are influenced by the equatorial anomaly region. In scenario V the rays are nearly meridional, the LEO satellite is in Northern mid latitudes, the GNSS satellite is in the Southern hemisphere at the border to higher latitudes. For scenarios S and T we expect not much of a plasmaspheric influence and a modest influence only from the ionosphere above points P. For scenarios U and V we expect substantially higher influence from both iTEC and pTEC.

Table III shows the properties of the ground-grazing occultation ray and the uppermost ray considered (800 km above the ground-grazing ray) for these four scenarios.

2.3 Electron density model

We have adopted the global electron density model NeUoG developed in Graz (Leitinger et al. 1996). It gives electron density as a function of solar activity (R_{12} or 10,7 cm solar radio flux (S_F), season (month of the year), Universal Time (or Local Time), geographic latitude, longitude, and height. It was prepared in such a way to be used along arbitrarily chosen point to point ray paths.

Recently a magnetic field aligned plasmasphere was added (NeUoG-plas, compare Leitinger and Kirchengast 1997).

2.4 Procedures and strategies

Step one: Choose one of the scenarios, solar activity (S_F value), month of the year and UT.

Step two: Apply an integration program which takes as input from a data file the end points of the occultation rays for 81 ray perigee heights from $h_t = 0$ to $h_t = 800$ km (10 km spacing). For each ray hTEC, iTEC, and pTEC are calculated through NeUoG-plas and stored in an output file. The vertical electron density profile vertically above the "midpoint" (ground-grazing occultation tangent point with $h_t = 0$) is calculated and stored in an other output file.

Step three: Invert a modified hTEC-profile to gain a horizontally averaged electron density profile. A very simple modification was adopted: subtraction of a constant value which was chosen to be 90% of iTEC+pTEC for the uppermost ray. For the inversion procedure it was assumed that all rays are parallel and coplanar which is a reasonable formal assumption. The 81 h_t values define the lower and

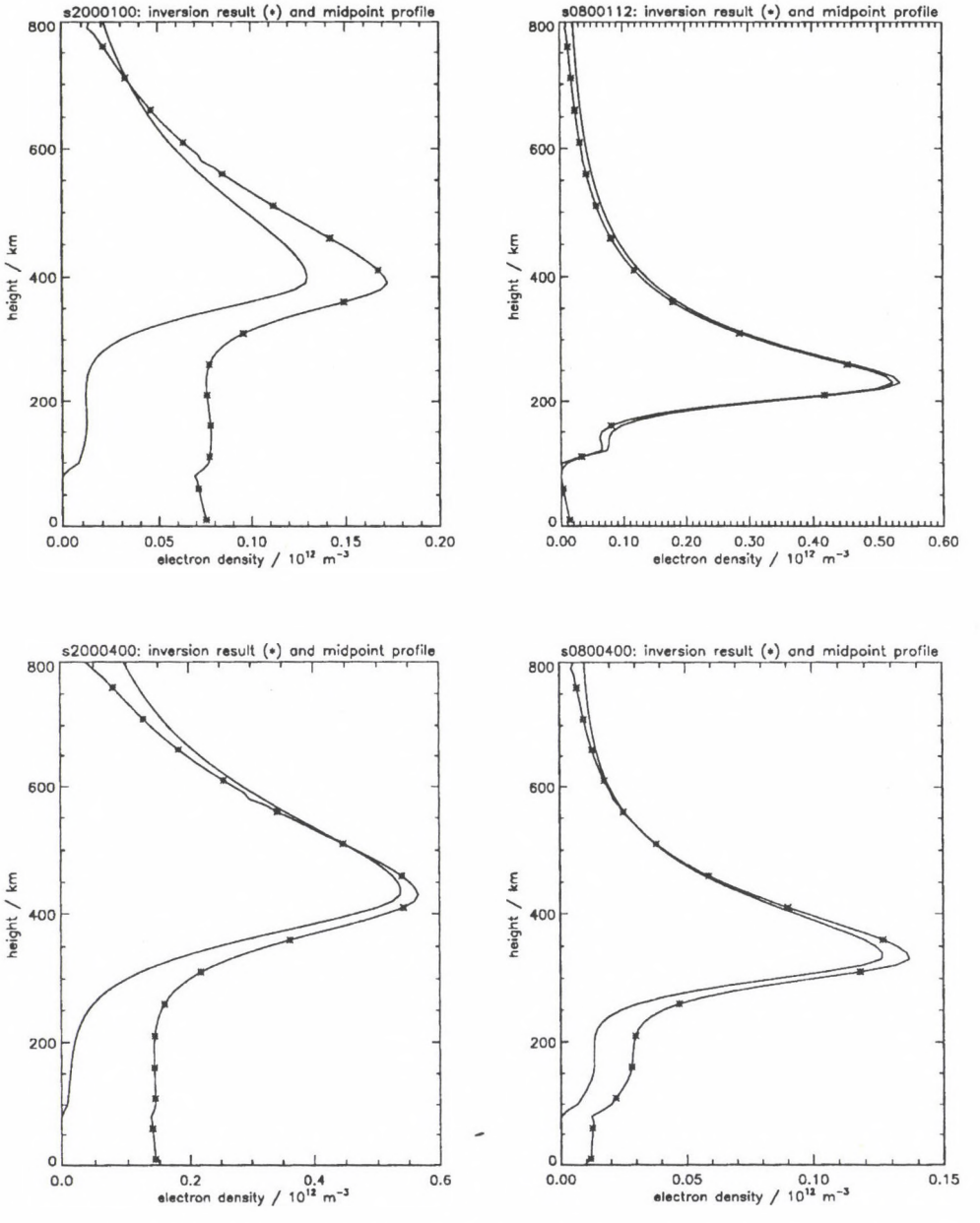


Fig. 2. Comparison of inversion result (*) and midpoint profile for scenario S cases. Nighttime cases (0 UT, 1.1 LT at the occultation tangent point) except top right (12 UT, 13.1 LT). High solar activity (left hand side: $S_F = 200$ units) and low solar activity (right hand side: $S_F = 80$ units)

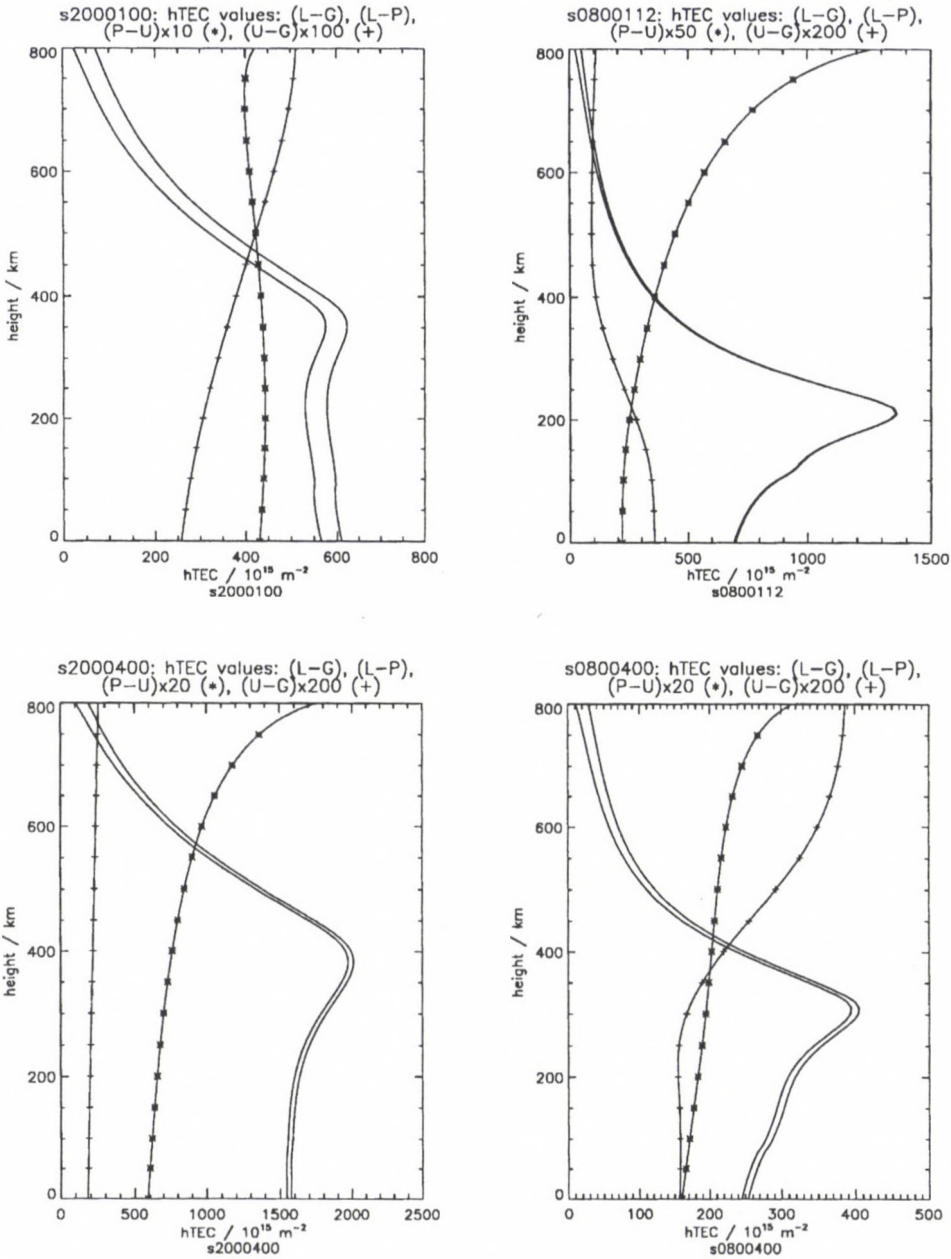


Fig. 3. Electron content along the occultation rays L–G for the scenario S cases of Fig. 2: from LEO to GNSS (points L to G – hTEC), from LEO to point P at LEO height (L–P – oTEC), for the uppermost ionosphere (P–U – iTEC), for the plasmasphere (U–G – pTEC). Electron content in units of $10^{15}\mathrm{m}^{-2}$, vs. height of the ray perigee in km

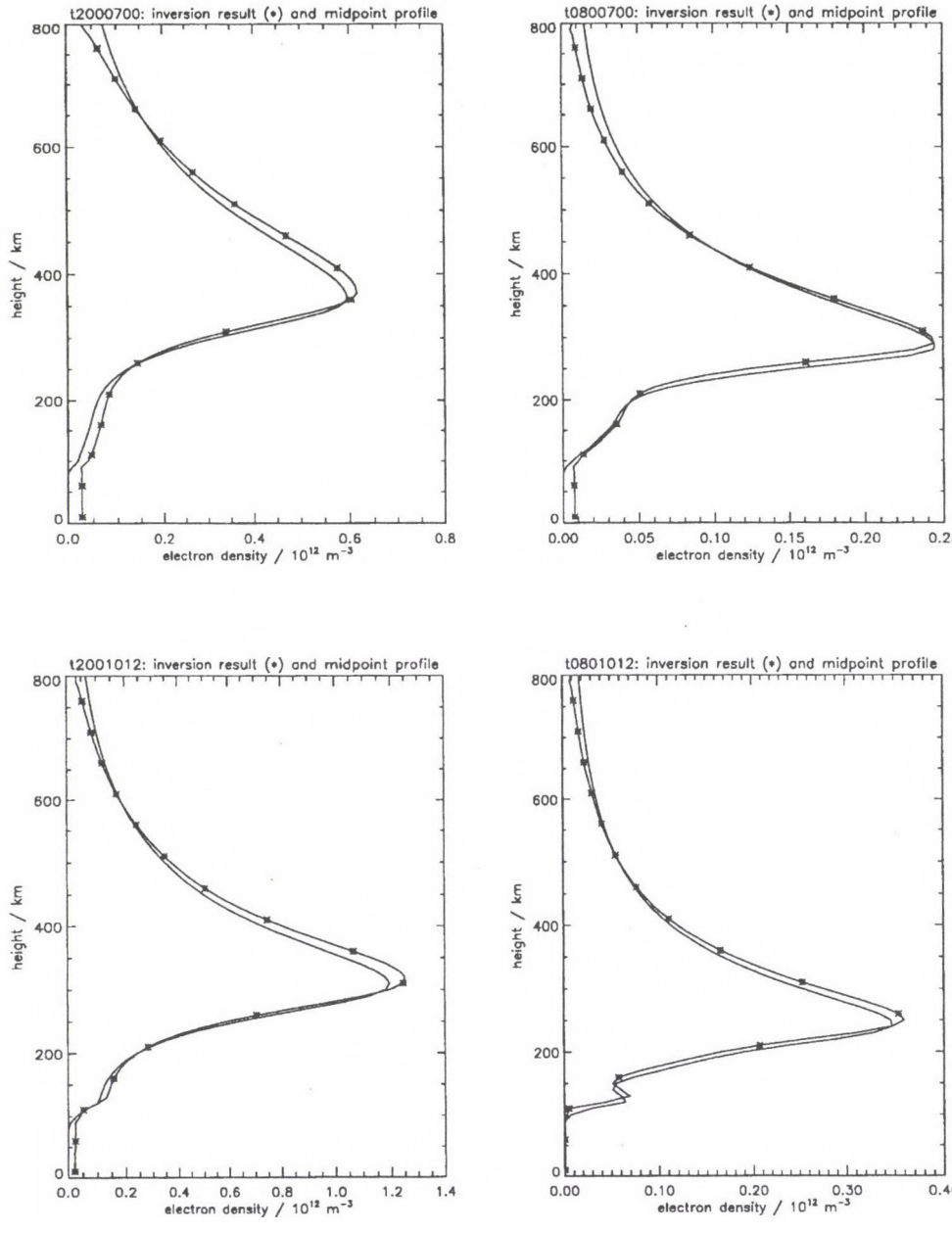


Fig. 4. Comparison of inversion result and midpoint profile for scenario T cases. Nighttime cases (top: 0 UT, 21.1 LT at the occultation point) and daytime cases (bottom: 12 UT, 9.1 LT). High solar activity (left hand side: $S_F = 200$ units) and low solar activity (right hand side: $S_F = 80$ units)

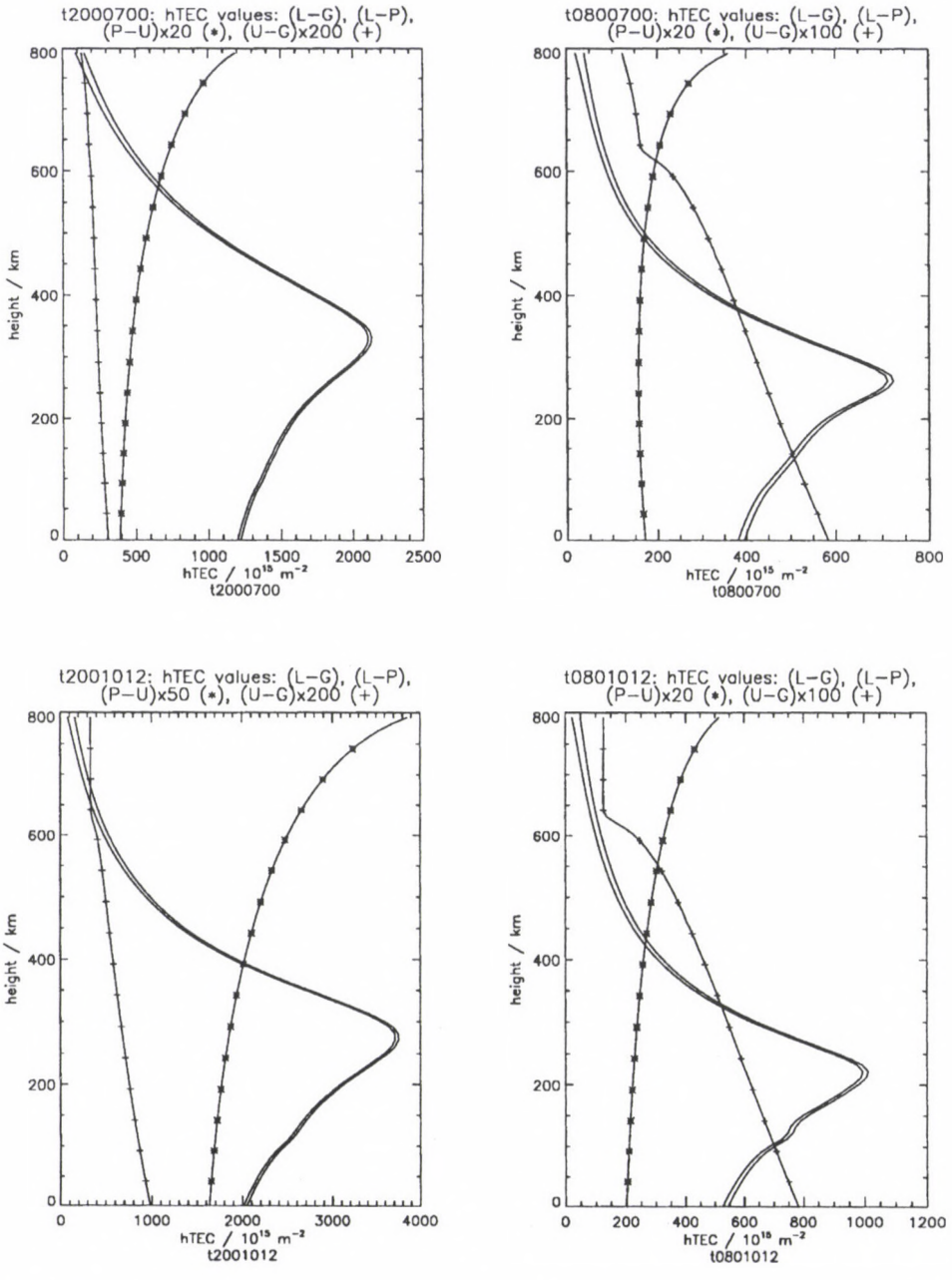


Fig. 5. Electron content along the occultation rays L–G for scenario T cases of Fig. 4: from LEO to GNSS (points L to G – hTEC), from LEO to point P at LEO height (L–P – oTEC), for the uppermost ionosphere (P–U – iTEC), for the plasmasphere (U–G – pTEC). Electron content in units of 10^{15}m^{-2}, vs. height of the ray perigee in km

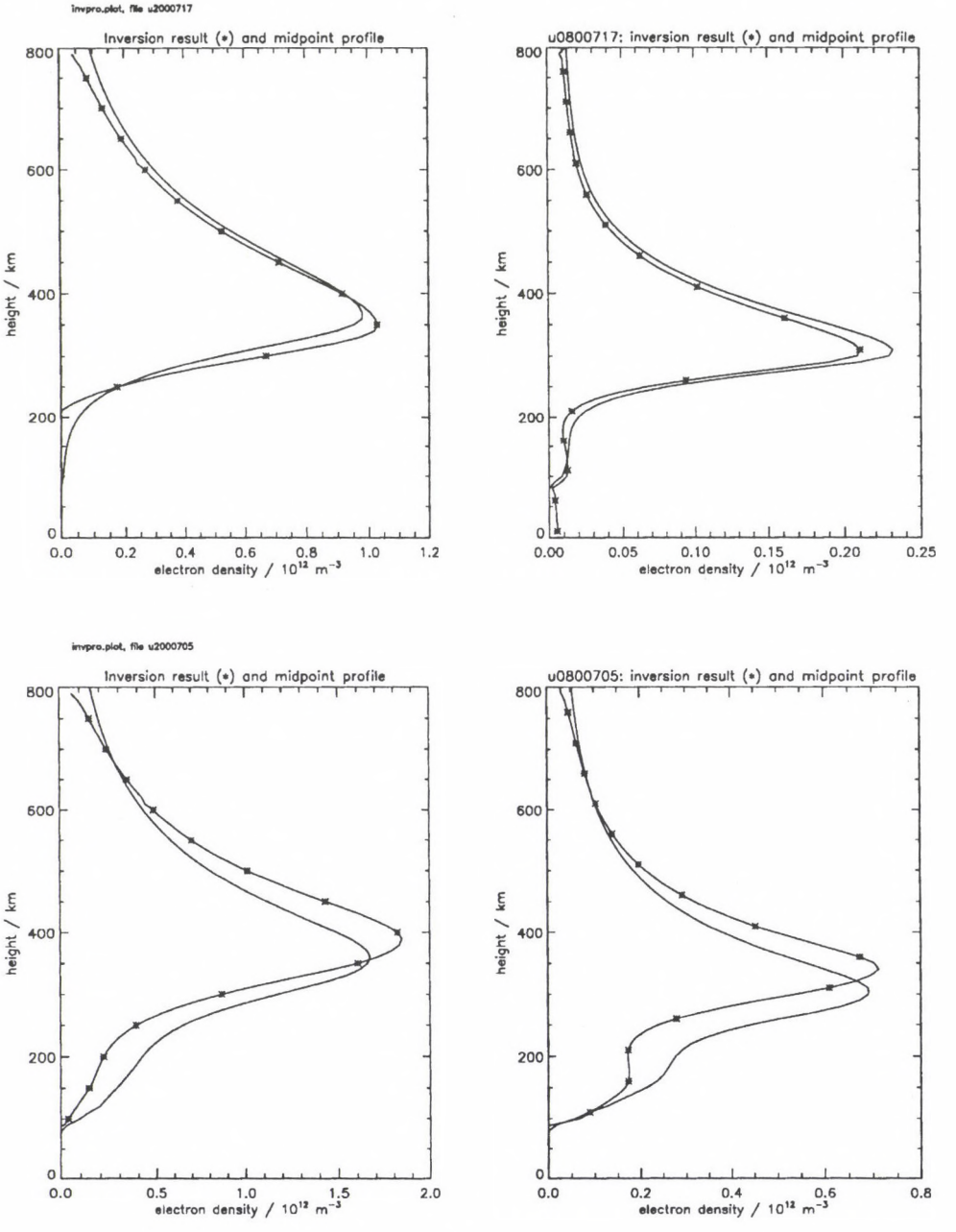


Fig. 6. Comparison of inversion result and midpoint profile for scenario U cases. Nighttime cases (top: 17 UT, 1.8 LT at the occultation point) and daytime cases (bottom: 5 UT, 13.8 LT). High solar activity (left hand side: $S_F = 200$ units) and low solar activity (right hand side: $S_F = 80$ units)

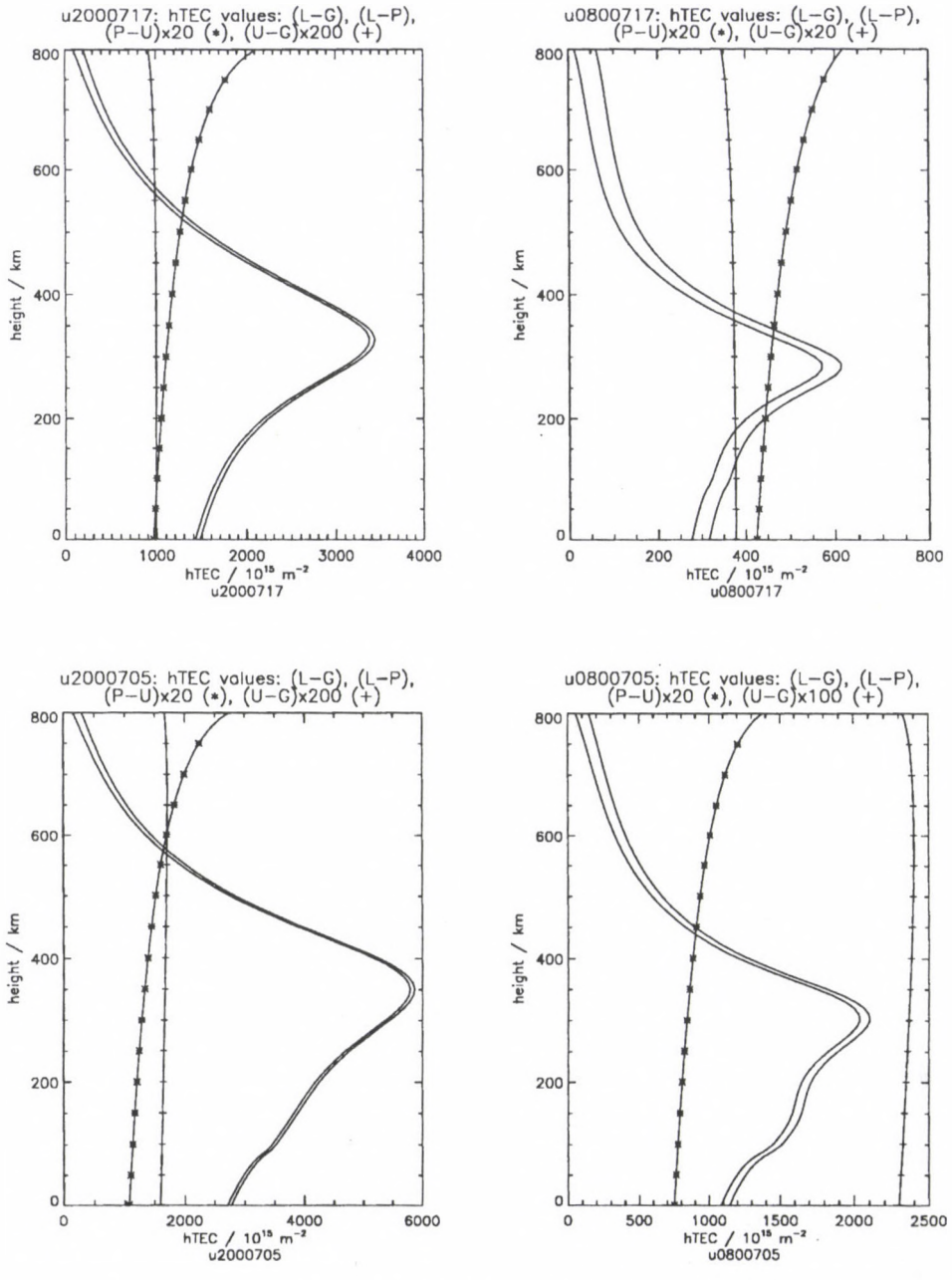


Fig. 7. Electron content along the occultation rays L–G for scenario U cases of Fig. 6: from LEO to GNSS (points L to G – hTEC), from LEO to point P at LEO height (L–P – oTEC), for the uppermost ionosphere (P–U – iTEC), for the plasmasphere (U–G – pTEC). Electron content in units of $10^{15}\mathrm{m}^{-2}$, vs. height of the ray perigee in km

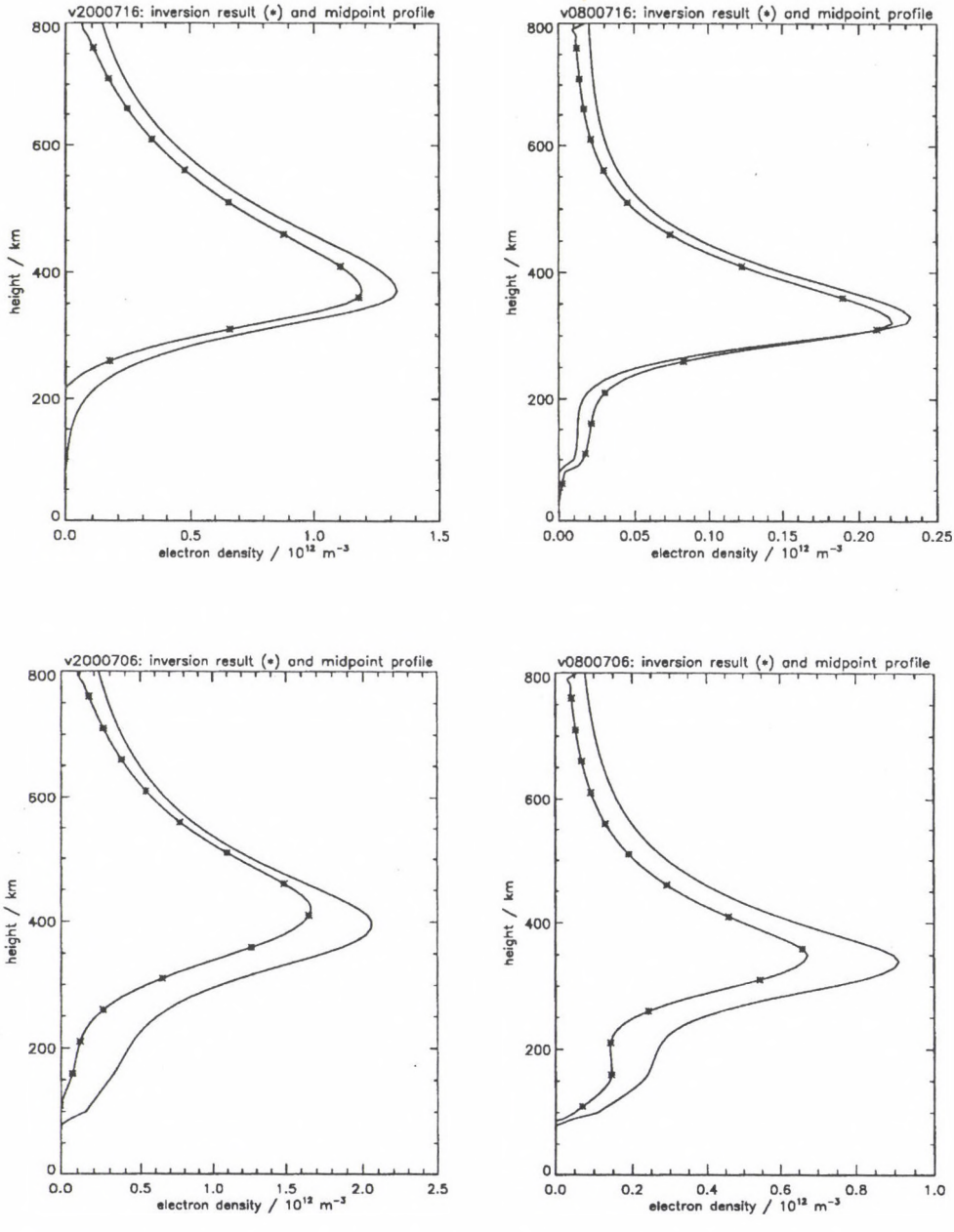


Fig. 8. Comparison of inversion result and midpoint profile for scenario V cases. Nighttime cases (top: 16 UT, 0.2 LT at the occultation point) and daytime cases (bottom: 6 UT, 14.2 LT). High solar activity (left hand side: $S_F = 200$ units) and low solar activity (right hand side: $S_F = 80$ units)

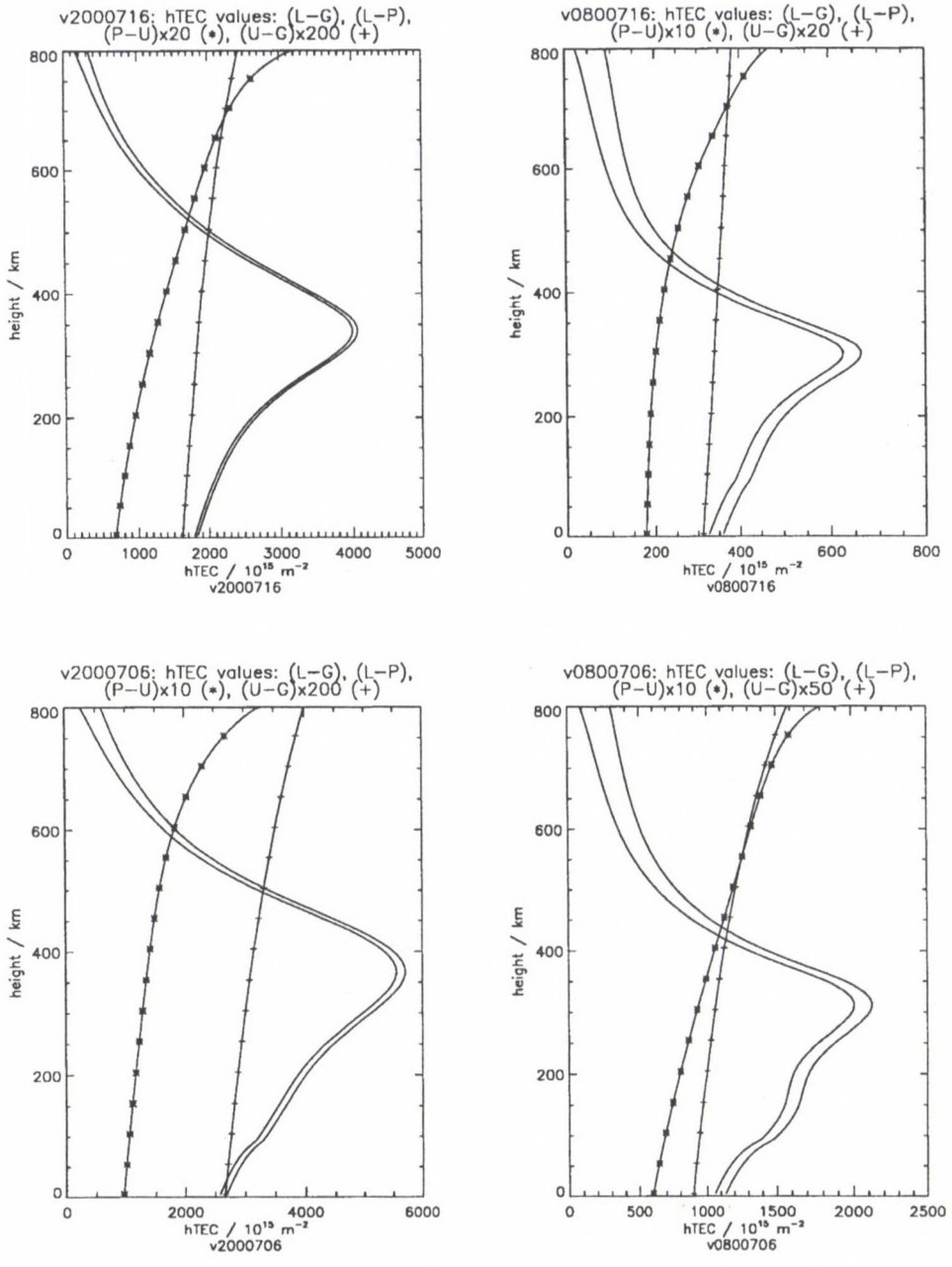


Fig. 9. Electron content along the occultation rays L–G for scenario V cases of Fig. 8: from LEO to GNSS (points L to G – hTEC), from LEO to point P at LEO height (L–P – oTEC), for the uppermost ionosphere (P–U – iTEC), for the plasmasphere (U–G – pTEC). Electron content in units of 10^{15}m^{-2}, vs. height of the ray perigee in km

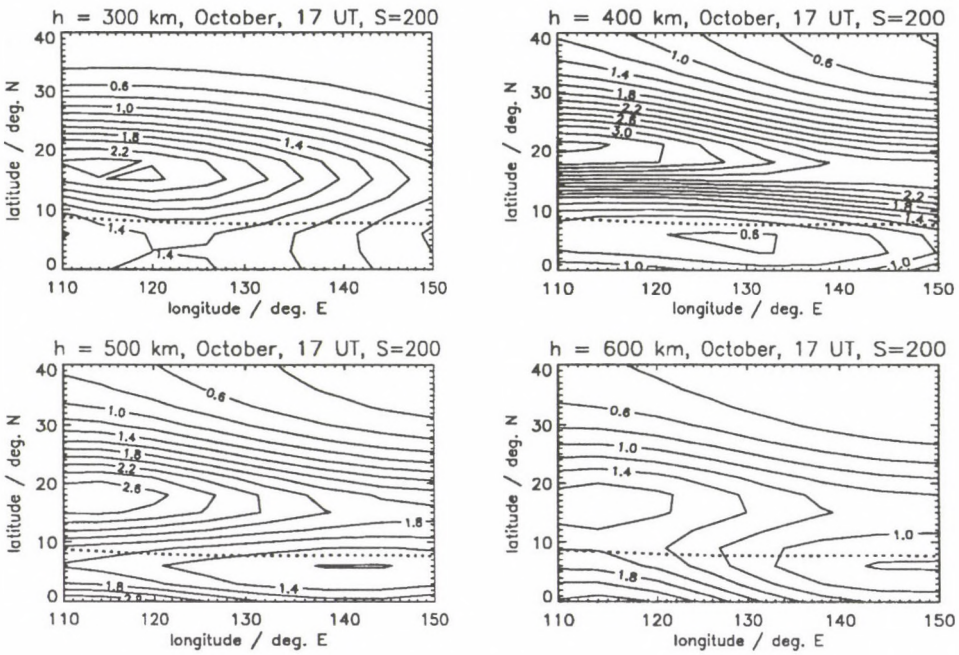


Fig. 10. Contour lines of electron density at heights of 300, 400, 500 and 600 km (left to right and top to bottom) from NeUoG-plas for October, 17 UT, $S_F = 200$ units. Region of importance for scenario U. Electron density in units of 10^{12} m^{-3}. Dotted line: CCIR dip equator

upper borders of 80 spherical layers. The procedure assumes a constant electron density gradient in each layer. A linear equation system is gained with a triangular design matrix solving for electron densities from the top down. Some artificial fluctuations can occur in the uppermost part which fade out after a few steps. Small errors accumulate downwards and the results for the lowest electron densities (small values when compared with peak values) become unreliable (in general below a height of 100 to 150 km).

Selection of input values needs a strategy to keep the number of simulations within reasonable limits. We have restricted to two levels of solar activity: $S_F = 80$ ("low solar activity") and $S_F = 200$ ("high solar activity"), to four months (January, April, July, October) and to local time of the ground-grazing occultation tangent point around noon and around midnight.

3. Sample results

In this section we present "typical" results from simulations based on our electron density model NeUoG-plas. Some electron density profiles for each of the four scenarios S, T, U, V are shown in the accompanying paper (Leitinger and Kirchengast 1997). For our occultation simulations the plasmapause was chosen to be at an "equivalent" L-value of 5 (compare Leitinger and Kirchengast 1997 for the mean-

ing of "equivalent" L-values). Therefore for all four scenarios the GNSS satellite is beyond the plasmapause. The rays for scenario S do not enter the plasmasphere at all, the rays for scenario T enter the plasmasphere from the LEO side near a height of 4000 km and cross the plamapause near $h = 13000$ km. The scenario U and V occultation rays cross from the ionosphere into the plasmasphere and cross the plamapause near $h = 11000$ km and $h = 7000$ km, respectively.

Sample results from the simulations are shown in Figures 2 through 9. In Figures 2, 4, 6 and 8 the inversion results (horizontally averaged height profiles of electron density) are compared with midpoint electron densities, "midpoint" meaning the ray perigee of the ground-grazing occultation ray ($h_t = 0$; compare Fig. 1). In Figs 3, 5, 7 and 9 electron content is shown for the occultation rays. The four panels of Fig. 3 correspond to the four panels of Fig. 2, etc.

The samples have been selected to show "bad" cases rather than "good" ones. This is especially true for two scenario S cases (left hand sides of Figs 2 and 3).

4. Discussion

So far the results of our simulations can be summarized as follows.

1. The influence of the plasmasphere is distinctly less severe than the influence of the uppermost ionosphere (Figs 3, 5, 7, 9). Higher LEO orbits would give substantially smaller iTEC influence. Major improvements can be expected from hTEC observations made before the LEO satellite reaches the h_t apex (L=P). Near co-planarity of the orbits (like for scenarios S and T) allows direct use of the iTEC+pTEC values gained in this way, diverging orbits (like for scenarios U and V) probably need some model assistance.

2. Horizontal gradients can act in such a way that the inversion results loose close relation to the midpoint profile. In our model ionosphere NeUoG-plas strong variations of the height profile with latitude and longitude only appear in the surroundings of the equatorial anomaly (see Fig. 10). In mid- and high latitudes such strong gradients can appear under severely disturbed conditions only. In strong-gradient cases evaluation of the occultation data should be done with incorporation of additional information. One possibility is the use of tomographic reconstruction (see Leitinger 1996, Leitinger et al. 1997b), an other is adaptive modelling or data assimilation (e.g., Daley 1991, Rodgers 1998).

3. When the occultation rays remain at mid- and high latitudes (scenarios S and T) most of the inversion results are "good" in the sense that they are representative for the midpoint profile. Even in the two "bad" cases shown for scenario S (Figs 2 and 3, lefthand sides) the peak heights match very well.

More simulations will be carried out in the near future including disturbed cases and model assisted inversion.

References

Daley R 1991: Atmospheric Data Analysis, Cambridge Univ. Press, Cambridge

Leitinger R 1996: In: Modern Ionospheric Science, H Kohl, R Rušter, K Schegel eds, Eur. Geophys. Soc., Katlenburg-Lindau, Germany, 346–370.

Leitinger R, Kirchengast G 1997: Easy to use global and regional ionospheric models — A report on approaches used in Graz, *Acta Geod. Geoph. Hung.* (present issue)

Leitinger R, Titheridge J E, Kirchengast G, Rothleitner W 1996: *Kleinheubacher Ber.*, 39, 697–704.

Leitinger R, Ladreiter H-P, Kirchengast G 1997: *Radio Sci.*, 32, 1657–1669.

Rodgers C D 1998: Inverse Methods for the Atmosphere: Theory and Practise. Cambridge Univ. Press, Cambridge

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 395–405 (1997)

MAPPING THE IONOSPHERIC F REGION BY MEANS OF SATELLITE TOMOGRAPHY

T NYGRÉN[1], E D TERESHCHENKO[2], B Z KHUDUKON[2], O V EVSTAFIEV[2], M LEHTINEN[3], M MARKKANEN[3]

In November 1995 a satellite tomography campaign was arranged in Scandinavia using a chain of four receivers extending from Tromsø in northern Norway to a receiving site close to Helsinki, Finland. Difference Doppler measurements were carried out by receiving signals from Russian navigational satellites flying approximately parallel to the chain. Electron density maps on the vertical plane between the satellites and the receivers were then calculated from these observations by means of stochastic inversion. The data set consists of 257 satellite passes during a period of 16 days. A computer video showing the meridional section of the F region above Scandinavia has been produced by interpolating the inversion results to 30-min time resolution. In this paper a set of selected observations is shown indicating the great variation of the auroral and sub-auroral F region with local time as well as with geomagnetic activity.

Keywords: difference Doppler method; electron density; ionosphere; tomography

1. Introduction

Ever since the suggestion by Austen et al. (1986, 1988), tomographic inversion has been used in determining the ionospheric electron density from total electron content measurements carried out by the difference Doppler or Faraday rotation methods. The conventional inversion methods in satellite radiotomography are various iterative algorithms, which have been compared in several studies (e.g. Raymund 1994, 1995, Kunitsyn et al. 1995). An alternative approach is to find the most probable values of the unknown electron densities by means of stochastic inversion. Different formulations of stochastic inversion have been introduced e.g. by Fremouw et al. (1992) and Markkanen et al. (1995). Satellite radiotomography has now been developed into a level which allows its use in geophysical research. This is convincingly demonstrated by comparisons with incoherent scatter observations (Mitchell et al. 1995, 1997). In the present paper, results of satellite tomography are shown obtained by the method of Markkanen et al. (1995). A more detailed description of this algorithm is given by Nygrén et al. (1997).

The difference Doppler measurements, which are most commonly utilised by satellite radiotomography, are usually carried out using beacon satellites flying close to 1000 km altitude. Two coherent radio waves at different frequencies, e.g. at 150 and 450 MHz, are received on the ground level and their phase difference is recorded. The phase difference is due to the dispersive properties of the ionospheric

[1]Department of Physical Sciences, University of Oulu, FIN-90570 Oulu, Finland
[2]Polar Geophysical Institute, 15 Khalturina, 183023 Murmansk, Russia
[3]Sodankylä Geophysical Observatory, FIN-99600 Sodankylä, Finland

1217-8977/97/$ 5.00

plasma. It can be easily shown that the phase difference is proportional to the integral of electron density along the ray from the satellite to the receiver. When the observations are made continuously at a number of sites in a chain on the ground level while a satellite is flying along the chain, the integrals will be determined along a great number of rays crossing each other in the ionosphere. This means that the experiment is suitable for calculating the electron density in the vertical plane by means of tomographic inversion.

2. Observations and data analysis

A satellite tomography campaign was arranged in Scandinavia during the period 14–29 October, 1995. The receiver chain consisted of four sites at Tromsø (69.662°N, 18.940°E) Norway, Esrange (67.877°N, 21.064°E) Sweden, and Kokkola (63.837°N, 23.058°E) and Kärkölä (60.584°N, 23.985°E) Finland. These sites lie approximately along a same magnetic meridian at distances of 200–450 km. Difference Doppler measurements were carried out using signals from Russian navigational satellites, which fly at 1000 km altitude approximately parallel to the chain during their southward passages. The path orientation of a northward passage is different, and therefore a fifth receiver was installed at Sodankylä (67.371°N, 26.623°E) Finland.

The total number of satellite passages registered during the campaign is 370. Since the measurement was not always successful at all sites, and only a part of the satellite flights were from north to south, the number of passages suitable for tomographic analysis is necessarily smaller. After a careful study, 257 passages were accepted for the analysis. These data were then inverted using the same analysis package as in Markkanen et al. (1995) and Nygrén et al. (1997), which is based on the stochastic method. In the analysis, an inversion grid is defined on the vertical plane above the receiver chain, and the inversion results are electron densities at the grid points. Due to the curvature of the Earth, the grid elements have annular rather than rectangular shapes. The horizontal size of the elements was 42.75 km on the ground level, except at the edges where larger sizes were used. The vertical element size was 33.3 km below 100 km and above 700 km, and 25 km between these two heights.

The tomographic analysis gives plots of the ionospheric electron density on the vertical plane above the chain at intervals varying from 20 min to a couple of hours. The average time separation of successive plots is about 1.5 h. A computer video was made by interpolating the results to 30-min time resolution. This gives a possibility to see the behaviour of the ionospheric electron density during a period of more than two weeks, speeded up to a 5-min presentation. One must, however, remember that the video contains interpolation over gaps of a few hours, so that it is not capable of describing the ionosphere correctly at all times.

In order to give a general picture of the observations, selected plots of the tomographic reconstructions are presented in this paper. These are chosen to demonstrate the variability of the ionospheric F region at auroral and sub-auroral latitudes. The figures shown are not the interpolated ones used for the video but the original reconstructions which describe the ionosphere more accurately than the video does.

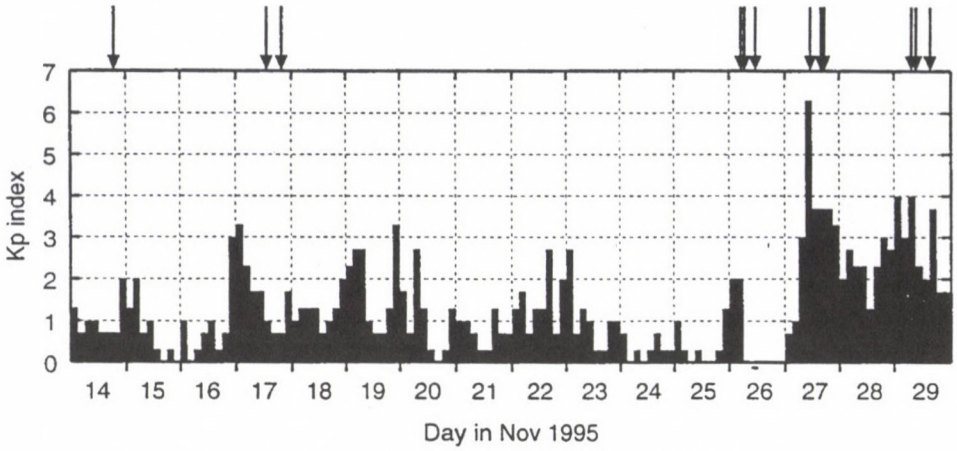


Fig. 1. K_p index during the tomographic campaign in November 1995. The times of observations shown in Figs 2–5 are indicated by arrows

The magnetic activity is low during most of the campaign period, but near the end the activity strongly increases. Therefore the effect of magnetic activity on the ionospheric F region can be visualised using these data.

3. Results

The behaviour of the K_p-index during the campaign period is portrayed in Fig. 1. This shows that, until 27 Nov, the general level of magnetic activity was generally quite low, with $K_p \leq 1$ most of the time and reaching only occasionally close to the level $K_p = 3$. A sudden change took place on 27 Nov. during the last three days of the campaign usually $1 \leq K_p \leq 4$, only in one case $K_p = 1^-$ and on one occasion $K_p = 6^+$. The arrows on top of the panel indicate the times of the reconstructions shown in later figures.

Figure 2 portrays the development of the F layer from early morning to noon during a period of low magnetic activity. The top panel shows a broad F layer trough with a width of the order of 10° and the F layer peak at the edges of the troughs at 300–400 km altitude. In the middle of the trough, a weaker cloud of ionisation with a peak density of the order of 10^{10} m^{-3} is observed at essentially lower heights, at 200 km or even below. The latitudinal width of this cloud is about 5°. This example contains the main features of the trough during the midnight or early morning hours, which are continuously repeated during the whole campaign whenever the magnetic activity is low. Especially the low-altitude cloud is a persistent feature. The trough appears already early in the evening, at least after 17 UT, and before midnight it can be even wider than that shown in Fig. 2.

The middle panel in Fig. 2 presents the northward expansion of higher F region electron density from the south during the morning hours. In this process the trough will be gradually filled while the high electron densities at its northern edge are

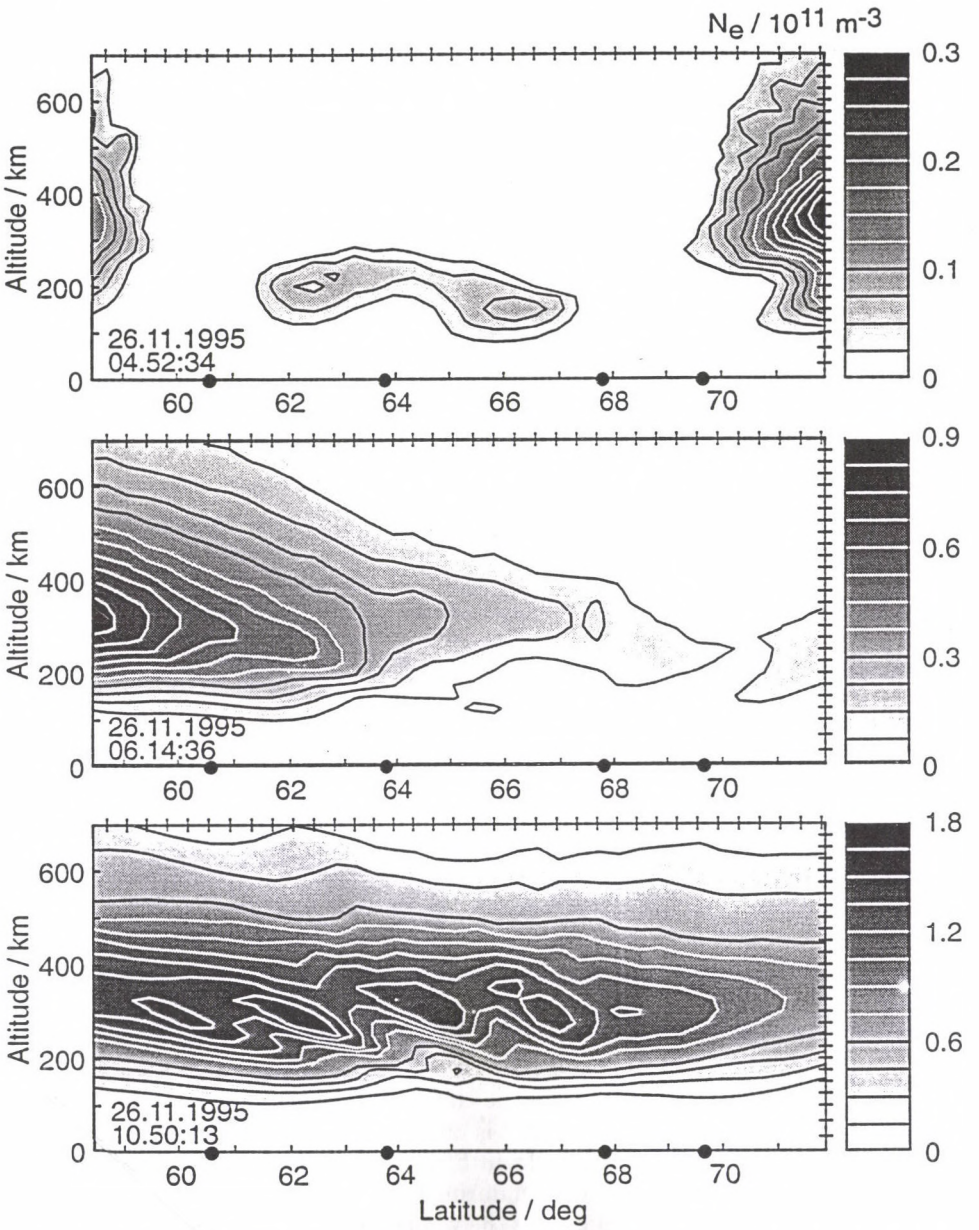


Fig. 2. Tomographic reconstructions of the ionospheric electron density above the receiver chain during satellite passages starting at 0452:34, 0614:36 and 1050:13 UT on 26.11.1995, in terms of gray scale and contour plots. The circles on the ground level indicate the receiver sites. Notice the difference of density scales in various panels

significantly reduced. Indications of an F region tilt are also visible. The sequence of individual plots actually indicate that the low-lying cloud at the middle of the trough gradually merges with the southward edge thus forming the tilt.

The bottom panel of Fig. 2 shows a quiet daytime horizontal F layer with its peak at a fixed altitude of about 300 km. The peak electron density is higher than 1.5×10^{11} m^{-3} but, in the northernmost part of the figure, the density is clearly smaller. The most fascinating feature in this panel is a set of tilted wave fronts at 300 km altitude revealing the presence of travelling ionospheric disturbances. The horizontal wave length of these disturbances is of the order of 250 km so that they are obviously due to medium scale atmospheric gravity waves. Assuming the source of the waves to be at lower heights, the tilt of the wave fronts indicates southward propagation. During the whole campaign, gravity waves were quite commonly observed in the daytime, especially at pre-noon hours. One should notice that the absence of the wave fronts in the northernmost part of the figure may be an artifact caused by the fact that the measurements do not contain information from rays oriented in the direction of the wave fronts. This possibility was convincingly demonstrated by Saksman et al. (1997).

Figure 3 portrays the behaviour of the F layer during a slightly more active period. The top panel contains a typical horizontal daytime layer with a maximum above the northernmost part of the receiver chain. This maximum is very commonly observed at daytime, quite often structured by travelling ionospheric disturbances such as those in the bottom panel of Fig. 2. The intensity of the travelling ionospheric disturbances is variable, but it is usually weaker than in Fig. 2.

The middle and bottom panel of Fig. 3 show troughs which are so wide that their southern edges are hardly visible in the figures. In these cases the layer peaks at the northern edge of the trough are at clearly lower heights than in Fig. 2. A plasma enhancement at the centre of the trough is visible in the bottom panel but this time actually at a greater altitude than the peak of the northern edge. A prominent feature in these figures is the remarkable steepness of the northern edge. In the middle panel, where the steepness is greatest, the edge is oriented in the direction of the geomagnetic field, which suggests that the edge and its steepness is controlled by particle precipitation sharply restricted to a narrow latitudinal range. In both cases there is a narrow region just to the south of the northern edge where the electron density is even lower than further in the south.

Figures 4 and 5 contain snapshots from the geomagnetically active period at the end of the campaign. The top panel of Fig. 4 is a daytime F layer from the most active interval with $K_p = 6^+$. In this case the trough is visible in the north already at noon time. The trough is narrow, only about 2° in width, and much higher electron densities are encountered within the trough region than in the previous examples. A plasma enhancement below 200 km altitude is found close to the southern edge of the trough between 68° and 70° latitudes. A denser contour plot would indicate that this enhancement does exhibit a maximum so that it looks like a cloud separate from the upper layer. The upward bulge of the topmost contour indicates that plasma enhancements are present also at great altitudes above the trough.

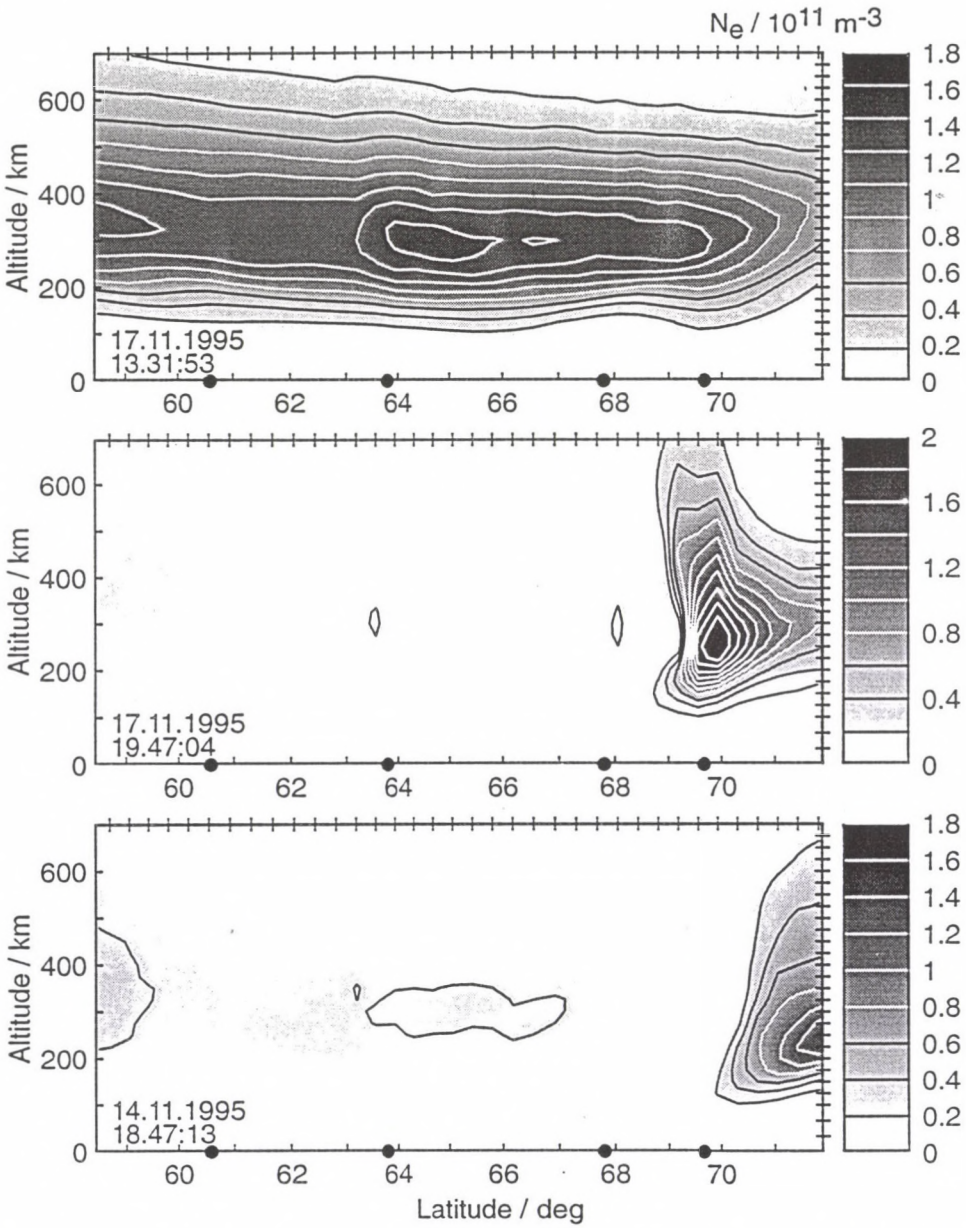


Fig. 3. Same as Fig. 2 for satellite passage times 1331:53 and 1947:04 UT on 17.11.1995 and 1847:13 UT on 14.11.1995

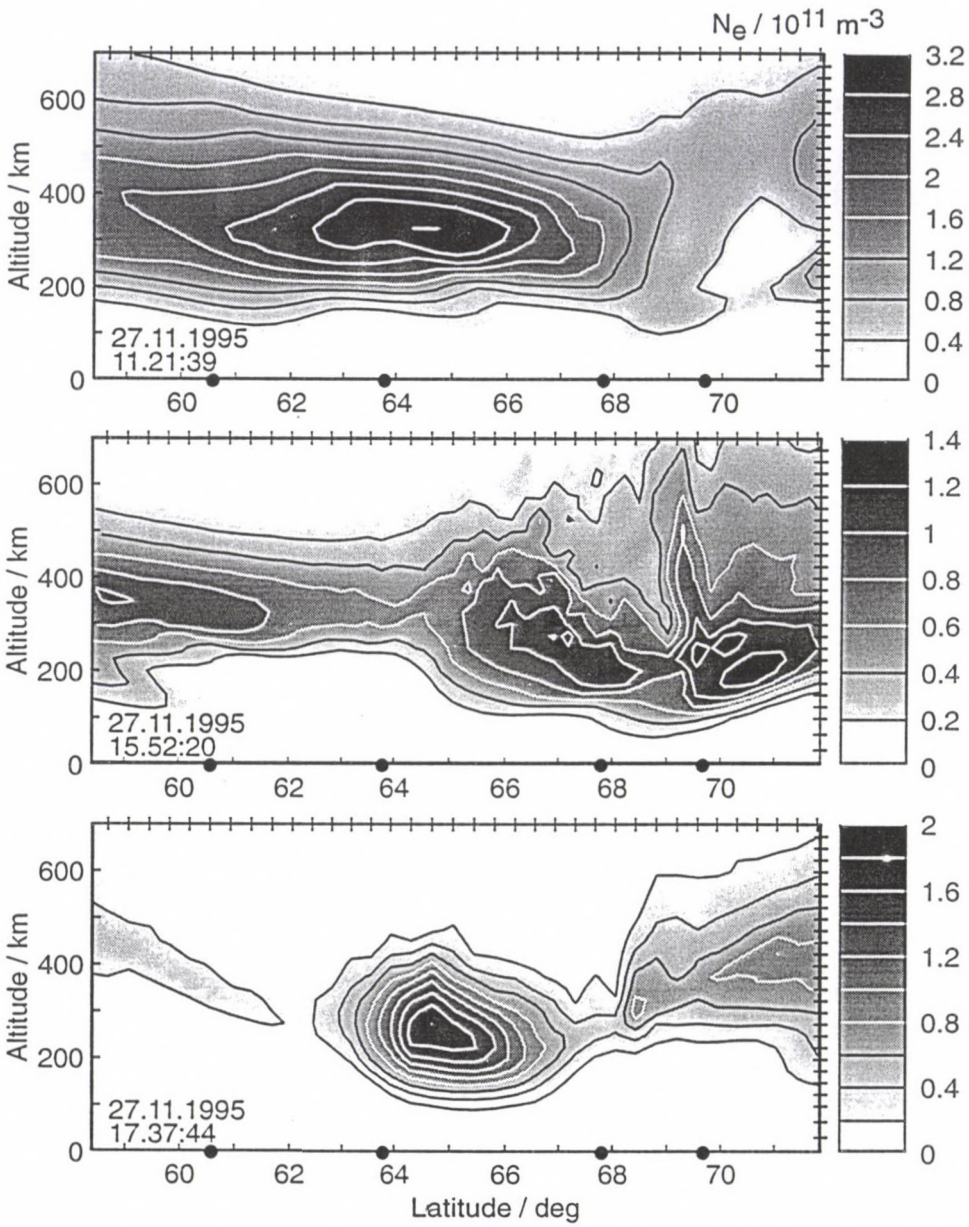


Fig. 4. Same as Fig. 2 for satellite passage times 1121:39, 1552:20 and 1737:44 UT on 27.11.1995

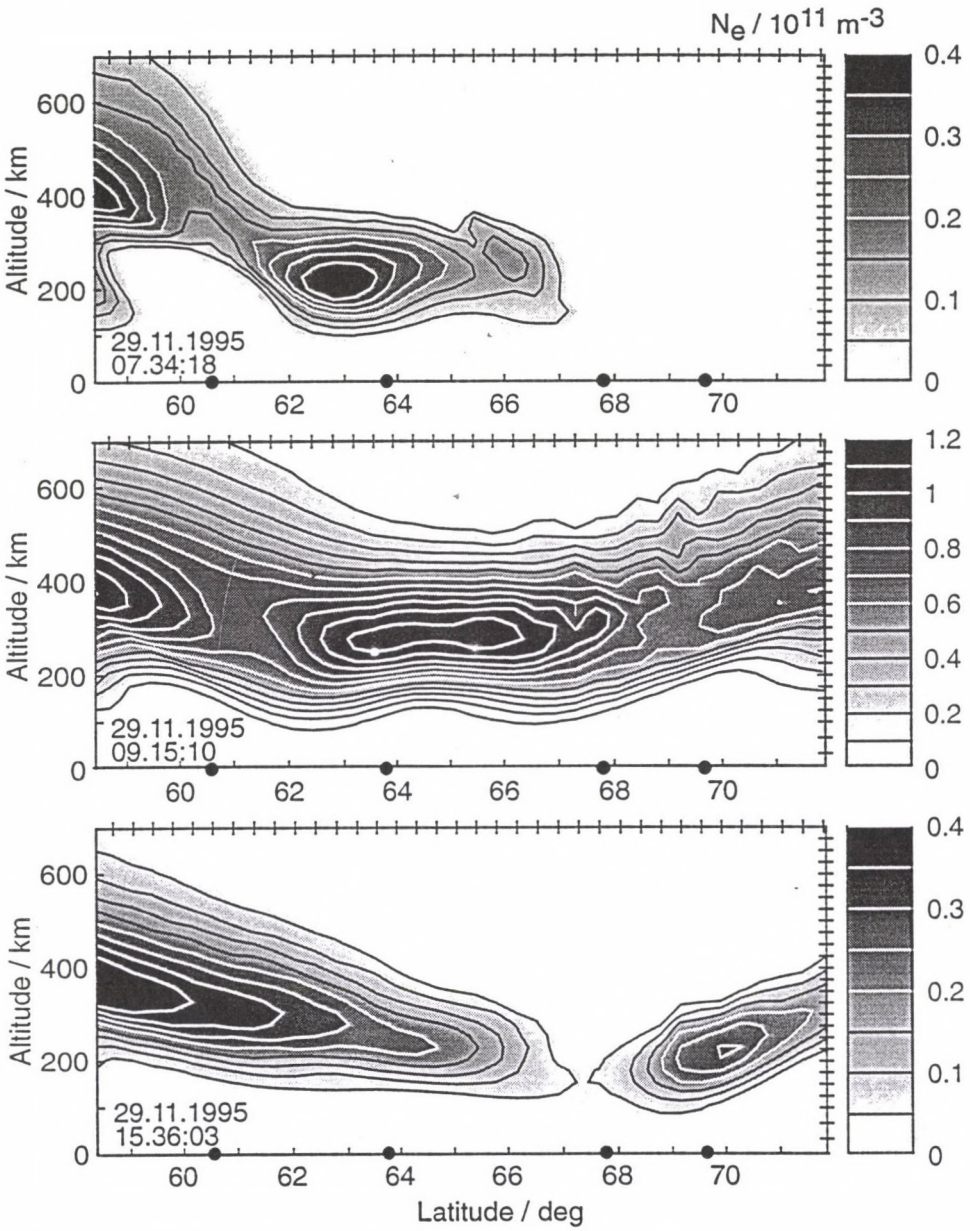


Fig. 5. Same as Fig. 2 for satellite passage times 0734:18, 0915:10 and 1536:03 UT on 29.11.1995

In the middle panel of Fig. 4, portraying the situation about 4.5 h later, the trough has moved about 7°–8° to the south. To the north of the trough, the F region consists of large patches and, just below 70° latitude, a thin plasma finger reaches great altitudes pointing in the direction of the local geomagnetic field. The main feature in the bottom panel is the large density enhancement within the latitude range 63°–67°. This blob is a long-lasting feature, which is observed in several satellite passes approximately during two hours.

On 29 November the magnetic activity has continued for a couple of days. Figure 5 contains morning and daytime F layers from this period. These examples are quite different from the previous ones; in all cases the F region contains, within the field of view, strong tilts which seem to be associated with a wave-like ionospheric disturbance. From the middle panel one can obtain a very rough estimate of 1500–2000 km for the disturbance wave length. Although no definite proof can be based only on these data, a tentative explanation could be that, during the preceding period of magnetic activity, a sufficient amount of energy has been fed to the neutral atmosphere to generate large scale gravity waves, which are revealed by their vigorous effect on the ionospheric F layer.

4. Discussion

The power of ionospheric radiotomography is convincingly demonstrated by the above examples. By means of a relatively inexpensive equipment (if we exclude the costs of satellites), snapshots of the ionospheric F region electron densities can be obtained from regions which are much wider than those which can be mapped by incoherent scatter radars, for instance. The field of view is determined by the length of the receiver chain. By installing longer chains, larger areas than those shown in the above figures can be easily mapped with the same satellites. It is obvious that this method could provide very valuable information if automatic or semiautomatic receiving sites connected to computer networks were established to carry out measurements on continuous basis. The availability of suitable satellites on long terms may, however, limit the interest in building such facilities.

In spite of its benefits, ionospheric tomography has also some limitations. The temporal resolution is determined by the satellite flight time within a field of view, which is of the order of 15–20 min. If large changes take place in the ionosphere during this interval, the method is no more capable of giving correct results. The availability of suitable satellite passages also restricts the temporal resolution since changes occurring between satellite passages may remain unobserved. Another limitation is provided by the fact that the satellites do not fly exactly along the receiver chain, and therefore one has to assume that the ionosphere contains no horizontal gradients in the direction perpendicular to the chain.

Limitations of a different type are associated with the mathematical principles of the tomographic method itself. It can be shown that the data collected by a few ground based receivers are so defective that their mathematical inversion is all but unique; in other words, an infinite number of solutions satisfying the observations exist (see e.g. Saksman et al. 1997). In the present inversion method regularization

variances are used in guiding the solver to a result which is likely to be close to the correct one. The problem is not so serious as it might look at first sight, since seriously incorrect regularization variances lead to clearly incorrect results. In analysing the present observations, for instance, the whole data set was inverted in a routinelike manner using several vertical profiles of regularization variances. The results shown in the present paper were obtained by a single regularization profile which was estimated to lead to the most plausible results on the average. What is important, however, is that the main features of the ionosphere remained the same in all inversions which were not clearly in error. The above results indicate, for instance, that a single regularization profile is capable of leading to a great variety of layer thicknesses and peak altitudes. Therefore, even a routine analysis is capable of giving at least a correct qualitative picture of the spatial ionospheric structures, although errors in the absolute values of the electron density may occur. More reliable results can be obtained by a careful choice of the regularization variances. The accuracies thus gained are presented by Markkanen et al. (1995) and Nygrén et al. (1997).

The present tomographic analysis package has been designed in such a way that it allows the inclusion of other type of data in the inversion problem. If, for instance, the ionospheric electron density profile is determined with a good accuracy by an incoherent scatter radar, these observations with appropriate data variances can be merged to the data set, and the inversion result above the radar will be mainly determined by the radar measurement. Bottomside profiles obtained from ionosondes can be utilised for the same purpose. A good knowledge of electron density profile at a single site will help in choosing the correct regularization variances and it will also improve the accuracy of the inversion result within a wider region, probably within the whole field of view. Until now, these possibilities have not been used in their full power.

Like any other geophysical observational method, satellite radiotomography has its benefits and limitations. Therefore, whenever possible, is should be used together with other instruments for getting the best possible picture of the state of the ionosphere.

Acknowledgements

The authors are grateful to S M Chernyakov, J Pirttilä, P Pollari, E Saviaro and T Ulich for assistance in the tomographic measurements.

References

Austen J R, Franke S J, Liu C H, Yeh K C 1986: In: Radio beacon contribution to the study of ionisation and dynamics of the ionosphere and corrections to geodesy, A Tauriainen ed., Oulu, Finland, Part 1, 25.

Austen J R, Franke S J, Liu C H 1988: *Radio Sci.*, 23, 299.

Fremouw E J, Secan J A, Howe B M 1992: *Radio Sci.*, 27, 721.

Kunitsyn V E, Andreeva E S, Popov A Y, Raznikov O G 1995: *Ann. Geophys.*, 13, 1263.

Markkanen M, Lehtinen M, Nygrén T, Pirttilä J, Helenius P, Vilenius E, Tereshchenko E D, Khudukon B Z 1995: *Ann. Geophys.*, 13, 1277.
Mitchell C N, Jones D G, Kersley L, Pryse S E, Walker I K 1995: *Ann. Geophys.*, 13, 1311.
Mitchell C N, Kersley L, Heaton J A T, Pryse S E 1997: *Ann. Geophys.*, 15, 747.
Nygrén T, Markkanen M, Lehtinen M, Tereshchenko E D, Khudukon B Z 1997: *Radio Sci.* (in press)
Raymund T D 1994: *Int. J. Imag. Syst. Tech.*, 5, 75.
Raymund T D 1995: *Ann. Geophys.*, 13 1254.
Saksman E, Nygrén T, Markkanen M 1997: *Radio Sci.*, 32, 605.

Acta Geod. Geoph. Hung., Vol. 32(3-4), pp. 407–416 (1997)

SINGLE STATION AND SINGLE SATELLITE METHOD OF GPS IONOSPHERIC DATA PROCESSING

L BÁNYAI[1]

The Single Station and Single Satellite Method is introduced as an alternative of the world-wide used method, where the multi-satellite GPS observations are analized for ionospheric purposes in one processing procedure.

According to the analogy of spherical distances, "ionospheric" distances are computed from the geographic latitude and local time of the ionospheric points along the satellite pass, therefore the TEC values can be modelled as single variable quantities using the Legendre polynomials.

The least-squares adjustment of the observations by the Gauss-Markov model shows very large correlation between the code bias and the zero as well as the even order terms of the polynomials, which leads to false solution usually over the 2nd order. The zero and odd order terms of the Legendre polynomials proved to be a good choice, where the inner accuracy of code bias is better than 1 cm and the larger changes of TEC profile can be described properly.

The comparison of the results with the procedure developed at DLR Remote Sensing Ground Station, Neustrelitz, Germany, shows the effectiveness of the method.

Keywords: code synchronization bias; Global Positioning System, GPS; Legendre polinomials; total electron content, TEC

1. Introduction

The main characteristic of the world-wide used methods for the ionospheric application of GPS measurements is the processing of multi-satellite observations in one data processing step. The observations are carried out at least at one station but usually the local (one country) or regional (e.g. one part of the International GPS Service for Geodynamics – IGS) network is preferred. The compilation of a world-wide total electron content (TEC) map needs a global network. These procedures may be characterized as Multi-Station and Multi-Satellite Methods. According to the rules of the IGS organization usually one day of observations with 30 second sampling rate are processed in one processing step.

In the multi-satellite approach the vertical TEC values are modelled as low order surface polynomials (Lanyi and Roth 1988, Wanninger et al. 1994) or modified harmonic series are used (van der Marel and Georgiadou 1994) where the variables are the geometric latitude (ϕ) and the local time (t) of the ionospheric points.

As an alternative of the above approach the Single Station and Single Satellite Method is developed and introduced in this paper.

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

1217-8977/97/$ 5.00

2. Single Station and Single Satellite Method

The basic observation equations are the geometry-free linear combinations with the synchronization error model given by Wanninger (1994) as follows:

$$R_2 - R_1 = c(dt^r_R + dt^s_R) + \frac{1}{\sin E'}\frac{A}{2}\frac{f_1^2 - f_2^2}{f_1^2 f_2^2} I_v + \varepsilon_R \tag{1}$$

$$\lambda_2\phi_2 - \lambda_1\phi_1 = (\lambda_2 N_2 - \lambda_1 N_1) + c(dt^r_\phi + dt^s_\phi) - \frac{1}{\sin E'}\frac{A}{2}\frac{f_1^2 - f_2^2}{f_1^2 f_2^2} I_v + \varepsilon_\phi \tag{2}$$

where

R	pseudorange or code observable (m)
ϕ	phase observable (cycle)
N	phase ambiguity (initial phase unknown)
f	frequency (1/s)
c	speed of light (m/s)
λ	wavelength (m)
I_v	vertical total electron content of the ionosphere
E'	elevation angle with respect to the ionospheric point computed from the single layer model (Fig. 1)
A	constant (80.6 $m^3 s^{-2}$)
dt^r_R	code synchronization delay difference of the receiver
dt^s_R	code synchronization delay difference of the satellite
dt^r_ϕ	phase synchronization delay difference of the receiver
dt^s_ϕ	phase synchronization delay difference of the satellite
ε	random observation errors and not modelled errors: multipath, phase center variation and higher order ionospheric effects

The indices 1 and 2 mean the L1 and L2 carriers, respectively.

The sum of Eq. (1) and Eq. (2) — the so called geometry- and ionosphere-free linear combination — can be defined as

$$\begin{aligned} R_2 - R_1 + \lambda_2\phi_2 - \lambda_1\phi_1 &= c(dt^r_R + dt^s_R) + (\lambda_2 N_2 - \lambda_1 N_1) + c(dt^r_\phi + dt^s_\phi) + \varepsilon \\ &= A_0 + A_1 \cdot (t - t_m) + \varepsilon\,. \end{aligned} \tag{3}$$

Removing the cycle slips this combination should be a constant value, however significant linear trend was experienced even in the case of P-code measurements (Bányai and Eper-Pápai 1996). The problem is solved by means of fitting a linear trend, where t_m is the mean observation time (UT) of the satellite pass, A_0 contains the mean biases at t_m and A_1 describes the slope of the line.

The subtraction of A_0 from Eq. (2) leads to the optimum ionospheric combination:

$$\lambda_2\phi_2 - \lambda_1\phi_1 - A_0 = -\frac{1}{\sin E'}\frac{A}{2}\frac{f_1^2 - f_2^2}{f_1^2 f_2^2} I_v - c(dt^r_R + dt^s_R) + \epsilon \tag{4}$$

where the phase ambiguities and biases are replaced by the more meaningful code synchronization biases. While the inner accuracy of this combinations can be characterized by the accuracy of phase measurements (2–3 mm) the outer accuracy can

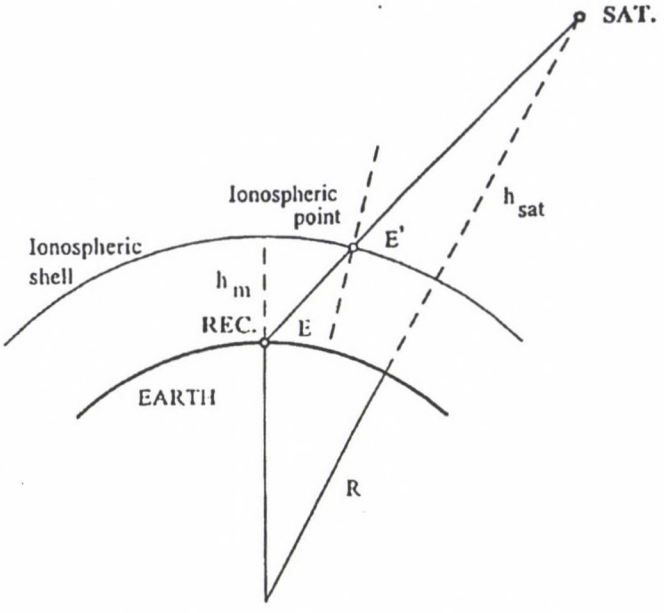


Fig. 1. The geometry of the single layer model

be characterizead by the accuracy of A_0 computed from the fit of the linear trend. The error of A_0 value constantly shifts the least-square estimate of the code biases. According to our experiences with TurboRogue GPS receiver in Sopron the standard deviations of linear fit are typically 5–10 cm (0.2–0.3 ns) in P-code and 10–30 cm (0.3–1.0 ns) in Cross-correlation tracking mode respectively. Because this combination contains all the not modelled errors it can be used to investigate the noise of the observation sites as well.

One example of different combinations can be found in Figs 2a-2c. In Fig. 2a the geometry-free linear combinations can be found where the sign of phase measurements is changed and shifted by A_0, which shows the asymmetry between the phase and the code measurements. In Fig. 2b the geometry- and ionosphere-free linear combinations and their linear trend are presented. In Fig. 2c the linear trend is removed from the code combinations.

In the case of single satellite observations the ionospheric points do not describe a surface, therefore the vertical ionospheric effects cannot be modelled with the polynomials used in multi-satellite approach. According to the analogy of spherical distance, "ionospheric" distance is introduced, where the geographic longitude is replaced by local time:

$$\cos(\vartheta_{AB}) = \sin(\phi_A)\sin(\phi_B) + \cos(\phi_A)\cos(\phi_B)\cos(t_A - t_B)\,.$$

The subscripts A and B denote two neighbouring ionospheric points, ϕ is a geographic latitude and t is a local time scaled to radian. From the series of ϑ_{AB} values along one satellite pass a new single variable can be composed which includes

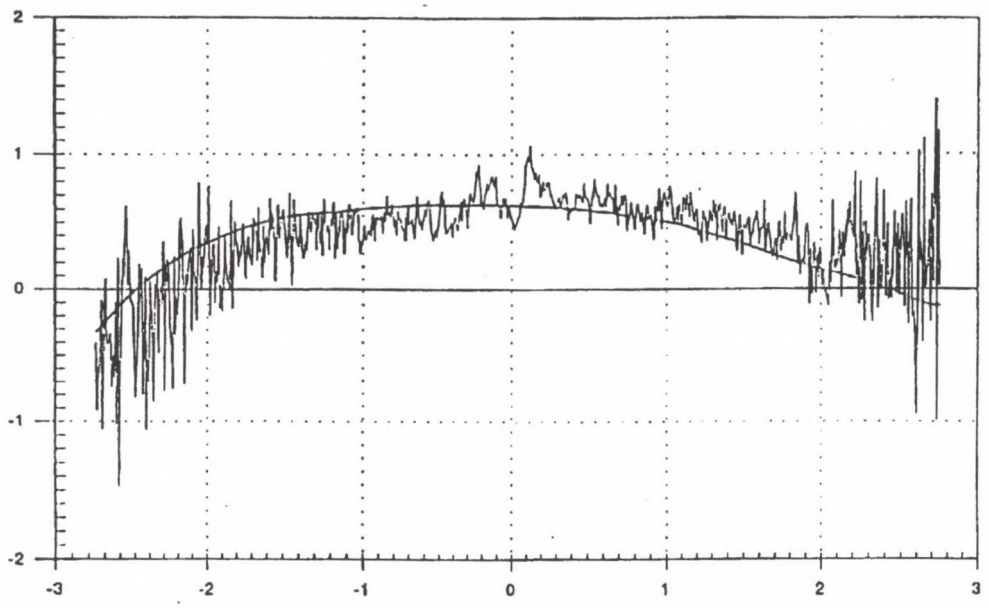


Fig. 2a. Code and phase geometry-free linear combinations (m). The sign of phase measurements is changed and shifted by A_0 (smooth line)

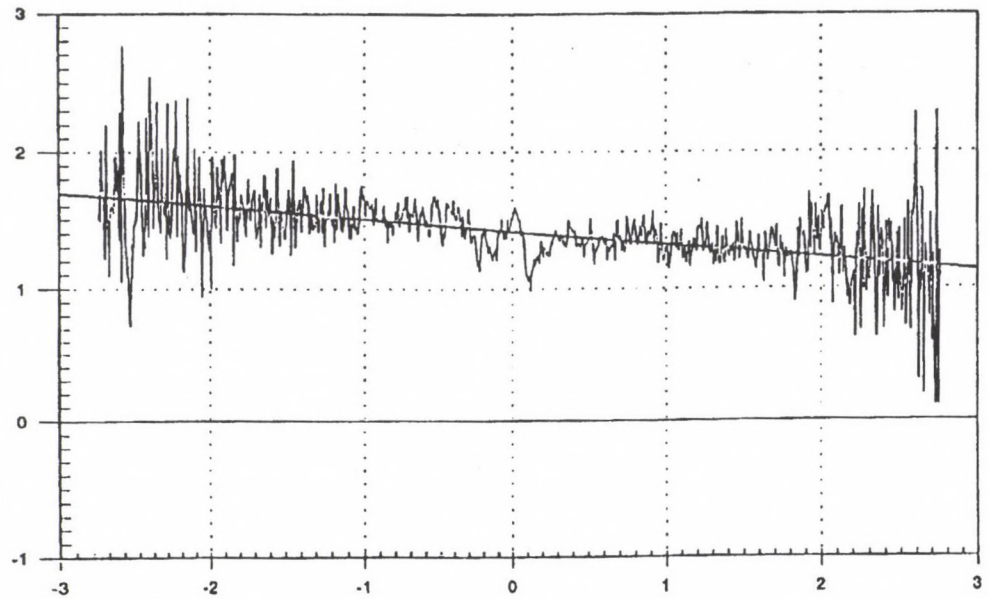


Fig. 2b. Geometry- and ionosphere-free combination together with the linear trend (m)

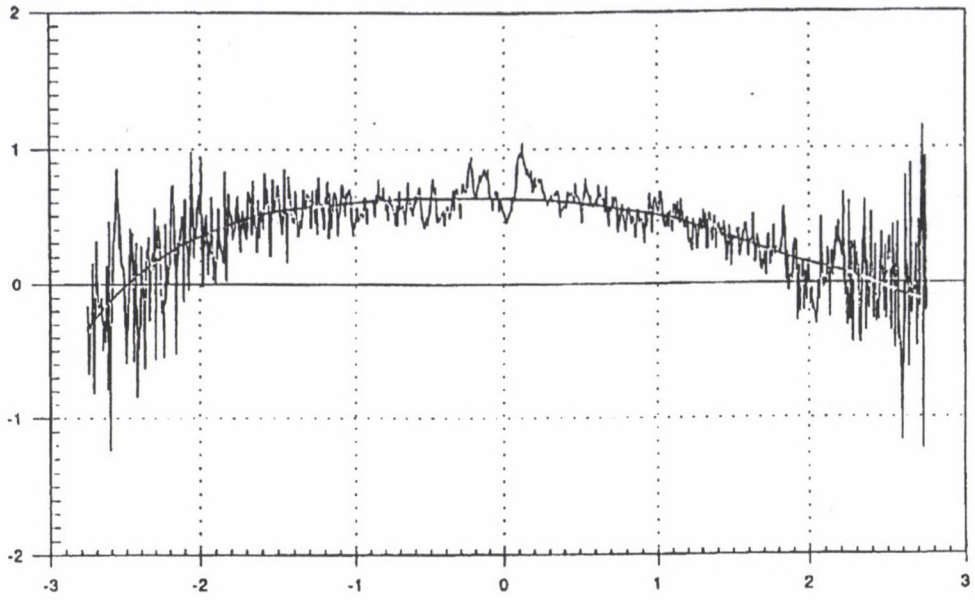


Fig. 2c. Optimum ionospheric combination with changed sign (smooth line) and geometry-free code combination after the linear trend is removed (m). Sopron 10.10.1994, SAT 12, cross-correlation TurboRogue measurements

the information from the original surface system. The origin of the variable ($\vartheta = 0$) is chosen at the mean observation time (t_m).

Before the least-squares adjustment of the single satellite observation it is reasonable to rearrange and simplify our observation equations as

$$v = \frac{1}{\sin E'} P(\vartheta) + CB - L$$

where v is the residual, $P(\vartheta)$ is the Legendre polynomial representation of the vertical ionospheric effects on the combination Eq. (4), CB is the sum of the receiver and satellite code biases and L is the observation term, all expressed in unit of meter.

The unknown coefficients of the Legendre polynomials and the CB bias can be determined by the Gauss-Markov modell of least-squares adjustment because the derivatives of the coefficients vary with $1/\sin E'$. The CB and the zero order cofficient can be separated effectively if there is a large change in the E' value. Therefore the observation of a high maximum elevation and full satellite pass are needed.

3. Investigation of the new method

For the practical investigations of the introduced method a program system was developed.

In the preparatory part the geometry-free linear combinations are subtracted from the binary data files in a satellite by satellite sense.

In the control part the observations below 20 degrees of elevation angle with respect to the observation site are deleted, the data gaps, the cycle-slips and the symmetry of the observations are controlled as well as the optimum ionospheric observations are determined.

Before the adjustment the parameters of the ionospheric points i.e. the geographic latitude and longitude, the local time as well as the elevation angle are computed. In the computations usually the $h_m = 400$ km single layer height is chosen (Fig. 1).

In the last step the CB bias and coefficients of the choosen Legendre polynomials are separated by the Gauss-Markov modell of least-squares adjustment.

One representative example of the test computations can be found in Table I where the standard deviation of unit weight (σ_o), the mean square error (mse) of the unknowns and the correlations between the CB and the other coefficients (corr) are summarised.

In the columns of the table the order of the Legendre polynomials vary from zero to five. In spite of the fact that σ_o values converge, the CB and the zero order cofficient (a_o) do not converge, moreover, in the last two cases the sign of a_o become false. After the linear solution the "mse" values become larger and larger, which is the consequence of the increasing high correlation between the unknowns. Even in the case of the zero order solution the correlation between the CB and a_o coefficient is very high, which shows the ill-condition of the basic modell.

Since the even order terms show a higher correlation the adjustments were repeated using only zero and odd order Legendre polynomials. The results are summarised in the lower part of Table I. In that case the solutions show a proper convergence and the CB and the odd order coefficient are practically uncorrelated.

For the comparison between the introduced and Multi-Station and Multi-Satellite Method three days of observations were processed in the DLR Remote Sensing Ground Station, Neustrelitz, Germany, by E Sardon and E Engler. According to the literature, their method applies the highest, second order surface polynomial to describe the vertical TEC values (Sardon et al. 1994, Sardon 1995). For the separation of the code biases and the coefficients of the polynomials a Kalman filter technique is applied, where the preliminary unknowns and their *a priori* variances have to be taken into account as well as multi-station and multi-satellite data are processed together.

The TurboRogue GPS observations in cross-correlation tracking mode recorded in Sopron were processed in the DLR together with the observations from fourteen IGS stations. The comparison of the estimated code biases (CB) for some satellites plus the GPS receiver can be found in Table II where the standard deviations of unit

Table I. Single Station and Single Satellite adjustment of SAT 12 (Sopron, 23.05.94 , P-code tracking) using different order of Legendre polynomials

	σ_0	CB	a_0	a_1	a_2	a_3	a_4	a_5
	±0.1147	−0.2410	1.8146					
mse		±0.0280	±0.0220					
corr			−0.9791					
	±0.0427	−0.3111	1.8696	0.3305				
mse		±0.0105	±0.0082	±0.0052				
corr			−0.9793	−0.1047				
	±0.0268	1.1564	1.2950	0.2910	1.6198			
mse		±0.0471	±0.0190	±0.0035	±0.0515			
corr			−0.9900	−0.3685	0.9902			
	±0.0246	0.7117	1.4694	0.0931	1.1314	−0.2069		
mse		±0.0603	±0.0239	±0.0180	±0.0601	±0.0186		
corr			−0.9946	−0.6291	0.9949	0.6890		
	±0.0126	6.0854	−1.8761	0.9599	2.9272	0.8144	−2.4188	
mse		±0.1290	±0.0788	±0.0221	±0.0540	±0.0255	±0.0562	
corr			−0.9962	0.9457	0.9078	0.9627	0.9713	
	±0.0112	14.785	−7.2517	−0.8186	5.9781	−1.5806	−6.2344	−1.7296
mse		±0.695	±0.4293	±0.1430	±0.2449	±0.1921	±0.3049	±0.1372
corr			0.9999	−0.9552	0.9965	−0.9603	−0.9992	−0.9862
	±0.1147	−0.2410	1.8146					
mse		±0.0280	±0.0220					
corr			−0.9791					
	±0.0427	−0.3111	1.8696	0.3305				
mse		±0.0105	±0.0082	±0.0052				
corr			−0.9793	−0.1047				
	±0.0283	−0.3056	1.8635	−0.1293		−0.4512		
mse		±0.0071	±0.0056	±0.0164		±0.0157		
corr			−0.9799	−0.0549		−0.0260		
	±0.0192	−0.3240	1.8788	2.1272		2.3950		1.1958
mse		±0.0048	±0.0038	±0.0779		±0.0981		±0.0412
corr			−0.9802	−0.0399		−0.0408		−0.0504

CB – the code bias (m),
σ_o – the standard deviation of unit weight (m),
mse – the mean square error of the unknowns (m),
corr – the correlation between the CB and the other coefficients

weight (σ_o) from the linear fit of the geometry- and ionosphere-free combinations are also presented. The data sets refer to the days September 22–24, 1995.

Table II. Comparison of code biases (CB), September 22–24, 1995. σ_o is the outer accuracy of GPS observations

SAT	DLR solution CB (ns)	SOPRON CB (ns)	solution σ_o (m)
	−7.143	−6.300	0.601
1	−7.398	−7.120	0.424
	−7.376	−7.120	0.331
	−8.779	−9.373	0.446
4	−9.043	−7.947	0.359
	−9.057	−9.650	0.287
	−8.228	−10.426	0.198
6	−8.556	−10.688	0.175
	−8.577	−9.935	0.195
	−5.917	−4.086	0.315
7	−6.192	−6.256	0.289
	−6.306	−4.479	0.301
	−6.600	−7.141	0.370
14	−6.857	−7.403	0.202
	−6.854	−9.116	0.219
	−6.314	−6.974	0.390
15	−6.550	−8.096	0.280
	−6.500	−7.016	0.281
	−5.193	−5.151	0.337
16	−5.430	−5.001	0.272
	−5.452	−4.892	0.297
	−5.983	−6.993	0.248
17	−6.482	−7.125	0.199
	−6.377	−7.692	0.195
	−6.728	−7.215	0.306
19	−6.977	−8.865	0.229
	−7.158	−6.456	0.235
	−6.378	−6.864	0.130
20	−6.602	−7.028	0.119
	−6.678	−5.968	0.115
	−7.586	−7.537	0.269
22	−8.035	−7.567	0.241
	−7.785	−8.600	0.243
	−10.099	−9.164	0.274
25	−10.417	−8.429	0.284
	−10.265	−7.807	0.265
	−8.853	−8.132	0.203
26	−9.089	−8.657	0.209
	−9.233	−8.675	0.236
	−8.673	−7.495	0.435
31	−8.982	−7.786	0.270
	−9.076	−7.230	0.228

In spite of the totally different estimation techniques and the significantly different data sets, the results — with a few exceptions — agree surprisingly well. The smoother change of the code biases from day to day estimated by the DLR solution is a consequence of the special feature of the Kalman filter technique, however the larger change estimated by our solution is in accordance with the σ_o values characterizing the outer accuracy of the GPS measurements. The larger differences between the two solutions may be the consequence of the preliminary unknowns, which play a special role in the case of the Kalman filter technique.

The Single Station and Single Satellite Method was used for the comparison of TurboRogue and Trimble SSi GPS receivers for ionospheric investigations as well (Bányai and Gianniou 1997). In the "near calibration" arrangement the method proved to be a good procedure to estimate the differences of the receiver code synchronization biases, the differences of the antenna phase center variations and for the comparison of different receiver noise level.

Conclusions

The introduced new Single Station and Single Satellite Method of GPS ionospheric data processing proved to be a proper device for both the TEC determination and calibration of "Rogue" type Geodetic GPS Receivers.

In comparison with the multi-satellite approach the following disadvantages can be established. For the optimal solution full and symmetric satellite pass is needed. The lower the change in the elevation angle of the observations the higher the correlation between the CB and the zero order coefficient of the polynomial, which provides a weeker solution. Moreover the lower observations are more noisy. The advantages of the new method are the relatively simple data processing and the possibility of higher order vertical TEC modelling by the odd order terms of Legendre polynomials.

Under normal circumstances the method can be used for ionospheric investigations and for the self calibration of the receiver and the observation site. It can also be useful as the prepocessing step of the multi-satellite and multi-station approach or can be implemented in other geodetic softwares for the modelling of ionospheric effects along a single satellite pass.

One of the most promising application may be the ionospheric comparison of GPS receivers in the near calibration arrangement, where the receiver noise, the difference of the code biases and the antenna phases can be estimated. Introducing the "travelling receiver" concept the GPS receivers of permanent ionospheric stations could be compared effectively.

Acknowledgements

This study was supported by the European project CEE-CIPA CT93-0130 and the State Scientific Research Fund (OTKA) through grants T4310 and A125. The help of Dr. Esther Sardon and Dr. Evelin Engler in the comparison with other method is greatly appreciated.

References

Bányai L, Eper-Pápai I 1996: *Bollettino di Geodesia e Scienze Affini*, 1, 9–24.
Bányai L, Gianniou M 1997: *ZfV*, 122, 136–142.
Bányai L, Kalmár J, Bencze P 1994: *Annali di Geofisica*, 37, 233–240.
Davies K, Hartmann G K, Leitinger R 1977: *J. Atmos. Terr. Phys.*, 9, 572–580.
Hartmann G K, Leitinger R 1984: *Bull. Geod.*, 58, 109–136.
Klobuchar L A, Doherty P H, Davies K 1994: In: Proc. Int. Beacon Satellite Symposium, L Kersley ed., Aberystwyth, UK, 1–4.
Lanyi G E, Roth T 1988: *Radio Sci.*, 23, 483–492.
Sardon E 1995: GPS satellite and receiver instrumental biases estimated using IGS data and data from Sopron. Technical report.
Sardon E, Rius A, Zarraoa N 1994: *Radio Sci.*, 29, 577–586.
van der Marel H, Georgiadou Y 1994: In: Proc. Int. Beacon Satellite Symposium, L Kersley ed., Aberystwyth, UK, 5–8.
Wanninger L, Sardon E, Warnant R 1994: In: Proc. Int. Beacon Satellite Symposium, L Kersley ed., Aberystwyth, UK, 13–16.
Wanninger L 1994: Der Einfluss der Ionosphäre auf die Positionierung mit GPS. Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, No. 201.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 417–425 (1997)

DIFFERENCE BETWEEN THE LENGTH OF THE TRUE PROPAGATION PATH AND THE GEOMETRICAL PATH OF RADIO WAVES TRANSMITTED BY GPS SATELLITES

P BENCZE[1], F MÄRCZ[1], J SZENDRŐI[1]

Formerly the propagation of radio waves transmitted by GPS satellites in a stratified ionosphere represented by ionospheric models has been studied. For the investigation of the propagation in an ionosphere containing also irregularities, it was necessary to increase the resolving power of the modelling. It has been found that vertical steps of 5 km have to be used in the computations to a height of 300 km, then steps of 50 km till 1000 km and steps of 500 km in the plasmasphere are necessary. The change of the length difference between the true path and the straight line (geometrical) path is discussed.

Keywords: GPS satellites; ionospheric effect; modelling of radio wave propagation; positioning

Introduction

The effect of the ionosphere on positioning was quite different during the period, when the NNSS system was mainly used, from the present situation, when the GPS satellites are operated. Though, the GPS satellites are using frequencies (1575.42 MHz and 1227.60 MHZ) much higher, than the frequencies transmitted by the NNSS satellites (400 MHz, as well as 150 MHz), and this means less ionospheric effect, at the same time the propagation path in case of the radio waves transmitted by the GPS satellites is much longer (height of the orbit 20200 km), than in case of the NNSS signals (height of the orbit 1000 km). Thus, the use of GPS satellites has an advantage by decreasing the ionospheric effect transmitting at higher frequencies, but has also a disadvantage having a longer propagation path.

It is true that the simultaneous use of the two frequencies reduces greatly the influence of the ionosphere. However, on the one hand in the form of the refractive index expanded in series the terms of higher order than the second order are neglected, on the other hand integrating the refractive index, the integration is carried out along the geometrical path and not along the true propagation path (Hoffmann-Wellenhof et al. 1992). The importance of the difference between the true path and the geometrical path becomes obvious, if the height variation of the electron density, especially its temporal and spatial changes are considered. Concerning the ionospheric effect, namely spatial changes in the electron density of the ionosphere can cause the largest errors in positioning (Bányai et al. 1993, 1994). For the reduction of the above mentioned difficulties, two procedures can be suggested. The first procedure would be the consideration of the variability of the

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

1217-8977/97/$ 5.00

ionosphere at the planning of GPS measurements, carrying out observations in periods of minimum ionospheric variability. The other solution would be to develop methods for ionospheric correction, including terms characterizing the temporal and spatial variability of the ionosphere.

Modelling of the propagation of radio waves

The modelling of the propagation of radio waves transmitted by GPS satellites is necessary, if we want to determine the true path of the radio waves. The modelling consists of two parts (Bencze and Kovács 1994). The first part includes the calculation of the geographical (geocentric) coordinates of the ends of the consecutive steps along the true path, for which the change of the zenith angle is computed. The geographical coordinates must be transformed first to geocentric coordinates. Then, the geometrical distance of the satellite from the observation site is determined for the given zenith angle of the satellite. Knowing the geometrical distance of the satellite from the observation site, the geocentric latitude and longitude of the satellite can be calculated. If the distance of two consecutive points of the true path from the Earth's center are known, the inner angle at the end of a step is computed. Finally, the geocentric latitude and longitude of the end of a step can be found from the triangle, the angles of which are the geocentric latitude and longitude, respectively, the complementary angle of the zenith angle of the previous point and the inner angle mentioned above.

In the second part of the modelling, the change of the initial zenith angle (at the observation site) at the end of the steps is determined. For the determination of the change of the zenith angle the refractive index given by the Appleton-Lassen formula of the magneto-ionic theory is used (Alpert 1974). As it is known, the plasma-frequency, the gyrofrequency appear in this formula, which must be calculated for the frequencies transmitted by the satellite. The computation of the plasma frequency needs the knowledge of the distribution of the electron density with height and with time. For this purpose ionospheric models can be used (IRI 9). The intensity of the geomagnetic field required for the determination of the gyrofrequency can be found by means of subroutins given in the computer program of ionospheric models. The angle between the geomagnetic field and the direction of propagation necessary for the calculation of the transversal and longitudinal components of the gyrofrequency is determined by the direction cosines.

More accurate results can be obtained, if the formula of the refractive index referring to the group velocity is used for the computation of the change of the zenith angle due to refraction of radio waves in the ionosphere (Budden 1985).

However, the electron density can be found by ionospheric models to a height of 1000 km. In the plasmasphere an exponential decrease of the electron density with height can be assumed because of the lack of a detailed model similar to that of the ionosphere. The electron density distribution in the plasmasphere used here refers to the equilibrium state of the plasmasphere (Bauer 1973, also Hartmann and Leitinger 1984, Klobuchar and Leitinger 1993). The exponential decrease of the electron density based on topside sounding and *in situ* plasma measurements

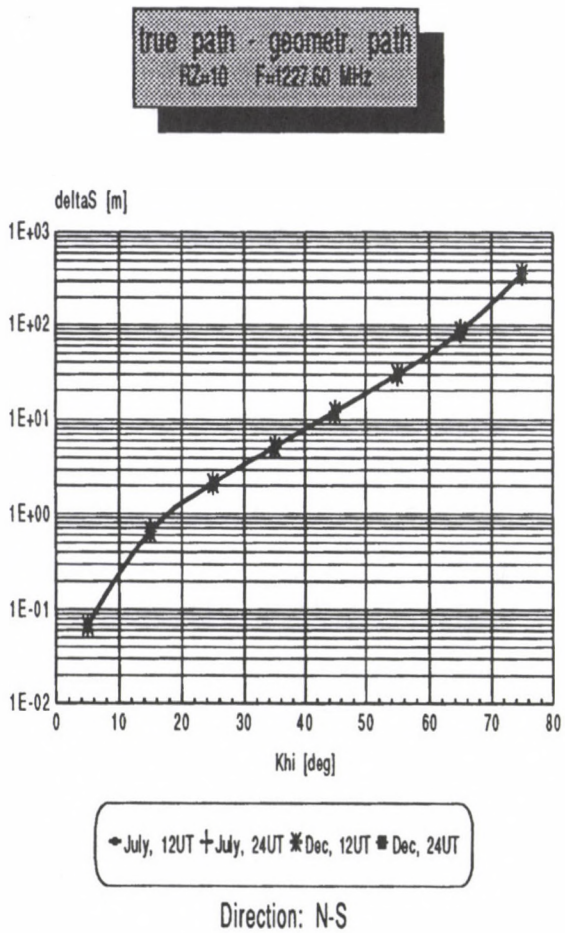


Fig. 1. Difference between the true path and the geometrical path as a function of the zenith angle of the satellite for N-S direction and low solar activity in case of the lower frequency

by satellites is expressed by an exponent of negative sign proportional to height and inversely proportional to the plasma scale height. It is to be noted that the plasmasphere reaches seldom equilibrium state. The electron density decreases, namely, during geomagnetically disturbed periods due to the dayside erosion of the magnetosphere and thus in consequence of the deflation of the magnetic flux tubes connected to the interplanetary magnetic field. During the decay of the geomagnetic activity, the geomagnetic field lines disconnected from the interplanetary field lines and reconnected with field lines of the conjugate hemisphere are refilled with plasma from the ionosphere. However, the refilling can mostly not be finished before a new disturbance begins especially during the years of maximum solar activity. Consequently, the formula used for the determination of the distribution of electron

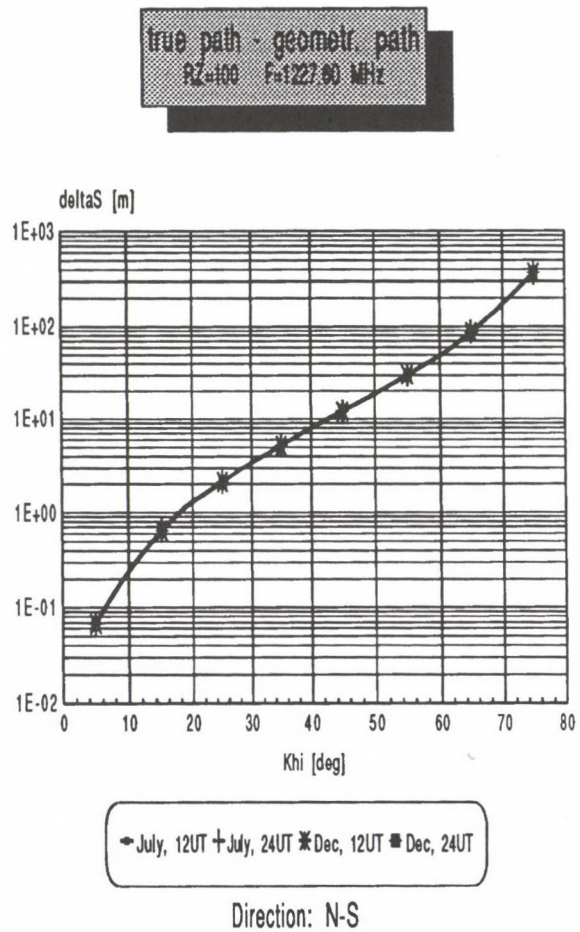


Fig. 2. Difference between the true path and the geometrical path as a function of the zenith angle of the satellite for N-S direction and high solar activity in case of the lower frequency

density in the plasmasphere describes the state of the plasmasphere characterized by maximum electron density. Thus, this state corresponds also to maximum influence of the plasmaspheric electron density on the propagation of radio waves.

Results and analysis

The results of the modelling are presented partly in Table I in numerical form, partly by Figs 1–4. The computations were carried out for the directions S-N, N-S, W-E, E-W and for July and December midday, as well as midnight. It has been assumed that these cases include the most characteristic states. In Table I the differences between the true path and the geometrical path are given for those cases, which correspond to maximum and minimum departures in the direction

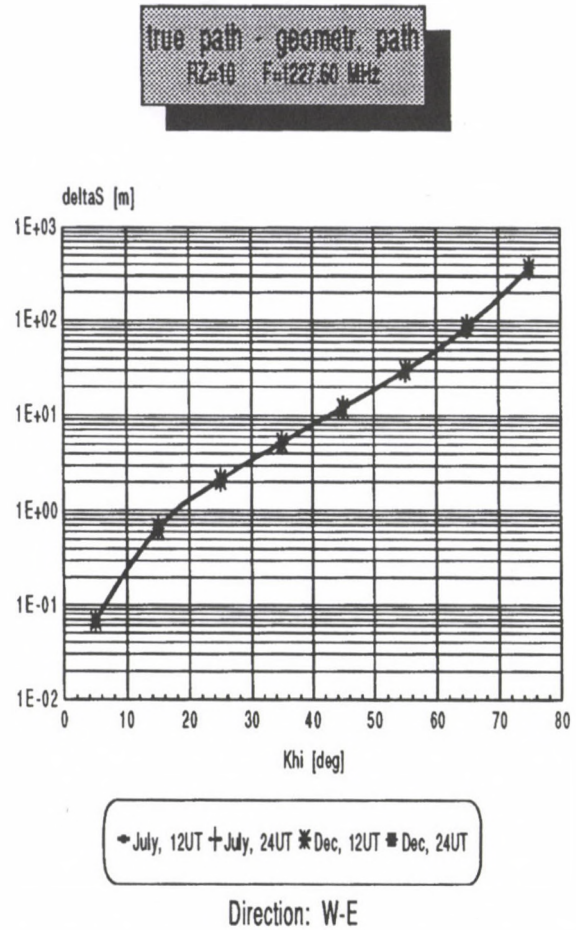


Fig. 3. Difference between the true path and the geometrical path as a function of the zenith angle of the satellite for W-E direction and low solar activity in case of the lower frequency

in question, as well as for different zenith angles and different solar activity. For the illustration of the effect of the frequency, data related to both frequencies are presented. It can be seen that the value of the maximum departure is of the order of 365.78 m at the zenith angle of 75°, its variation in the different directions reaches values of the order of some cm in case of low solar activity. In case of high solar activity, the variation of the difference between the true path and the geometrial path attains values of the same order in the different directions at this zenith angle, but the change of the differences is definitely larger.

However, from the point of view of positioning, the departure of the differences between the true path and the geometrical path referring to one of the frequencies from that obtained for the other is the most interesting. The largest departures

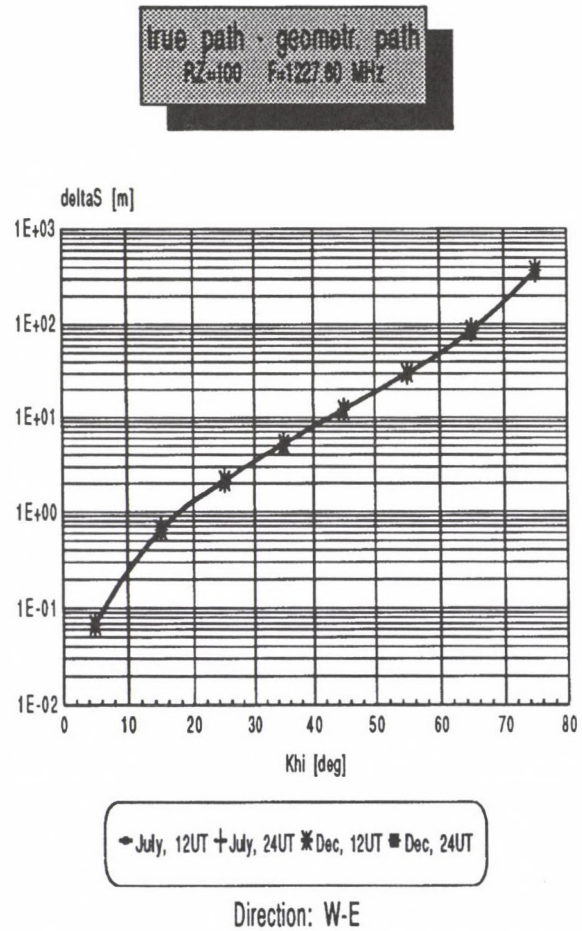


Fig. 4. Difference between the true path and the geometrical path as a function of the zenith angle of the satellite for W-E direction and high solar activity in case of the lower frequency

are found at the zenith angle 75° reaching values of the order of 1.5 cm in the direction N-S and 0.8 cm in the E-W direction. The difference between the true path and the geometrical path is smaller in case of the higher frequency (1575.42 MHz), than in case of the lower frequency (1227.60 MHz), if the S-N, N-S directions are considered, however, somewhat greater in case of the higher frequency, than in case of the lower frequency for the directions W-E, E-W. In Table I, values of the differences are given to a limit of thenth of mm.

In Figs 1–4 the differences between the true path and the geometrical path are plotted as a function of the zenith angle for different times, different directions and for the smaller frequency. Those cases were selected, where this difference was maximum, that is the N-S direction and the W-E direction for low (RZ=10) and

for high (RZ=100) solar activity. Plotting the differences on the vertical axis in logarithmic scale, the curves can be approximated by a nearly linear function, i.e. the relation between the difference true path — geometrical path (ΔS) and the zenith angle can be given by the equation

$$\Delta S = 0.0811 \exp(0.11132\chi)$$

where χ is the zenith angle.

Another relation describing the connection of the difference true path — geometrical path obtained at the higher frequency with the same difference determined for the lower frequency

$$\Delta S(F1) - \Delta S(F2) = -2.9319 \cdot 10^{-6} \exp(0.08504\chi)\,[m]$$

where F1 =1575.42 MHz and F2=1227.60 MHz.

Considering the numerical values, it is clear that the curves referring to different times in the same figure can not be distinguished.

Conclusions

From the above discribed results and their analysis, the following conclusions can be drawn.

The departure of the maximum values from the minimum ones of the difference between the true path and the geometrical path of the radio waves transmitted by GPS satellites attains values of several cms at each of the frequencies as shown by Table I.

The deviation of the maximum values from the minimum values of these differences is larger at the lower frequency, than at the higher frequency in case of S-N, N-S propagation, but somewhat greater at the higher frequency, than at the lower frequency in case of W-E, E-W propagation.

The magnitude of the difference between the true path and the geometrical path increases with increasing zenith angle at both frequencies and in all directions of propagation.

The departure of the maximum values from the minimum values of these differences is greater at high solar activity than at low solar activity at both frequencies and in all directions of propagation.

The deviation of the differences between the true path and the geometrical path obtained for the one frequency from that determined for the other frequency reaches values of the order of 1.5 cm in the direction N-S and 0.8 cm in the E-W direction.

From the point of view of positioning, the accuracy can be increased on the one hand by avoiding geomagnetically disturbed periods. On the other hand by the development of correction procedures, which include terms related to the temporal and spatial variability of the ionosphere.

Table I

x	F1		F2	
	RZ=10	RZ = 100	RZ=10	RZ = 100
		maximum values		
	direction: N-S	time: 12 07	direction: N-S	time: 12 07
5	0.0674	0.0674	0.0674	0.0674
15	0.6509	0.6510	0.6510	0.6510
25	2.0948	2.0949	2.0949	2.0951
35	5.1966	5.1969	5.1968	5.1974
45	12.1510	12.1517	12.1515	12.1527
55	29.9191	29.9207	29.9201	29.9228
65	87.1087	87.1124	87.1111	87.1174
75	365.7848	365.7964	365.7923	365.8122
		minimum values		
	direction: S-N	time: 24 12	direction: S-N	time: 24 12
5	0.0674	0.0674	0.0674	0.0674
15	0.6509	0.6510	0.6510	0.6510
25	2.0948	2.0949	2.0949	2.0951
35	5.1966	5.1969	5.1968	5.1973
45	12.1510	12.1517	12.1515	12.1526
55	29.9191	29.9205	29.9200	29.9225
65	87.1083	87.1116	87.1105	87.1160
75	365.7822	365.7907	365.7877	365.8020
		maximum values		
	direction: W-E	time: 24 12	direction: W-E	time: 24 12
5	0.0674	0.0674	0.0674	0.0674
15	0.6509	0.6509	0.6509	0.6509
25	2.0947	2.0947	2.0947	2.0947
35	5.1963	5.1963	5.1963	5.1963
45	12.1504	12.1504	12.1504	12.1504
55	29.9177	29.9177	29.9177	29.9177
65	87.1051	87.1051	87.1051	87.1051
75	365.7735	365.7735	365.7734	365.7734
		minimum values		
	direction: E-W	time: 12 07	direction: E-W	time: 12 07
5	0.0674	0.0674	0.0674	0.0674
15	0.6509	0.6510	0.6509	0.6510
25	2.0948	2.0949	2.0948	2.0950
35	5.1965	5.1967	5.1966	5.1969
45	12.1506	12.1509	12.1508	12.1513
55	29.9179	29.9182	29.9181	29.9186
65	87.1047	87.1043	87.1044	87.1038
75	365.7678	365.7620	365.7640	365.7544

Acknowledgements

This study was supported by the State Commission for Technical Development through the Hungarian Space Office by the grant T010. It is also supported by the State Scientific Foundation OTKA project No. T 02311.

References

Alpert Ya L 1974: Radio Wave Propagation and the Ionosphere, Vol 2. Consultants Bureau, New York

Bányai L, Kalmár J, Bencze P 1993: In: PRIME Studies with Emphasis on TEC and Topside Modelling. Wissenschaftlicher Bericht No 2, Teil 2. Inst. f. Meteorologie und Geophysik, Universität Graz, 419.

Bányai L, Kalmár J, Bencze P 1994: *Annali di Geofisica*, 37, 233–240.

Bauer S J 1973: Physics of Planetary Ionospheres. Springer Verlag, Berlin

Bencze P, Kovács K 1994: In: Proc. Int. Beacon Satellite Symposium, L Kersley ed., University of Wales, Aberystwyth, UK, 362–365.

Budden K G 1985: The Propagation of Radio Waves. Cambridge University Press, Cambridge

Hartmann G K, Leitinger R 1984: *Bull. Geod.*, 58, 109–136.

Hoffmann-Wellenhof B, Lichtenegger H, Collins J 1992: Global Positioning System. Theory and Praxis. Springer Verlag, Wien

IRI 9, International Reference Ionosphere. National Space Data Center, World Data Center-A, Greenbelt, USA.

Klobuchar J A, Leitinger R 1993: In: PRIME Studies with Emphasis on TEC and Topside Modelling. Wissenschaftlicher Bericht No 2, Teil 1. Inst. f. Meteorologie und Geophysik, Universität Graz, 13–20.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 427–437 (1997)

MODELLING OF THE PROPAGATION OF RADIO WAVES TRANSMITTED BY GPS SATELLITES IN AN IONOSPHERE WITH IRREGULARITIES

P BENCZE[1], J SZENDRŐI[1], K KOVÁCS[1]

The results concerning the modelling of the propagation of radio waves transmitted by GPS satellites in the presence of sporadic E layers are described. The greatest deviations of the difference between the true path and the geometrical path found for the higher frequency from the difference true path-geometrical path obtained for the lower frequency attain values of the order of 26 cm at the zenith angle 75° of the satellite in the N-S direction (summer, night and low solar activity conditions) and about 7 cm in the W-E direction (summer and winter night, as well as low solar activity conditions).

Keywords: GPS satellites; ionospheric sporadic E layers; modelling of radio wave propagation; positioning

Introduction

The modelling of the propagation of radio waves in a homogeneous ionosphere applied to the radio waves transmitted by GPS satellites have shown that the difference between the deviations of the true path-geometrical path found for the higher frequency from that obtained for the lower frequency can attain values of the order of cm (Bencze et al. 1997). This order of magnitude is already important in case of positioning. However, the consideration of the ionosphere is not only important from the point of view of its temporal and spatial variation — as it has been emphasized in the former paper, but also from the point of view of the irregularities occurring in the ionosphere.

The irregularities occurring in the ionosphere can be devided into two groups. One of the groups include the sporadic layers, which-contrary to the thick or steady layers, as the E, F1 and F2 layers, are of sporadic occurrence and can be observed in the E region. Therefore, they are called sporadic E layers, or briefly Es layers. Es layers occur at all latitudes, though they are due to different processes. The other group is formed by field aligned irregularities; that is by irregularities elongated along the geomagnetic field lines having diameters of the order of 100 m perpendicular to the field lines, but extension of the order of 10 km along the field lines. This type of irregularity appears mostly at low and high latitudes.

Modelling the propagation of radio waves transmitted by GPS satellites in an ionosphere containing irregularities, first the effect of sporadic layers has been investigated because of their more frequent occurrence at mid-latitudes. Sporadic layers can be observed all day. Considering their height distribution, sporadic layers occur

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

1217-8977/97/$ 5.00

in the whole E region. They can be observed for several hours in course of which their height can decrease due to the process creating the sporadic layers. According to the wind-shear theory of mid-latitude sporadic E, Es layers are produced by the vertical shear of the horizontal wind, which can be attributed to the propagation of atmospheric gravity waves. The formation of Es layers is enabled by the force affecting the motion of ions carried with the wind in the geomagnetic field. Thus, the ions are diverted upwards, where the eastward component of the horizontal wind-varying sinusoidally corresponding to an atmospheric wave — is decreasing and downwards, where the westward wind is increasing. Consequently the diverted ions form a layer, where the zonal wind changes direction. In this paper only the zonal component of the wind shall be considered, since the effect of the vertical shear of the zonal component is much greater in the height range important from the point of view of Es layers, than the vertical shear of the meridional component. The long life time of Es layers can be explained by ions, which have a very small recombination coefficient. It has been shown that ions of meteoritic origin have recombination coefficients demanded by the long lifetime.

Modelling of the effect of sporadic E layers on the propagation of radio waves

For the modelling of the effect of sporadic E layers on the propagation of radio waves first a model of Es layers is necessary, giving the ion density distribution in the Es layer. Such model of a mid-latitude Es layer can be constructed by assuming the sinusoidal variation of the zonal component of the horizontal wind corresponding to an atmospheric wave. The characteristics of the Es layer, its maximum ion density $n_{i\,\max}$ and height depend on the change of the zonal wind speed with height. The distribution of ion density n_i with height in a sporadic E layer can be expressed by the equation (Axford 1963).

$$n_i = n_{i\,\max} \exp\left(\frac{XU_{yo}}{Q_i D_a} \cdot \frac{\lambda_z}{2\pi} \cos\frac{2\pi}{\lambda_z} z\right)$$

where $n_{i\,\max}$ is the maximum ion density, $X = B_x/B = \cos I$ the ratio of the north component of the geomagnetic field to the total intensity, U_{yo} stands for the amplitude of the atmospheric wave, $Q_i = \nu_{in}/\omega_i$ the ratio of the ion-neutral collision frequency ν_{in} to the ion gyrofrequency ω_i, D_a is the ambipolar diffusion coefficient

$$D_a = \frac{2kT_i}{m_i \nu_{in}} .$$

Here k is Boltzmann's constant, T_i stands for the ion temperature, which equals the neutral temperature in this height range, m_i is the ion mass. The above quantities $(X, \nu_{in}, \omega_i, T_i, m_i)$ can be determined by means of atmospheric models. Furthermore, λ_z and z are the vertical wavelength of the atmospheric wave and the vertical distance reckoned from the center of the Es layer upwards and downwards, respectively.

For the determination of the distribution of the ion density in the Es layer the computation of the amplitude and vertical wavelength of the atmospheric wave is needed, which can be carried out by substituting for U_{yo} in the equation of the ion density distribution the expression of U_{yo} obtained from the equation of the atmospheric wave. Thus, the relations

$$U_{yo} = \left[\frac{Q_i D_a}{\cos I} \cdot \ln\left(\frac{n_{i\,\max}}{n_{io}}\right)\left(\frac{dU_y}{dz}\right)_{\max}\right]^{\frac{1}{2}}$$

and

$$\lambda_z = \left[\frac{4\pi^2 Q_i D_a}{\cos I} \ln\left(\frac{n_{i\,\max}}{n_{io}}\right) \frac{1}{\left(\frac{dU_y}{dz}\right)_{\max}}\right]^{\frac{1}{2}}$$

are obtained. As it can be seen, the ion density distribution depends on the amplitude and vertical wavelength of the atmospheric wave creating the Es layer. On the other hand, both quantities are connected with the vertical shear of the zonal wind dU_y/dz. The wind shear is increased by the increase of the amplitude of the atmospheric wave and vice versa. Similarly, the wind-shear increases with decreasing vertical wavelength of the atmospheric wave. Thus, the ion density distribution is determined both by the amplitude and the vertical wavelength of the atmospheric wave forming the factor of the cosine function in the argument of the exponential function. This would mean that the ion density distribution does not depend on the wind-shear. However, the vertical wavelength is included in the argument of the cosine function, consequently controlling the shape of the Es layer. The thickness of the Es layer is the smaller, the smaller the vertical wavelength, i.e. the greater is the wind-shear. Investigations have shown that the thickness of Es layers indicate a seasonal variation showing summer minimum and a solar cycle variation, where the minimum values appear in the years of maximum solar activity.

Considering the other quantities appearing in the ion density distribution of Es layers, $n_{i\,\max}$ can be computed from the blanketing frequency f_bEs obtained by ionosonde measurements. n_{io} represents the background ion density, that is the ion density in the absence of wind-shear, which is computed by means of ionospheric models for the given geographical and geophysical situation (time, geographical latitude and longitude, solar and geomagnetic activity).

In course of the computations carried out for the modelling of the effect of Es layers it has been assumed that there are more Es layers one above the other the height difference between them being 5 km. This assumption is based on the conditions that more sporadic E layers can be formed at points separated by one vertical wavelength, from which only the lowest can be observed by vertical sounding of the ionosphere (ionosonde). The other layers are namely blanketed by the lowest layer. Thus, taking six Es layers located between 95 and 120 km separated in height by 5 km, the greatest effect has been modelled, which can occur at all.

The radio waves transmitted by GPS satellites traverse also the plasmasphere. For the determination of the effect of the plasmasphere on the propagation the same

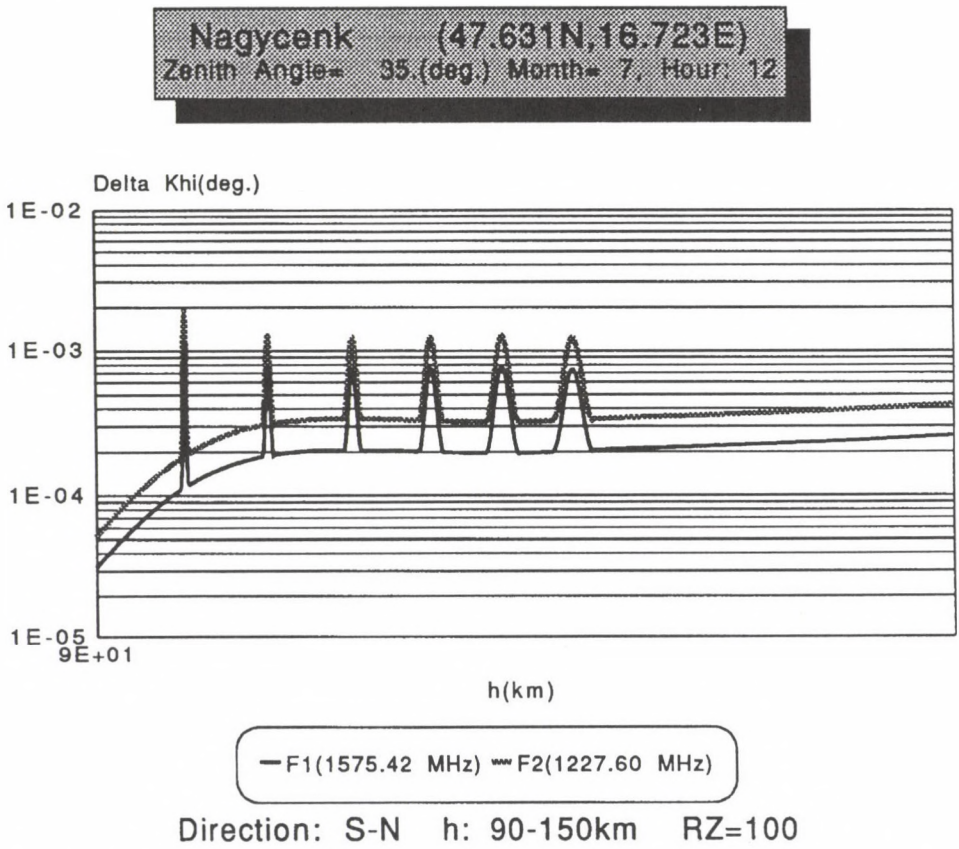


Fig. 1. Change of the zenith angle of the satellite due to a step along the propagation path referring to summer day conditions in S-N direction and for high solar activity in case of both frequencies

procedure has been used, as that described in the paper Bencze et al. (1997) and based on Bauer (1973).

Results and analysis

The main results of the modelling of the radio wave propagation in the presence of Es layers are given in numerical form in Table I. Table I presents the differences between the true path and the geometrical path for both frequencies in case of different zenith angles of the satellite and referring to low and high solar activity, respectively. The data are grouped according to maximum and minimum values and the appropriate directions. The data indicate that these differences are greater in case of the lower frequency except the W-E direction. The Table shows that the maximum difference true path-geometrical path is obtained at the greatest zenith angle of the satellite 75° considered here and its value attains the value 366.19 m.

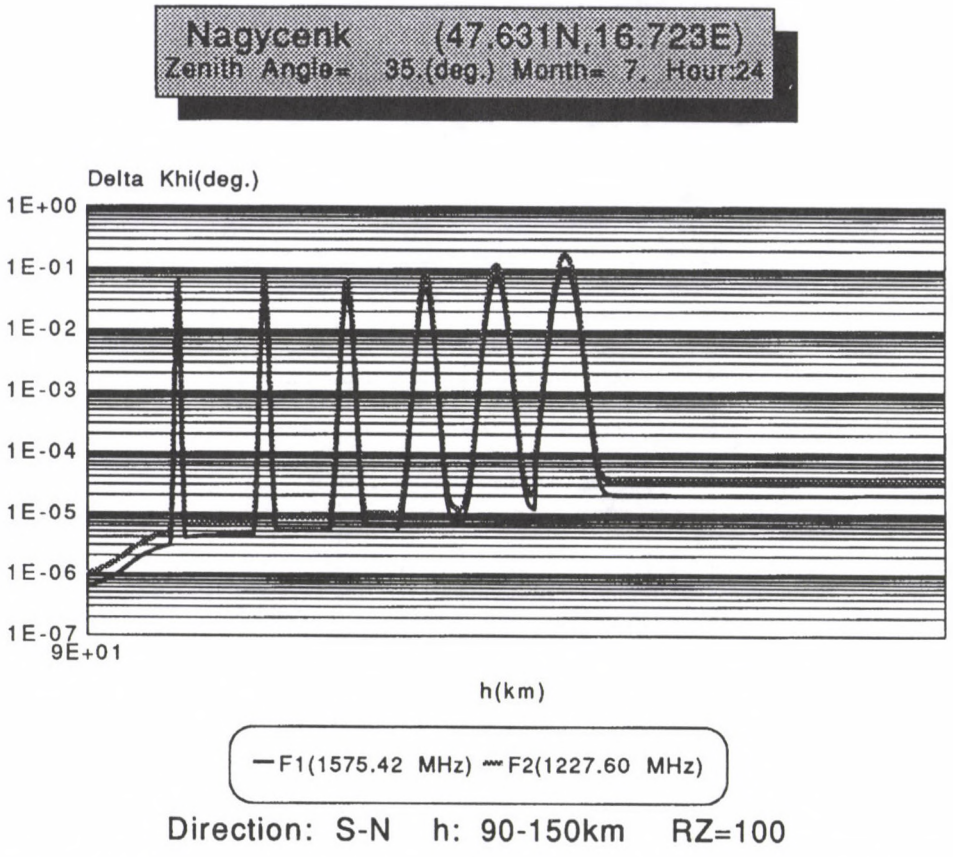


Fig. 2. Change of the zenith angle of the satellite due to a step along the propagation path referring to summer night conditions in S-N direction and for high solar activity in case of both frequencies

The variation with solar activity indicates that this difference is generally greater by the order of 10 cm in case of low solar activity.

Considering the more important quantity, that is the deviation of the difference between the true path and the geometrical path determined for the higher frequency from the difference true path-geometrical path obtained for the lower frequency, this deviation can amount to 26 cm at the zenith angle 75° in case of N-S propagation for summer night conditions and low solar activity, 7 cm in the W-E direction for winter night and low solar activity.

In Figs 1 and 2 the changes of the zenith angle of the satellite due to a step along the propagation path are presented referring to summer day and night conditions in N-S direction, as well as high solar activity. The Figures show the sporadic E layers emerging from the background of the zenith angle variation. An interesting picture is provided by comparing the true path of the radio waves in a homogeneous

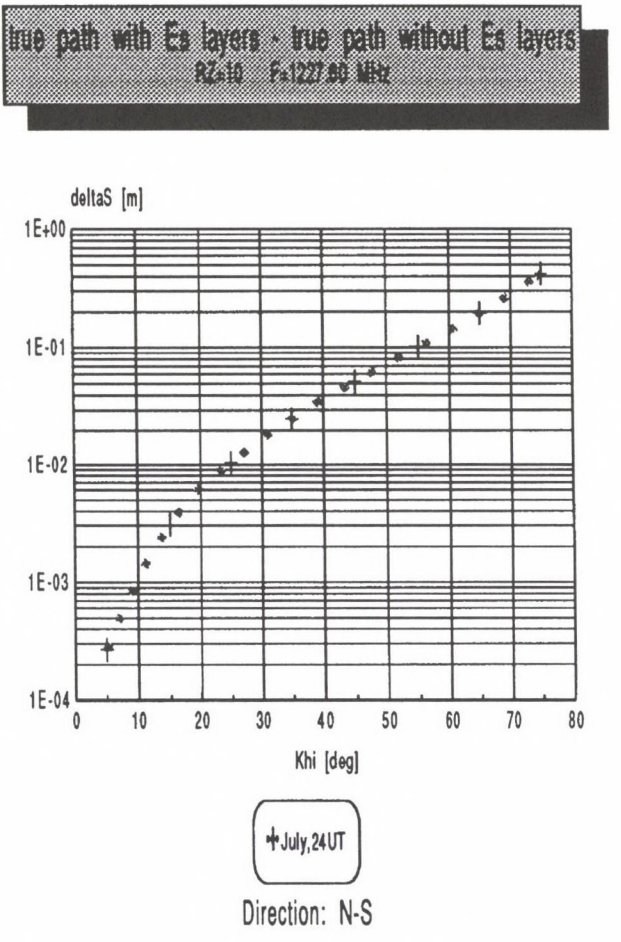


Fig. 3. The difference between the true paths of radio waves in an ionosphere with Es layers and that in a homogeneous ionosphere for summer night, in N-S direction, for low solar activity referring to different zenith angles of the satellite in case of the lower frequency

ionosphere with the true path of radio waves in an ionosphere containing Es layers (Figs 3–6). Figure 3 shows the difference between the true path of radio waves in an ionosphere with Es layers and that in a homogeneous ionosphere for summer night, in N-S direction and for low solar activity referring to different zenith angles of the satellite, in case of the lower frequency. This curve indicates a smooth variation with the zenith angle. In Fig. 4 the change of this difference is presented for summer night, in N-S direction, referring to different zenith angles of the satellite in case of the lower frequency, but for high solar activity. This curve shows that the differences are smaller at all zenith angles of the satellite in case of high solar activity than for low solar activity. The change of the difference in question is illustrated for summer night and low solar activity referring to different zenith angles of the satellite in case

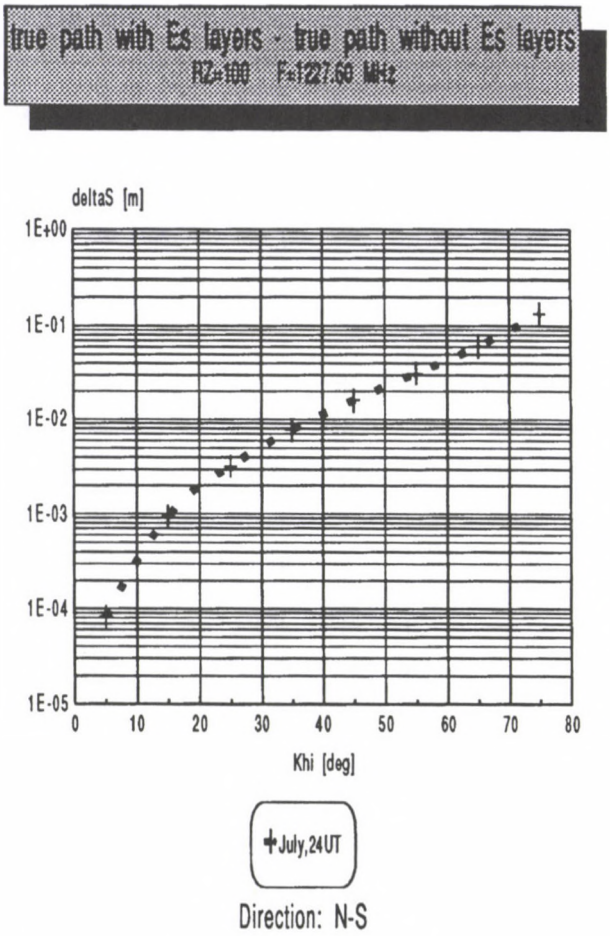


Fig. 4. The difference between the true paths of radio waves in an ionosphere with Es layers and that in a homogeneous ionosphere for summer night, in N-S direction, for high solar activity referring to different zenith angles of the satellite in case of the lower frequency

of the lower frequency, but for W-E direction by Fig. 5. The shape of the curve is strange the differences increasing with increasing zenith angle of the satellite to about 40°, then decreasing with the further growth of the zenith angle to about 65° and increasing again with increasing zenith angle. The same behaviour can be seen in Fig. 6, referring to the same conditions, but for high solar activity, however the values are smaller in case of high solar activity, than in case of low solar activity.

Conclusions

The modelling of the propagation of radio waves transmitted by GPS satellites in an ionosphere in case of the presence of sporadic E layers resulted in the following main conclusions.

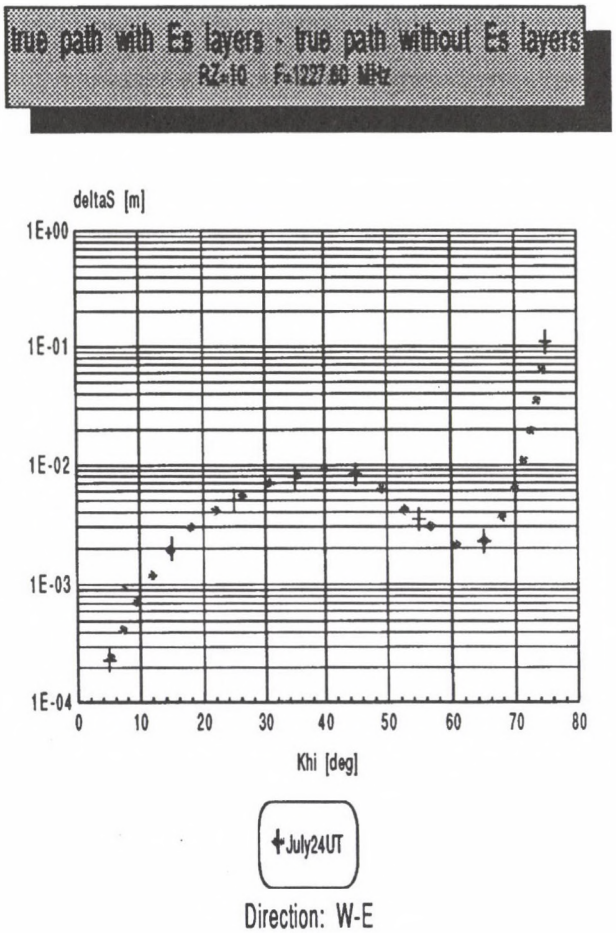


Fig. 5. The difference between the true paths of radio waves in an ionosphere with Es layers and that in a homogeneous ionosphere for summer night, in W-E direction, for low solar activity referring to different zenith angles of the satellite in case of the lower frequency

The ionosphere containing Es layers can increase the deviation of the difference between the true path and geometrical path referring to the higher frequency from the difference true path-geometrical path obtained for the lower frequency. This circumstance is important from the point of view of the definition geomagnetically disturbed period. Geomagnetically disturbed periods are not needed namely, for the occurrence of Es layers at mid-latitudes. For the existence of Es layers appropriate dynamical conditons are necessary, which produce the vertical shear of the horizontal wind.

The effect of Es layers on the difference between the true path and the geometrical path depends on the frequency and the direction, on time and the zenith angle of the satellite, furthermore on solar activity, as it can be seen in Table I.

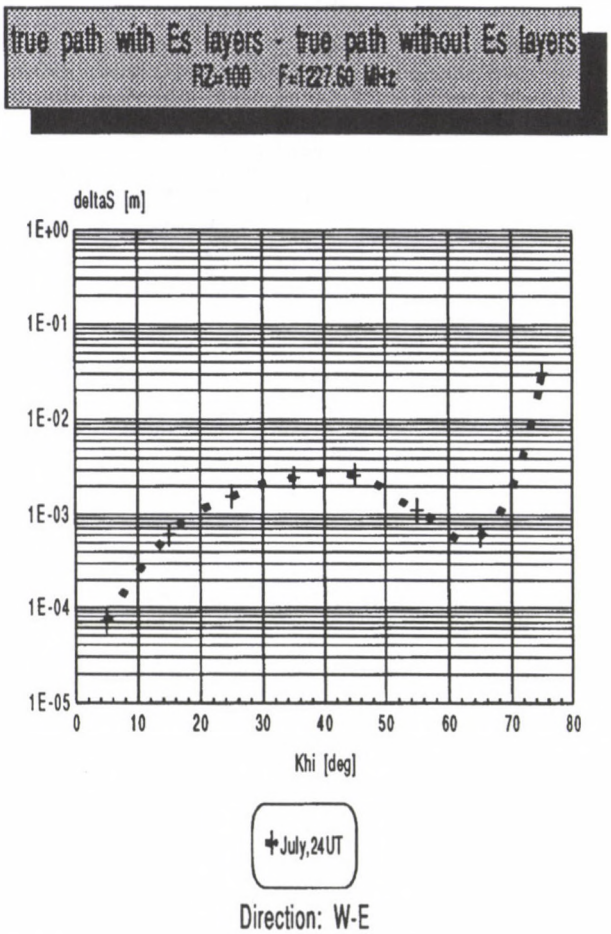


Fig. 6. The difference between the true paths of radio waves in an ionosphere with Es layers and that in a homogeneous ionosphere for summer night, in W-E direction, for high solar activity referring to different zenith angles of the satellite in case of the lower frequency

The magnitude of the difference true path-geometrical path decreases with solar activity.

The deviation of the difference between the true path and the geometrical path obtained for the higher frequency from the difference true path-geometrical path referring to the lower frequency amounts to 26 cm in the N-S direction at low solar activity and to 8 cm at high solar activity ($\chi = 75°$). These values are 7 cm and 2 cm, respectively in case of W-E propagation.

From the point of view of positioning, thus it is not only necessary to take into account the level of geomagnetic activity, but also to consider the presence of Es layers.

Table I

	F1		F2	
	RZ=10	RZ=100	RZ=10	R=100
		maximum values		
χ	N-S direction	date: 0724	N-S direction	date: 0724
5	0.0675	0.0674	0.0676	0.0674
15	0.6520	0.6513	0.6540	0.6519
25	2.0985	2.0959	2.1051	2.0979
35	5.2056	5.1992	5.2215	5.2041
45	12.1696	12.1564	12.2023	12.1665
55	29.9549	29.9293	30.0180	29.9490
65	87.1776	87.1279	87.2993	87.1664
75	365.9317	365.8244	366.1917	365.9088
		minimum values		
χ	N-S direction	date: 1212	N-S direction	date: 1212
5	0.0674	0.0674	0.0674	0.0674
15	0.6509	0.6509	0.6509	0.6510
25	2.0948	2.0948	2.0948	2.0948
35	5.1965	5.1966	5.1966	5.1968
45	12.1508	12.1509	12.1511	12.1514
55	29.9186	29.9189	29.9192	29.9199
65	87.1073	87.1079	87.1088	87.1104
75	365.7805	365.7821	365.7851	365.7893
		maximum values		
χ	W-E direction	date: 1224	W-E direction	date: 1224
5	0.0675	0.0674	0.0676	0.0674
15	0.6516	0.6511	0.6529	0.6515
25	2.0966	2.0953	2.0997	2.0963
35	5.1993	5.1973	5.2043	5.1988
45	12.1535	12.1514	12.1588	12.1530
55	29.9190	29.9181	29.9212	29.9188
65	87.1059	87.1053	87.1075	87.1057
75	365.8120	365.7847	365.8859	365.8051
		minimum values		
χ	W-E direction	date: 0712	W-E direction	date: 0712
5	0.0674	0.0674	0.0674	0.0674
15	0.6509	0.6510	0.6510	0.6510
25	2.0948	2.0949	2.0948	2.0950
35	5.1965	5.1967	5.1966	5.1969
45	12.1507	12.1509	12.1508	12.1513
55	29.9180	29.9182	30.9181	29.9186
65	87.1048	87.1044	87.1045	87.1039
75	365.7681	365.7626	365.7645	365.7554

Acknowledgement

These investigations were supported by the State Comission for Technical Development through the Hungarian Space Office by the grant T 010.

References

Alpert Ya L 1974: Radio Wave Propagation and the Ionosphere Vol. 2. Consultants Bureau, New York

Axford W I 1963: *J. Geophys. Res.*, 68, 769–779.

Bauer S J 1973: Physics of Planetary Ionospheres. Springer Verlag, Berlin

Bencze P, Märcz F, Szendrői J 1997: *Acta Geod. Geoph. Hung.* (present issue)

Budden K G 1985: The Propagation of Radio Waves. Cambridge University Press, Cambridge

Hartmann G K, Leitinger R 1984: *Bull. Geod.*, 58, 109–136.

Hoffmann-Wellenhof B, Lichtenegger H, Collins I 1992: Global Positioning System. Theory and Praxis. Springer Verlag, Wien

Klobuchar J A, Leitinger R 1993: In: PRIME studies with emphasis on TEC and topside modelling. Wissenschaftlicher Bericht No 2, Teil 1. Inst. f. Meteorologie und Geophysik, Universität Graz, 13–20.

Acta Geod. Geoph. Hung., Vol. 32(3-4), pp. 439–449 (1997)

ON THE IMPLEMENTATION OF A GPS BASED METHOD FOR THE REAL-TIME MEASUREMENT OF THE TOTAL ELECTRON CONTENT OF THE IONOSPHERE

E ENGLER[1], N JAKOWSKI[1], A JUNGSTAND[1], D KLÄHN[1], E SARDON[1], S SCHLÜTER[1]

Due to the recent developments towards an European Navigation Satellite System and the actual effords in aviation, a rising interest in the real-time monitoring of the ionosphere is noticed. The crucial task for the estimation of the total electron content (TEC) in real-time is in the separation of the ionospheric delay from the total error budget. Parameters affecting the total error budget have been identified to be the number of epochs before the first ambiguity fixing, elevation cut-off angle, the utilization of either predicted or real-time satellite/receiver instrumental biases, and the handling of multipath effects.

In the paper, an operable method for the real-time estimation of TEC is presented. The method is based on the usage of a L1/L2 GPS receiver, preprocessed satellite and receiver biases, and an appropriate ionospheric background model for the preparation of ionospheric maps. A rating of the influences of various processing parameters on the accuracy of the real-time TEC measurements will be given. The evaluation of the accuracy has been obtained by comparing the real-time measurements with TEC obtained by postprocessed data of a receiver at the same location and a subset of the IGS network for reference. The accuracy of the real-time method has been proven to be better than 4 TECU in any case. Depending on actual computer technologies, TEC measurements can be taken up to a sampling rate of 1/3 Hz.

The work on the real-time TEC station was done in the framework of the projects DIRES and RAKO under financial support of the German Space Agency (DARA) at DLR Remote Sensing Ground Station Neustrelitz.

Keywords: ionospheric mapping; ionospheric propagation error; real-time TEC measurement; total electron content, TEC

Introduction to GPS Ionospheric Propagation Error (IPE)

The development of the real-time TEC measurement techniques and equipment originally is intended for the improvement of the accuracy of single user GPS positioning and navigation service. Besides, there is considerable interest in the monitoring of the actual state of the ionosphere in other fields of scientific and engineering research, mainly ionospheric research, space weather forecast and monitoring, aircraft flight navigation, and wide area differential GPS navigation networks (WAAS, EGNOS).

The ionospheric propagation error (IPE) as measured by GPS is due to the dispersive propagation of modulated L-band signals at the two GPS carrier frequencies of 1.57542 GHz (L1) and 1.2276 GHz (L2) in the ionosphere and plasmasphere. As

[1]DLR Remote Sensing Ground Station, Neustrelitz, Germany

1217-8977/97/$ 5.00

well known, dispersion introduces a delay on carrier and modulation phases. Neglecting terms of higher than third order in frequency, the delay introduced into the GPS measurements on the code and carrier phase is given by

$$\Delta_{r/\phi} = \pm\frac{40.3}{f^2} m^3 s^{-2} \mathrm{TEC}\,. \tag{1}$$

Due to the typical hight of about 25000 km and the actual position of an individual GPS satellite a plasmaspheric contribution of about $1\ldots2\cdot10^{16}$ el/m^2 can be included in the measurements. Introducing the more comfortable notation 1 TECU$\equiv 1\cdot10^{16}$ electrons/m^2 and inserting L1 or L2 frequency into (1) we get the propagation error per TECU, that is

$$\frac{\Delta_{r/\phi L1}}{1\ \mathrm{TECU}} \approx \pm 0.162\ \mathrm{m} \qquad \frac{\Delta_{r/\phi L2}}{1\ \mathrm{TECU}} \approx \pm 0.267\ \mathrm{m}\,.$$

These numbers have been proven to be useful if translating from the world of ionospheric research to the world of navigation.

In the next part a short outline of the basic algorithm for the estimation of total electron content from two-frequency GPS measurements as presented in 1995 by the authors (RAKO, Engler et al. 1995, Jungstand et al. 1995, Klobuchar 1987) is given. In the next but one part, basing on the mapping feature added to the last version of the real-time program, a comparison with postprocessed TEC maps will illustrate the accuracy reached by the method. At the end, an outlook for improvements will be given, too.

The real-time TEC measurement algorithm for L1/L2 P-code GPS receivers

Defining the physical quantities to be considered in the followings the notations below are used

k	satellite number or GPS receiver channel number
$P_k(f)$	measured code range at frequency f
$\phi_k(f)$	measured phase range at frequency f
A_k	sum of geometric range, tropospheric error, and clock error
TEC_k	total electron content in the path of observation (slant TEC)
$E_k(f)$	sum of all instrumental delays, multipath, and random noise errors for code measurement
$\xi(f)$	sum of all instrumental delays, multipath, and random noise errors for phase measurement
$N(f)$	phase ambiguity
$\lambda(f)$	wavelength (c/f).

$$\alpha \equiv \frac{40.3\ \mathrm{m}^3}{s^2}, \quad \text{and} \quad \beta \equiv \alpha\frac{f_1^2 - f_2^2}{f_1^2 f_2^2}$$

the fundamental equations for GPS measurements, using L1/L2 frequency P-code receivers reads

$$P_k(f) = A_k + 40.3\ \mathrm{TEC}_k/f^2 + E_k(f) \tag{2}$$

$$\phi_k(f) = A_k - 40.3\ \mathrm{TEC}_k/f^2 - \lambda(f)N(f) - \xi_k(f)\,. \tag{3}$$

The algorithm, as introduced in its first version in (RAKO, Engler et al. 1995) can be found by the following steps:

Step 1 (Aquisition): Evaluation of the mean values $\langle T_k^r - T_k^\phi \rangle_{T_s}$, $\langle A_k^r - A_k^\phi \rangle_{T_s}$ with

$$A_k^\phi \equiv \frac{f_1^2\phi_k(f_1) - f_2^2\phi_k(f_2)}{(f_1^2 - f_2^2)}\,, \tag{4}$$

$$A_k^r \equiv \frac{f_1^2 P_k(f_1) - f_2^2 P_k(f_2)}{f_1^2 - f_2^2}\,, \tag{5}$$

$$T_k^r \equiv P_k(f_1) - P_k(f_2)\,, \tag{6}$$

$$T_k^\phi \equiv \phi(f_1) - \phi(f_2)\,. \tag{7}$$

within a certain observation time T_s. In the course of error discussion we will show, that T_s depends on the mode of operation of the GPS system (antispoofing mode switched ON or OFF) and the minimum satellite elevation cut-off angle. In any case, T_s is limited to about 20 minutes at least.

Step 2 (Solving the ambiguities): Evaluation of the constant L1/L2 phase offset contributions using

$$\begin{aligned} N(f_1) &= INT\left\{\tfrac{1}{\lambda_1\lambda_2}\left[-\kappa_2\lambda_2\left(\langle T_k^r - T_k^\phi\rangle - (b_k(f_1) - b_k(f_2))\right) - \right.\right. \\ &\quad \left.\left. -\kappa_1\lambda_1\left(\langle A_k^r - A_k^\phi\rangle + (\kappa_1 b_k(f_1) - \kappa_2 b_k(f_2))\right)\right]\right\} \\ N(f_2) &= INT\left\{\tfrac{1}{\lambda_1\lambda_2}\left[\lambda_2\left(\langle T_k^r - T_k^\phi\rangle - (b_k(f_1) - b_k(f_2))\right) + \right.\right. \\ &\quad \left.\left. +\lambda_1\left(\langle A_k^r - A_k^\phi\rangle + (\kappa_1 b_k(f_1) - \kappa_2 b_k(f_2))\right)\right]\right\} \end{aligned} \tag{8}$$

INT() is the integer part of the argument. To this purpose, the instrumental biases $b(f_1)$ and $b(f_2)$ must be known.

Step 3 (Continuous Monitoring): If step 1 and step 2 have been performed successfully, slant TEC can be measured continuously by processing the actual phase measurements $\phi(f_1)$, $\phi(f_2)$ of each epoch, that is by inserting into

$$\mathrm{TEC}_k = \frac{1}{\beta}\left(\phi(f_1) - \phi(f_2)\right) + \Delta\Phi\,. \tag{9}$$

To get the simple form of Eq. (8), code and phase measurement disturbances must obey the following approximations within a certain limit:

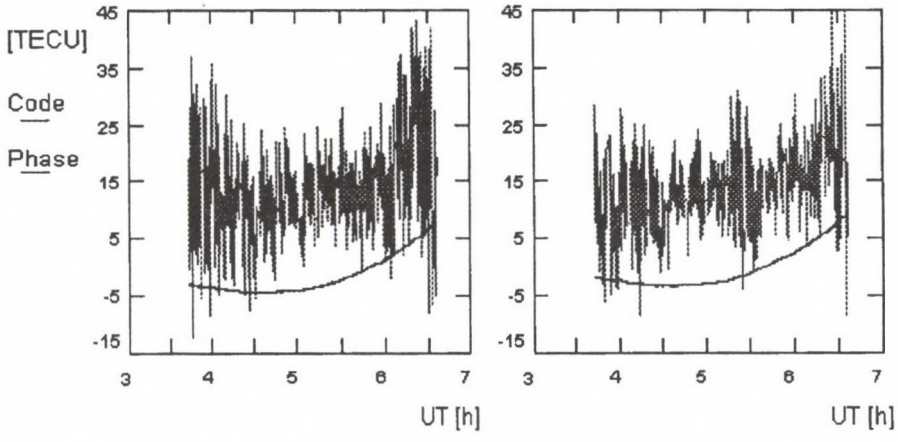


Fig. 1. Slopes of code and phase TEC measurements before fixing the ambiguites for two different receiver stations

— Code and phase noise:

$\langle f_1^2\zeta_{1k} - f_2^2\zeta_{2k}\rangle$	$\to \varepsilon$	Mean weighted phase noise difference must be small
$\langle \zeta_{1k} - \zeta_{2k}\rangle$	$\to \varepsilon$	Mean phase noise difference must be small
$\langle f_1^2 n_{1k} - f_2^2 n_{2k}\rangle$	$\to \varepsilon$	Mean weighted code noise difference must be small
$\langle n_{1k} - n_{2k}\rangle$	$\to \varepsilon$	Mean code noise difference must be small

— Multipath effects:

$\langle f_1^2 m_{1k} - f_2^2 m_{2k}\rangle$	$\to \varepsilon$	Mean weighted multipath error difference must be small
$\langle m_{1k} - m_{2k}\rangle$	$\to \varepsilon$	Mean multipath error difference must be small.

The restrictions on the multipath errors turned out to be the most restrictive ones. In principle the algorithm calibrates phase-TEC slope into the mean range-TEC slope by estimating the phase ambiguities after fixing period T_s taking advantage of the high precision of phase TEC derived from L1/L2 frequency measurements. Figure 1 illustrates this procedure for two different GPS satellites. The figure also gives a good impression on the high noise level, that disturbs the code measurements and the very low phase noise level. It can be seen that using phase measurement result in a very high relative accuracy. On the other hand, due to the high noise and multipath rates of the code measurements, Fig. 1 shows that it is totally hopeless to perform high accuracy TEC measurements by using code measurements only.

In the description of the algorithm we noticed, that it is necessary to know the instrumental code delays of all system components, that is of all GPS satellites and the GPS receiver. Inspecting Eq. (8), we conclude that it is necessary to know

Table I. Postprocessed (GILION) rms error of receiver biases, found for a 2-month period for different European IGS reference stations

	ds10	ds42	brus	bor1	mate	nzre	nzmo	onsa	pots	trom	wtz1
rms/ns	0.25	0.37	0.29	0.29	0.31	0.40	0.23	0.98	0.45	0.58	0.28

Table II. Postprocessed (GILION) rms error of GPS satellite biases, found for a 2-month period

PRN	1	2	4	5	6	7	9	12	14	15	16	17
rms/ns	0.16	0.14	0.12	0.11	0.12	0.11	0.10	0.11	0.12	0.15	0.15	0.18
PRN	19	20	21	22	23	24	25	26	27	28	29	31
rms/ns	0.14	0.13	0.16	0.18	0.16	0.13	0.16	0.12	0.13	0.22	0.14	0.17

the differential biases $\Delta b_k = b_k(f_1) - b_k(f_2)$ only*. In our case, we can process the differential biases for the real-time station by GILION, a TEC and GPS bias monitoring program (Coco et al. 1991, Sardón et al. 1994) operated since several years at our location. To take advantage of GILION, the GPS receiver intended to be used for the real-time TEC monitoring must be included for calibration as an additional station into the standard monitoring network (some European selection of the IGS). Table I gives an impression on the accuracy of the differential biases obtained. 'nzre' and 'nzmo' are the identifiers for two of the TurboRogue SNR8100 GPS receivers (reference and monitoring station) operated at our location.

Stability of satellite instrumental biases:

Besides, as can be seen from Table II, the stability of the differential biases turns out to be sufficiently good to be used for several months. This holds true as long as there is no hardware modification, either with the GPS receiver nor in the GPS satellite system. In case a new satellite is added to the system, a recalibration run, at least for the new satellite must be done. The same holds if one of the receivers was replaced or modified. Next it turned out that the differential biases differ for antispoofing ON/OFF mode, so two sets of differential bias data must be maintained. Due to the fact that antispoofing is switched off in very rare cases, the 'antispoofing on' data are the more important ones. Reviewing the data for two monts, the residual rms TEC error introduced by using postprocessed biases turned out to be 1.7 TECUs at least.

Next problem to answer is to get an estimate for the time needed to perform step 1 and step 2 of the algorithm, that is to define the time T_s needed to hold the

*Doing so, there is a residual error due to the term $\kappa_1 b_1 - \kappa_2 b_2$ that must be represented by the differential bias $\Delta b = b_1 - b_k$ only. As shown by Engler et al. (1995), Coco et al. (1991), Sardón et al. (1994) due to the statistical nature of b_1 and b_2 this can be done within the limits of previous approximations.

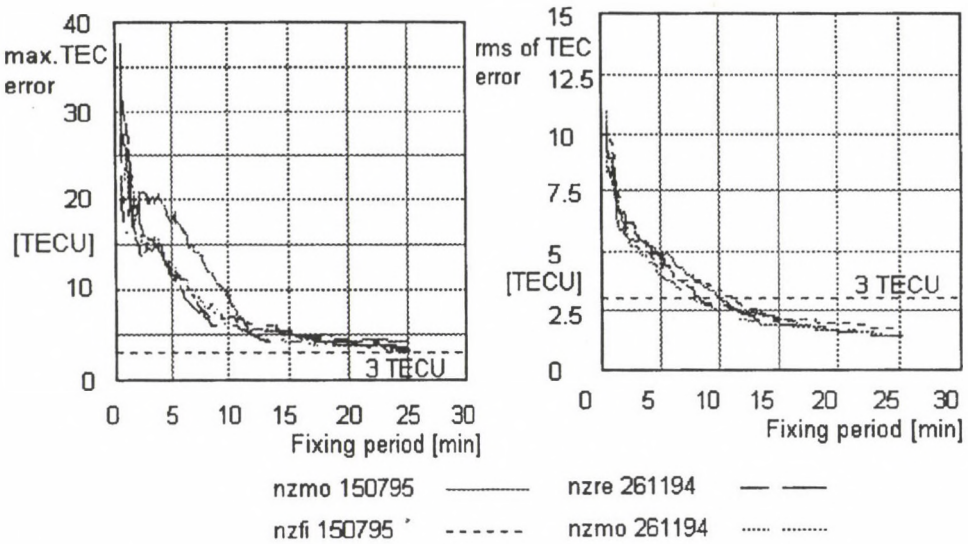


Fig. 2. Maximum TEC error and rms TEC error compared to postprocessed TEC measurements. The fixing period T_s was varied and an elevation cut-off angle of 20° was used in the algorithm (antispoofing ON)

errors below the measurement threshold. To this purpose a series of investigations have been performed. A summary is given in Fig. 2 for the case antispoofing was switched ON. We conclude that to achieve at a maximum TEC error of 4 TEC we need an aquisition time of 20 minutes. Then the rms error of TEC will be below 2 TECU. In case AS is switched OFF, an aquisition time of 10 min was found to keep the errors within the same limits. Similar investigations have been done to get an estimate for the lower elevation threshold. The elevation threshold influences the multipath level, code noise, and the number of phase cycle slips to deal with during the aquisition phase. To keep these disturbances below the error limits, 10 have been proven to be a good number in case AS is switched OFF, and 20° have been proven to be tolerable in case AS is switched ON (RAKO).

After the implementation of the algorithm, the measurement errors have been verifyed by several experiments. The results of these experiments have been compared to TEC estimates gained by the postprocessing program GILION.

The maximum TEC error was found to be

— $4 \cdot 10^{16}$ el/m^2 for quiet ionospheric situations

— $5 \cdot 10^{16}$ el/m^2 in case of ionospheric storms

relative to the GILION results.

The local range of validity for this results is

— $\leq$ 500 km if AS is switched ON

— $\leq$ 800 km if AS is switched OFF

related to the position of the real-time station.

Ionospheric mapping

Mapping can be done by supporting the real-time TEC measurements by an appropriate ionospheric background model. To this purpose presently the NTCM2 model for Central Europe was implemented. Due to the aproach proposed by Jakowski (RAKO) the vertical TEC at some grid point k, l of the map is computed by the mapping algorithm:

$$TEC_v(k,l) = ac(k,l) + m(k,l) + \overline{am}$$

with:

$$ac(k,l) = \frac{\sum_{j=1}^{N} a_j WF(k,l)}{\sum_{j=1}^{N} WF(k,l)},$$

and N the number of observed satellites. The weighting function is computed by

$$WF(k,l) = \exp\left(\frac{(r_k - r_l)^2}{2s^2}\right)$$

with $a_j = am_j - \overline{am}$, $am = \frac{1}{N}\sum_{j=1}^{N} am_j$, $am_j = ao_j - m_j$, $ao_j = MF^{-1} aos_j$, and the mapping function

$$MF = \left(1 - \left[\frac{r_E \cos\beta}{r_E + h}\right]\right)^{-\frac{1}{2}},$$

with

aos_j — measured slant TEC for j-th satellite

m_j — vertical TEC from model for pierce point of j-th satellite.

A comparison of postprocessed TEC maps and maps from the real-time station is given in Figs 3 and 4.

As can be seen from the figures, also for the time considered the difference in the maps is within the discussed error limits. The tendency of the real-time iso-TEC lines is in good agreement with the high precision results, too. To convince of the performance of the real-time TEC monitoring, a 24-hour series of visualized TEC maps, both postprocessed and real-time have been available for several month at our ftp-server. If the NTCM2 TEC model is used for the background model, the mapping feature is restricted to the local range of the validity of the NTCM2. This restriction can be suspended by taking advantage of global models, like the Klobuchar model (Engler et al. 1995). Then however, we must count for an additional error of about 0.5 to 1 TECU at the rim of the area observed by the real-time station.

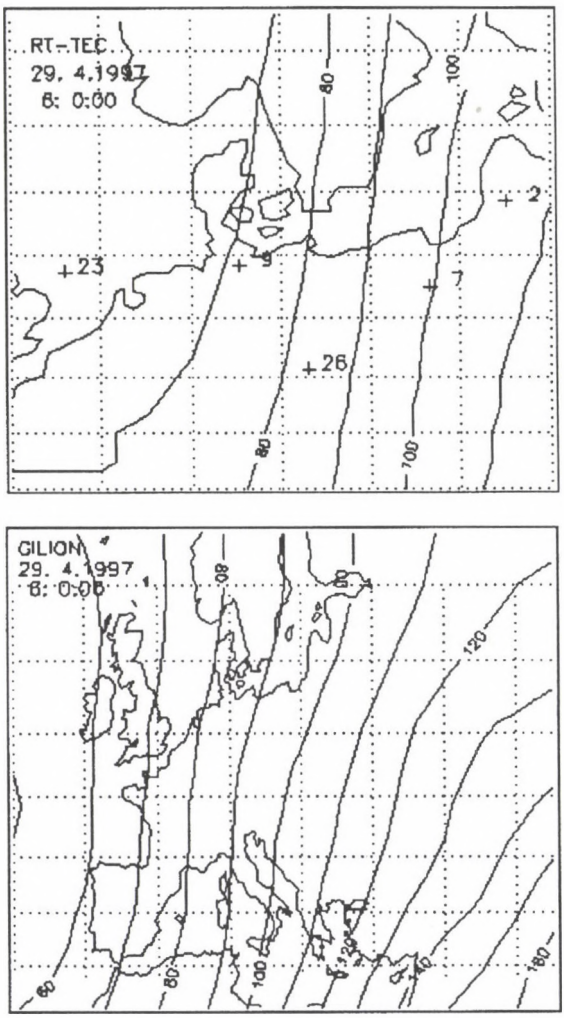


Fig. 3. Maps of postprocessed (GILION) and real-time derived TEC (RT-TEC) for Apr. 29, 1997, 6:00UTC. Notice the difference in geographic area covered by the maps. The iso-TEC lines are marked by 10*TEC value

HW/SW implementation of the algorithm

The algorithm was implemented into a real-time TEC processing software. The minimum hardware requirements to run the program are

— PC, 100 MHz 486
— Win 3.11 or Win95
— 16 MB RAM
— 400 MB HD
— 2 serial ports (UART 16550 AFN or eq.)

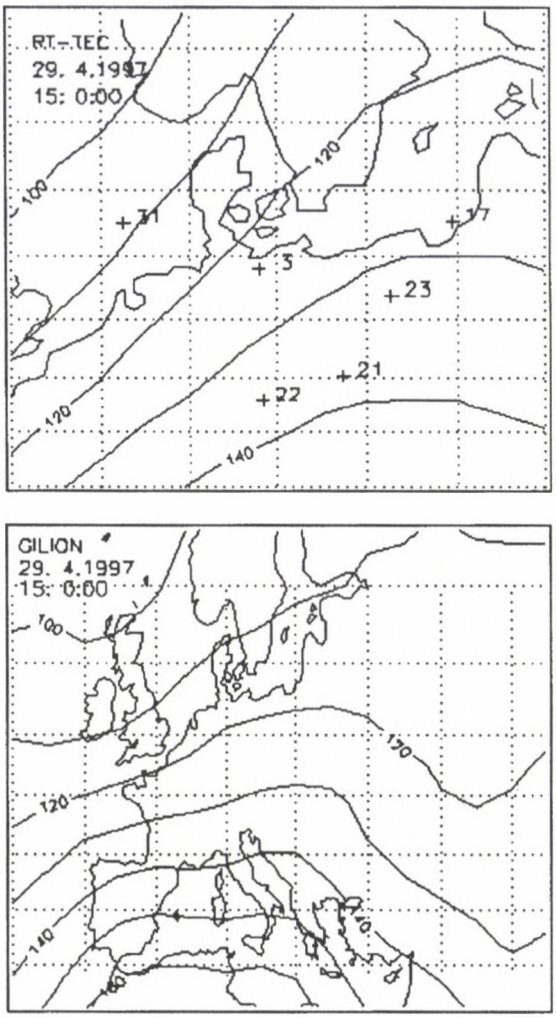


Fig. 4. Maps of postprocessed (GILION) and real-time derived TEC (RT-TEC) for Apr. 29, 1997, 15:00 UTC. Notice the difference in geographic area covered by the maps. The iso-TEC lines are marked by 10*TEC value

— Turbo Rogue 8000 series GPS receiver (a must).

While processing, the program stores the TEC map data on a disk. For the final visualization of the maps a data visualization program can be used. At DLR we are using IDL.

Besides the processing of the real-time TEC maps, the program offers some additional comforts, namely it

— permits to set all processing parameters (elevation cutoff, aquisition time, data file naming),

— initializes the RS232 serial communication lines and the GPS receiver (baud rate etc.),
— collects all GPS raw data and system informations (for data base storage and postprocessing),
— computes and collects all TEC measurements,
— processes a daily accumulated TEC-Map,
— and loggs all processing activities for later performance diagnosis.

Due to 38.4 kBaud maximum communication rate of the SNR8000 GPS receiver, the receiver specific TurboASCII data type, and the time needed for the numerical processing ($\approx$ 0.4 seconds/epoch), the measurement update rate is limited to 1/3 Hz on the 386/100 MHz computer system. To break this limitations a new version for Win95 is under actual consideration. Besides, in the new version the SNR8000 highly compressed TurboBinary format will be used to speed up overall performance.

Summary and conclusions

An algorithm for real-time TEC measurements was developed for L1/L2 P-code GPS receivers, able to cope with A/S. The maximum TEC error is about 4 TECUs, the rms of measurements is about 2 TECUs. The algorithm independently processes all satellites and receiver channels and was implemented into an easy-to-use real-time software. Verification was done by comparison to postprocessed high precision TEC measurements as obtained by GILION.

In our opinion, the algorithm and its implementation covers the requirements of most real-time TEC monitoring, mapping, and networking applications.

A further reduction of the measurement error can be achieved by reducing the residual multipath error, f.i. by using lookup tables for local multipath environment.

Last not least, a reduction in the residual errors and a reduction of the aquisition time T_s may be achived at by using more sophisticated statistical methods like Kalman filtering and fast ambiguity searching methods.

Acknowledgements

For the support of our work within the framework of the projects DIRES and RAKO our thanks are due to the German Space Agency (DARA). Moreover, our thanks are due to H G Kugland, K Hiersche and U Netzband for their steady work in the daily GPS data processing and archiving routines. Special thanks are due to J U Walther for keeping our network services in high operability.

References

Abschlußbericht Projekt RAKO an die Deutsche Agentur für Raumfahrtangelegenheiten, DARA (Final report of the project RAKO as given to the German Space Agency, DARA).

Coco D S, Cocker C, Dahlke S R, Clynch J R 1991: *IEEE Transactions on Aerospace and Electronic Systems*, 27, 931–938.

Engler E, Sardón E, Klähn D 1995a: In: Proc. ION GPS 95, Palm Springs (CA) 1183–1191.

Engler E, Sardón E, Jakowski N, Jungstand A, Klähn D 1995b: In: Proc. IGS-Workshop, Potsdam

Jungstand A, Engler E, Sardon E, Klähn D 1995: In: Proc. ION Technical Meeting 1995, Anaheim (CA), USA

Klobuchar J A 1987: Ionospheric time-delay algorithm for single frequency GPS users, IEEE Transactions on Aerospace and Electronic Systems. Vol. AES-23, No. 3, 325–331.

Sardón E, Rius A, Zarraoa N 1994: *Radio Science*, 29, 577–586.

Acta Geod. Geoph. Hung., Vol. 32(3–4), pp. 451–459 (1997)

RELIABILITY OF THE TEC COMPUTED USING GPS MEASUREMENTS — THE PROBLEM OF HARDWARE BIASES

R WARNANT[1]

In this paper, we outline the procedure used at the Royal Observatory of Belgium in order to compute the TEC using GPS measurements with a precision of 2–3 TECU. This procedure requires the determination of the so-called receiver and satellite *differential group delays*. The combined biases (receiver + satellite) are determined on a daily basis (i.e. one solution a day) using the geometry-free combination of code observations. The method is applied to a network of 7 permanently operating Turbo Rogue receivers; the reliability of these computed biases is discussed.

Keywords: hardware biases; GPS; ionosphere; total electron content, TEC

1. Introduction

The Global Positioning System has already proved to be a very useful tool to study the ionosphere. Indeed, GPS code and carrier phase measurements can be processed in order to determine the Total Electron Content (Lanyi and Roth 1988, Warnant 1996). In practice, the TEC can obtained from:

1. The so-called geometry-free combination of dual frequency code measurements, $P^i_{p,GF}$;

$$P^i_{p,GF} = P^i_{p,L1} - P^i_{p,L2} \,. \tag{1}$$

This equation can be rewritten in function of the Total Electron Content, TEC^i_p:

$$P^i_{p,GF} = -1.05 \cdot 10^{-17} TEC^i_p + (DG_p - DG^i) \tag{2}$$

with

TEC^i_p slant TEC measured along the path going from satellite i to receiver p;

DG^i, DG_p the satellite i and receiver p differential group delays;

$P^i_{p,L1}$, $P^i_{p,L2}$ the $L1$, $L2$ P-code measurements made by receiver p on satellite i.

When the Anti-spoofing is active (it is the case since January 31, 1994), the code observations have a precision ranging from a few decimeters to more than one metre. These measurements are not ambiguous but contain biases called receiver and satellite differential group delays. The existence of these biases is due to the fact that the two GPS frequencies undergo different propagation delays inside the receiver and satellite hardware.

2. The geometry-free combination of dual frequency phase measurements $\Phi^i_{p,GF}$;

$$\Phi^i_{p,GF} = \Phi^i_{p,L1} - \frac{f_{L1}}{f_{L2}} \Phi^i_{p,L2} \tag{3}$$

[1]Royal Observatory of Belgium, Avenue Circulaire 3, B–1180 Brussels, Belgium, e-mail: R.Warnant@oma.be

1217-8977/97/$ 5.00

or rewritten in function of the TEC:

$$\Phi^i_{p,GF} = -5.52 \cdot 10^{-17} TEC^i_p + N^i_{p,GF} \tag{4}$$

with

f_{L1}, f_{L2} the frequency of the $L1$, $L2$ carriers;
$\Phi^i_{p,L1}$, $\Phi^i_{p,L2}$ the $L1$, $L2$ carrier phase measurements made by receiver p on satellite i;
$N^i_{P,GF}$ a real ambiguity.

Phase measurements have a precision better than one millimetre but contain an initial ambiguity which is real in the case of the geometry-free combination. In the absence of cycle slips, $N^i_{P,GF}$ has to be solved for every satellite pass.

3. A combination of geometry-free code and phase measurements.

$$P^i_{p,GF} - \lambda_{L1}\Phi^i_{p,GF} = (DG_p - DG^i) - \lambda_{L1}N^i_{p,GF} \tag{5}$$

with λ_{L1} the $L1$ carrier wavelength.

This combination is used to solve the ambiguity $N^i_{P,GF}$ which is introduced in Eq. (4) in order to determine the TEC. This third method allows to combine the advantages of both measurement types: the TEC is obtained from the precise phase measurements but the information contained in the code observations is used to solve the ambiguity. Nevertheless, the procedure requires the determination of the receiver and satellite differential group delays. In most of the cases, these biases have to be computed:

— the satellite biases: they are measured by the manufacturer before the satellites are launched but these values are not valid any more when they are on their orbit (see, for example, Wilson and Mannucci 1993, Wanninger et al. 1994, Warnant 1996).

— the receiver biases: in the past, the old Rogue receivers (the so-called Big-Rogue and Mini-Rogue) had an auto-calibration function allowing to measure the receiver bias. Unfortunately, this function does not exist any more on the Turbo Rogue receiver. To our knowledge, no other receiver has this capability.

2. Computation of the biases

The Royal Observatory of Belgium has a network of 7 permanent GPS stations (Fig. 1). The station of Brussels is in continuous operation since April 1993. Dentergem, Dourbes and Waremme were installed in January 1994. In this section, we outline the method we have developed to study the hardware biases. The method has been applied to our Turbo Rogue network.

In fact, the error made in the determination of the differential group delays is the largest error source when computing the TEC using GPS measurements. It is clear that these biases cannot be neglected: for example the bias of one of our receivers (serial number 238) is +5.33 ns. The fact to neglect it would give an error of 16 TECU on the computed TEC.

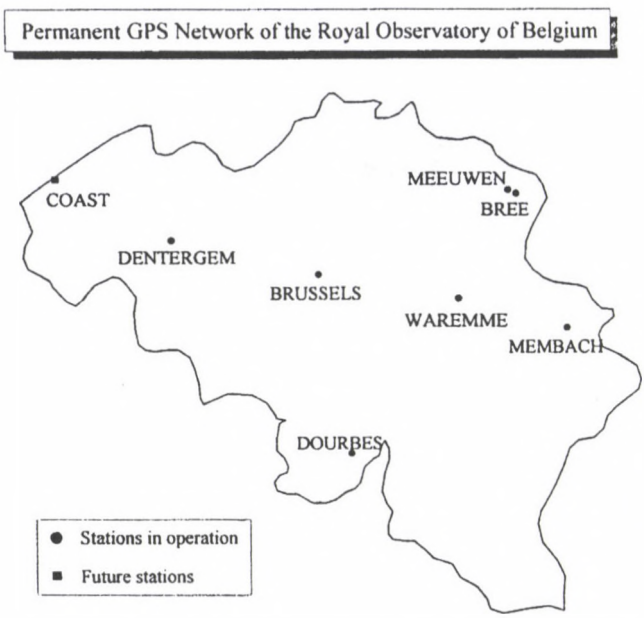


Fig. 1. The permanent network of the Royal Observatory of Belgium

In practice, the combined biases (receiver + satellite) are determined using Eq. (1) where the ionosphere is modelled by means of a simple polynomial in latitude and local time. Every model (i.e. polynomial) is computed using periods of about 6 hours of data. During this procedure, the "usual" assumptions are made:

— the ionosphere is concentrated in a spherical shell of infinitesimal thickness located at a height of 350 km; the intersection between this layer and the satellite line of sight is called the *ionospheric point*;

— "static" behaviour of the ionosphere on short periods (6 hours): the TEC only depends on latitude and local time;

— the receiver and satellite biases are constant on short periods. In the case of the Turbo Rogue biases, we have verified this assumption: during a period of a few hours, the biases remain constant within 0.2 ns (Warnant and Wanninger 1994, Warnant 1996).

The combined biases are determined on a daily basis (i.e. one solution a day). We only process night-time observations with an elevation mask of 20°: in most of the cases, the ionosphere is "quieter" during the night. This procedure which has been applied to our 7 Turbo Rogue GPS receivers has allowed us to study the long-term behaviour of the computed biases. The main results obtained with this network can be summarized as follows:

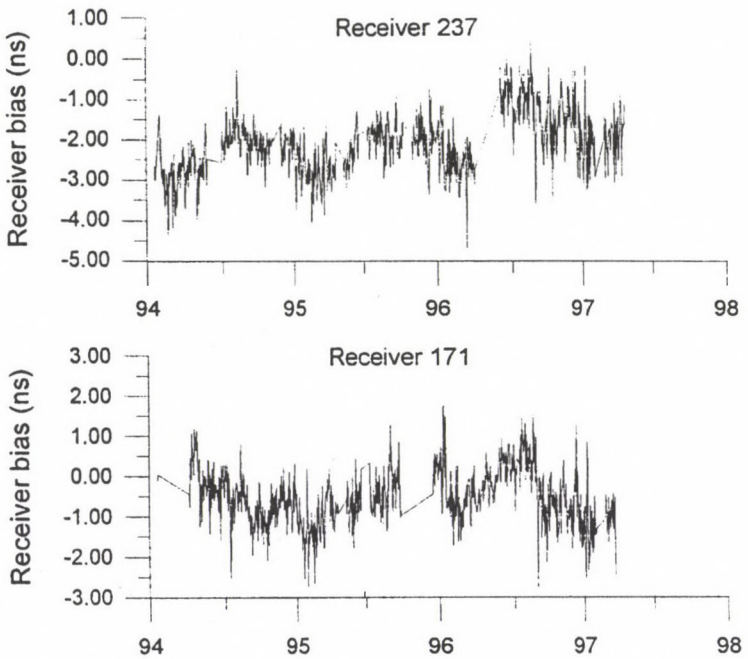


Fig. 2. Computed biases for Turbo Rogue receivers 237 and 171

— the Turbo Rogue bias changes as the Anti-Spoofing is activated or not: we have observed that, in most of the cases, the difference between the two values (i.e with Anti-Spoofing on or off) is about ± 0.5 ns. Since January 31, 1994, the Anti-spoofing is activated but sometimes, it is turned off on a few or all the satellites;

— the biases of two "identical" receivers even with two consecutive serial numbers can be very different from each other: for example, the biases of receivers 237 and 238 are respectively -2.37 ns and $+5.33$ ns;

— the bias depends on the temperature; the effect is visible during hot summer days;

— Figure 2 shows the computed differential group delays for 2 receivers (serial numbers 237 and 171) in a period of about 3 years. By looking at this figure, it can be seen that the bias changes (sometimes much) as soon as in the hardware is changed a little (after a repair, for example): in June 1996, we have replaced the microprocessor of receiver 237: you can see very clearly a jump in the computed biases after this change; a similar replacement has been made receiver 171 in December 1995 and a repair has been performed in September 1996.

In addition, the computed biases have a (short-term) day-to-day variability

of about 1 ns and have a periodic (seasonal) behaviour. Which is the origin of these variations in the *computed* biases? There are 2 possible explanations: the variations are due to a real change of the receiver bias or they are due to the fact that the use of our simple polynomial to model the ionosphere gives rise to residual errors which vary from day to day, from season to season, ... as in the case of the ionosphere. If this last explanation is true, then we would expect that the residual errors would be similar for the different receivers: from Fig. 1, it is clear that the seasonal behaviours of receivers 171 and 237 are very similar; it is also the case of the other receivers. If these seasonal variations were real changes of the biases, the effect would be different for the different receivers. It could be argued that this seasonal trend could be due to the environmental parameters in which the receivers are placed: the external temperature also depends on the season. Nevertheless, in our case, this explanation cannot be true: the receiver installed at Brussels which is placed in a room where the temperature remains constant within 2 or 3°K undergoes the same seasonal variations as the other receivers.

For these reasons, the value of the bias adopted to determine the TEC is obtained by computing the mean of all the daily solutions in a period the duration of which depends on data availability: it ranges from one month to more than one year. This technique has the advantage that it reduces the influence of the ionospheric residual errors: most of the effects are cancelled in the mean.

— in addition to these "artificial" variations, the receiver biases also undergo real (sometimes unexplained) changes; such variations can be detected by forming the difference between the biases computed for 2 different receivers: the difference has the advantage that it removes the common seasonal effects; Fig. 3 shows the difference between the delays computed for receivers 238 and 321. Receiver 321 was installed at Brussels where the temperature remains stable. In March 1996, the difference suddenly increases by about 2 ns, remains stable during a few days and then returns to the original value. In fact, the bias of receiver 238 has changed: we have seen it by forming differences with other receivers. We have no idea about the origin of this problem. We have explained in the previous paragraph that we adopt the long-term mean of all the daily computed biases to determine the TEC in order to reduce the residual ionospheric errors. In the case of receiver 238, the bias variation was real; as a consequence, we have seen a corresponding jump in the TEC coming from receiver 238 as compared with the other (Fig. 3). For this reason, it is clear that the receiver bias has to be very regularly controlled.

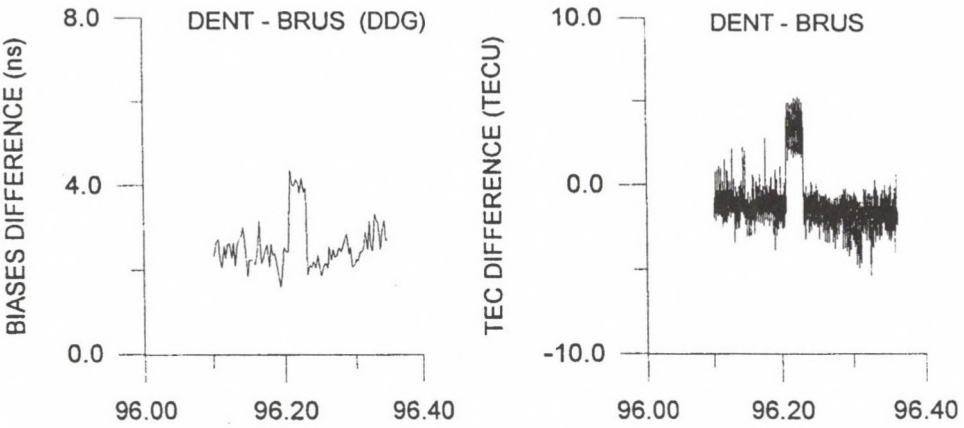


Fig. 3. Difference between the biases (left) and the TEC (right) computed for the receivers installed at Dentergem (sn 238) and Brussels (sn 321)

3. Determination of the TEC

When the biases have been determined, the TEC is computed as a function of latitude and local time (or longitude) of the ionospheric point. For example, the data collected at Brussels (latitude = 50.8°N, longitude = 4.4°E) allow to compute the TEC from about 35°N to 60°N in latitude and from −20°W to 25°E in longitude.

To obtain TEC profiles representative of the ionosphere above the observing station, we apply the following procedure:

— we select all the TEC values corresponding to an ionospheric latitude, L_{iono}, given by:

$$L_{\text{sta}} - 1.5° \leq L_{\text{iono}} \leq L_{\text{sta}} + 1.5°$$

where L_{sta} is the latitude of the observing station;

— we compute the mean of these TEC values in 15 minute periods.

Figure 4 shows the TEC profile above Brussels on October 30 1994 obtained by this method. The procedure outlined here above has been applied to a data set covering a period of more than 8 years: since May 1989, GPS measurements have been regularly performed at Brussels. Since April 1993, the measurements are continuous. Three additional permanent stations are in operation since January 1994 (Dentergem, Dourbes and Waremme); 3 new stations (Bree, Meeuwen and Membach) have been installed in 1997.

Figure 5 shows, as example, the results obtained at Brussels for two months (January 1994 and January 1997): in this figure, all the TEC profiles corresponding to the same month are represented on the same graph.

To verify the reliability (precision and accuracy) of the TEC determined by the method developed at ROB, two experiments have been performed:

1. *Precision*: we have compared the TEC obtained at Brussels with the TEC computed at Dentergem, Dourbes and Waremme (the typical distance between stations is about 70 km) during a period of more than 3 years: in most of the cases, the difference between the TEC at Brussels and the TEC in the other Belgian stations remains within 2–3 TECU (Fig. 6).

2. *Accuracy*: at Dourbes, an ionosonde which is the property of the Belgian Meteorological Institute is collocated with our GPS station. GPS and ionosonde de-

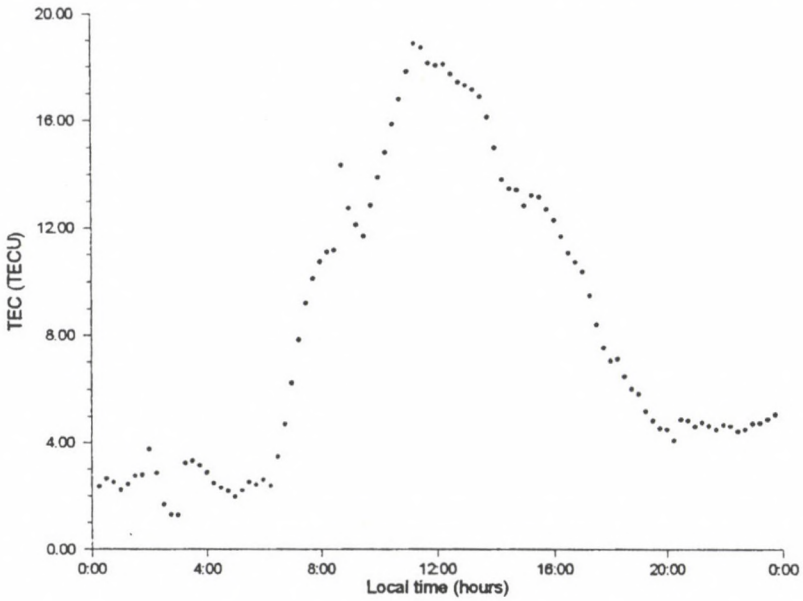


Fig. 4. TEC at Brussels on October 30 1994

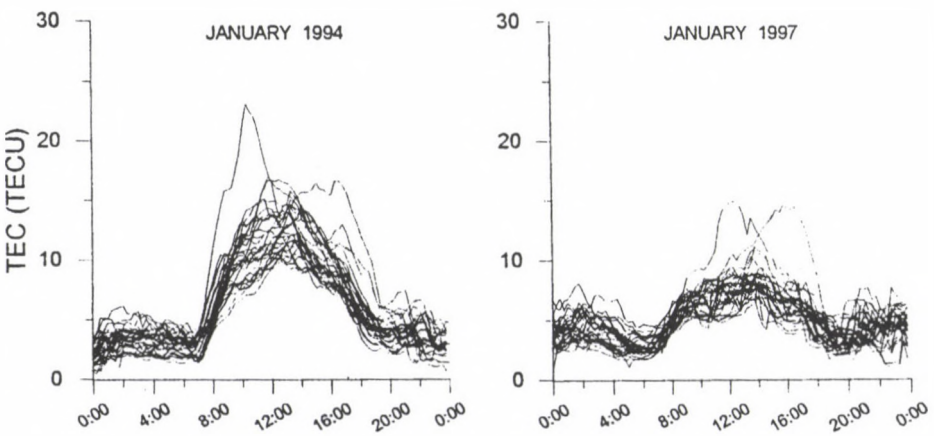


Fig. 5. TEC at Brussels during 2 month (January 1994 and January 1997)

rived TEC have been compared on a period of 2 years (Warnant and Jodogne 1997). Again, the two methods are in agreement within 2–3 TECU.

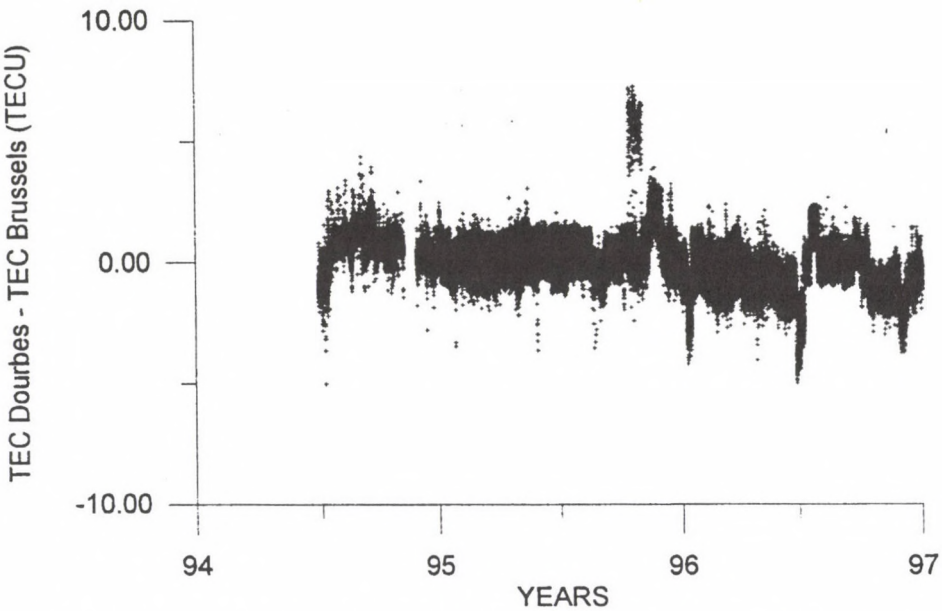


Fig. 6. Difference between the value of the TEC computed at Dourbes and Brussels

4. Conclusion

The paper has outlined the method developed at the Royal Observatory of Belgium in order to compute the Total Electron Content. This method requires the determination of the receiver and satellite differential group delays. The combined biases (receiver + satellite) are obtained on a daily basis from the geometry-free combination of code observations. The computed biases undergo day-to-day and seasonal variations; in most of the cases these variations are "artificial": they are due to ionospheric residual errors. For this reason, we adopt the mean value of all the computed biases on a long period to reduce the influence of these errors. Nevertheless, the biases also undergo "real" changes: they are sometimes due to the temperature (during hot summer days) but we have not been able to explain all these variations. For this reason, the receiver bias has to be regularly controlled. In addition, the consistency (precision) of our method has been verified: the TEC computed at Brussels has been compared to the TEC obtained at Dentergem, Dourbes and Waremme: they agree within 2–3 TECU.

References

Lanyi G E, Roth T 1988: *Radio Science*, 23, 483–492.

Wanninger L, Sardón E, Warnant R 1994: In: Proceedings of Beacon Satellite Symposium '94, Dépt. of Physics of University of Aberystwyth ed., 13–16.

Warnant R 1996: Etude du comportement du Contenu Electronique Total et de ses irrégularités dans une station de latitude moyenne. Application aux calculs de positions relatives par le GPS. Ph.D. Thesis (in French), Série Géophysique (N Hors-Série) de l'Observatoire Royal de Belgique, Bruxelles

Warnant R, Jodogne J-C 1998: A comparison between the TEC computed using GPS and ionosonde measurements. *Acta Geod. Geoph. Hung.* (in press)

Wilson B D, Mannucci A J 1993: In: Proceedings of ION GPS'93, Salt Lake City

Acta Geod. Geoph. Hung., Vol. 32(3-4), pp. 461-468 (1997)

STATISTICAL METHOD FOR DETERMINING CHARACTERISTICS OF THE DYNAMICS OF THE RADIO SIGNAL INTERFERENCE PATTERN

E L AFRAIMOVICH[1]

A statistical method is proposed for determining the characteristics of the radio signal interference pattern dynamics by measuring variations of phase derivatives with respect to time, and spatial coordinates as a function of time or frequency. These data are used to calculate corresponding values of the velocity vector. Subsequently, velocity and direction distributions are constructed and analyzed to verify the hypothesis of the presence of a predominant displacement. If it exists, then the pattern can be considered to be travelling, and the mean travel velocity can be determined from the velocity distribution. The use of this method is exemplified by an investigation of medium-scale travelling ionospheric disturbances (MSTIDs) and small-scale irregularities (SSI) from simultaneous polarization, angle-of-arrival and scintillation measurements of the ETS-2 radio signal at 136 MHz made in Irkutsk (52° N, 104° E) during December 1989 to December 1990. Method may be useful both in experiments with ionosphere-reflected radio waves and in transionospheric radio sounding operations. The new avenues for remote diagnostics of TIDs are based on exploiting standard measurements of transionospheric radio signal characteristics and coordinate-time measurements using multichannel receivers of the Global Positioning System (GPS).

Keywords: interference pattern; ionosphere; transionospheric sounding, travelling ionospheric disturbances; vertical incidence sounding

1. Introduction

TransIonospheric Radiointerferometer (TIR) (Afraimovich et al. 1991) is designed for simultaneous measurements of the main parameters of a 136 MHz radio signal from the geostationary satellite, ETS-2, located at 130° E (polarization rotation, angles of arrival, frequency Doppler shift, and amplitude — with a time resolution of 30 s and amplitude scintillations with a time resolution of 0.2 s). The comprehensive nature of our measurements offer a means to investigate the structure and dynamics of ionospheric irregularities simultaneously over a wide range of scales, from a few hundred meters to several hundred kilometers.

Three receiving points A,B,C of the radio interferometer were located at the apexes of a nearly right-angle triangle, with sides about 200 m long. The receiving antennas for channels A, B, and C were single polarization spiral antennas. In addition, the central point, A, incorporated an additional antenna, D, of opposite polarization. The set of antennas at A and D was used to measure the phase difference, $\Delta\phi$, between antennas A and D which is proportional to the total electron content (TEC), $I(t)$, along the line of sight to the satellite.

[1]Institute of Solar-Terrestrial Physics, POB 4026, Irkutsk, 664033, Russia, Fax: +7 3952 462 557, e-mail: afra@iszf.irk.ru

1217-8977/97/$ 5.00

The set of A, B, and C formed part of the system for measuring time variations of the phase difference $\Delta\phi$ along baselines CB and AB, proportional to values of the gradients I'_x and I'_y, where x and y are directed, respectively, eastward, E, and northward, N. Furthermore, amplitude scintillations with a time constant of 0.2 s were recorded simultaneously from the same antennas.

To implement the method described in Section 2, the continuous series $\Delta\phi_{AB}(t)$, $\Delta\phi_{CB}(t)$ and $\Delta\phi_{AD}(t)$ of at least two-days duration were filtered out in order to eliminate diurnal trends in the measured quantities. An eighth-order Butterworth filter was used in investigating MSTIDs. The period range 30–60 min, which lies within the generally accepted range of MSTIDs periods, was used.

Filtered series of $\Delta\phi_{AB}$, $\Delta\phi_{CB}$ and $\Delta\phi_{AD}$ were used to determine TID-induced TEC disturbances $\delta I(t)$ and variations in the phase derivatives, $\phi'_x = \Delta\phi_{CB}/\Delta x$ and $\phi'_y = \Delta\phi_{AB}/\Delta y$, where Δx and Δy are the reduced values of the baseline in the E and N directions. Variations of the phase time derivative $\phi'_t(t)$ were determined by differentiating $I(t)$ (Afraimovich et al. 1991).

2. Statistical measurement of the radio phase front velocity vector

Our statistical method developed for determining the velocity, v, and direction, ψ, of the radio signal phase front propagating in the antenna system plane is a generalization of a method proposed by Mercier (1986).

In the simplest form, space-time variations in phase, $\Delta\phi(x, y, t)$, of the transionospheric radio signal are proportional to TEC variations, $\Delta I(x, y, t)$, in the ionosphere, at each time, t, and can be represented in terms of a flat-Earth approximation of a travelling plane phase front

$$\Delta\phi(x, y, t) = k\Delta I(x, y, t) = \phi'_t(t)\Delta t + \phi'_x\Delta x + \phi'_y\Delta y\,, \tag{1}$$

where k is the proportionality factor, and ϕ'_t, ϕ'_x and ϕ'_y are the first phase derivatives with respect to time, t, with respect to the spatial coordinates x and y respectively.

The propagation velocity, $v(t)$, and the direction, $\psi(t)$, of the phase front at each time may be defined as

$$\begin{cases} W_y(t) &= \phi'_y(t)/\phi'_t(t) = |W|\cos\psi \\ W_x(t) &= \phi'_x(t)/\phi'_t(t) = |W|\sin\psi \\ v(t) &= |W|^{-1} = \left(W_x^2(t) + W_y^2(t)\right)^{-1/2}\,. \end{cases} \tag{2}$$

In radio astronomical interferometry, measurements are only made of the spatial derivatives of $\phi'_y(t)$ and $\phi'_x(t)$, which determine the direction modulus ψ 180° (Mercier 1986). Using the time derivative of the signal phase makes it possible to determine the unambiguous orientation of the wave vector of the phase disturbance in the range 0–360° at each particular instant of time.

In practice, we use instantaneous values of v and ψ, determined every 30 s, to construct — on a selected time interval — not only distribution functions of the azimuth $P(\psi)$ (as suggested by Mercier 1986) but also the velocity $P(v)$.

Ideally, when a "frozen" interference pattern travels in a given direction without changing its shape (specifically, a single travelling plane wave), the phase derivatives $\phi'_t(t)$, $\phi'_y(t)$ and $\phi'_x(t)$ would be either in phase or in anti-phase. Then the

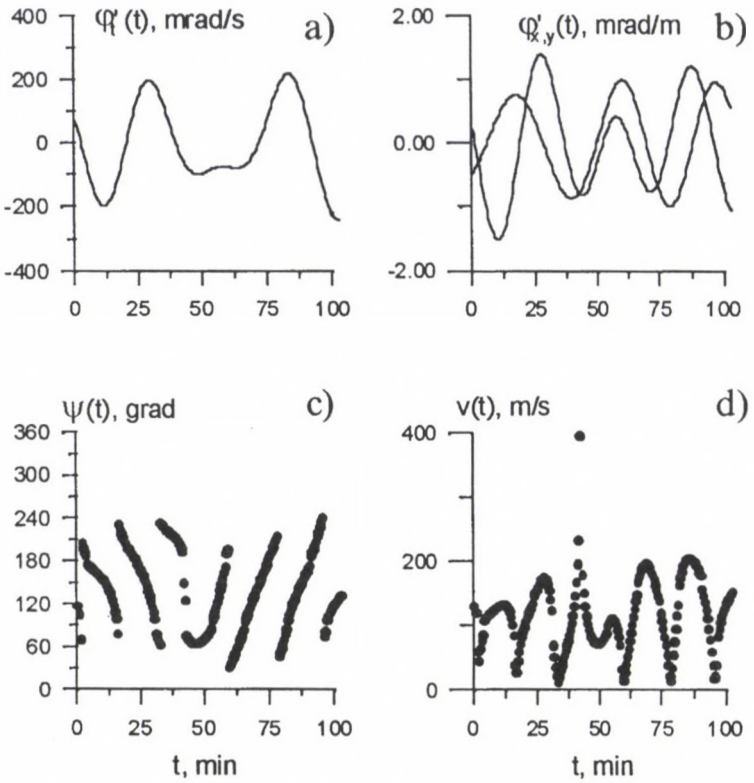


Fig. 1. Variations in space-time characteristics of the phase from 1000 LT to 1200 LT on November 10, 1990. a – $\phi'_t(t)$; b – $\phi'_x(t)$, dashed curve; $\phi'_y(t)$, solid curve; c – $\psi(t)$; d – $v(t)$

transformations (2) give a time-constant value of the azimuth and velocity, while distribution functions show a well-defined maximum at these values.

In real situations, there exist a plethora of factors that leads to more composite distribution functions. Primarily, this can be due to the finite width of the spectrum of directions of TID wave vectors. But the main difference of our proposed method from existing techniques is that its area of application is not limited to some generally accepted model for phase or amplitude perturbation of the interference pattern (for example, a single travelling plane wave). On the contrary, analyzing the distributions facilitates a choice between different models and makes it possible to assess the degree of their validity. Primarily, velocity and direction distributions are constructed and analyzed to verify the hypothesis of whether there is a predominant displacement direction of the interference pattern. If it exists, then the pattern can be considered to be travelling, and the mean travel velocity can be determined from the velocity distribution.

Figure 1 presents typical (of noon) two-hourly variations of the space-time characteristics of the phase, velocity and direction (a – $\phi'_t(t)$; b – $\phi'_x(t)$, dashed curve;

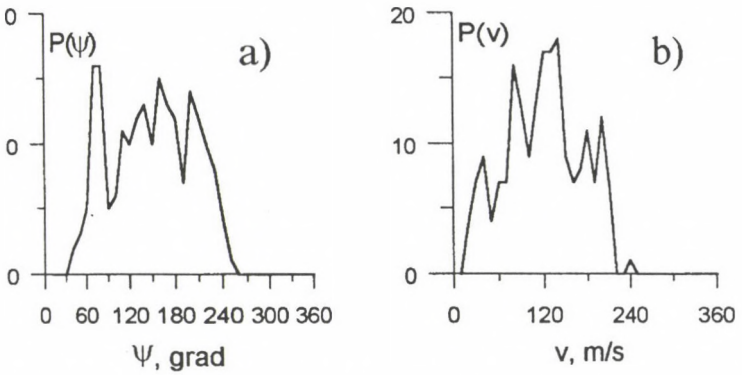


Fig. 2. Distributions of azimuths $P(\psi)$ – (a) and velocities $P(v)$ – (b), corresponding to the dependences $\psi(t)$ and $v(t)$ Fig. 1

$\phi'_y(t)$, solid curve; c – $\psi(t)$; and d – $v(t)$). Distributions of azimuths $P(\psi)$ and velocities $P(v)$, corresponding to the dependences $\psi(t)$ and $v(t)$ of Fig. 1, are shown in Fig. 2. The mean value of the velocity, v, for the daytime 2-hour interval is about 120 m/s and the mean value of the azimuth , ψ, is about 150°.

3. Determining the velocity vector of the amplitude diffraction pattern

An example of a typical 6-minute recording of amplitude scintillations from the outputs of three antennas A, B, and C for the ETS-2 signal is given in Fig. 3 (recorded October 6, 1990, beginning at 0749 LT). The figure illustrates a good likelihood of scintillations at the spaced antennas, with well-defined time shifts caused by the movement of a strongly anisotropic "frozen-in" SSI-induced amplitude diffraction pattern.

A multitude of independent counts of space-time derivatives of the signal phase in terms of the proposed method can be obtained with the spaced-antenna reception method in the frequency domain when a complex spectrum of scintillations is analysed for each of the antennas, i.e., the amplitudes $S_A(\omega)$, $S_B(\omega)$ and $S_C(\omega)$ and phase $\Phi_A(\omega)$, $\Phi_B(\omega)$ and $\Phi_C(\omega)$ spectra, where ω is angular frequency.

In this case, the phase derivatives ϕ'_t, ϕ'_x and ϕ'_y are now determined not as functions of time t, as is the case with radio interferometry (Section 2), but as a function of frequency, ω.

$$\begin{cases} \phi'_t &= \omega \\ \phi'_x &= (\Phi_C - \Phi_B)/dx \\ \phi'_y &= (\Phi_A - \Phi_B)/dy\,. \end{cases} \tag{3}$$

Corresponding values of the velocity $v(\omega)$ and direction $\psi(\omega)$ are calculated by formulas (2).

Figure 4 presents frequency dependencies for space-time characteristics of the phase of the ETS-2 radio signal scintillation spectrum, corresponding to the record in Fig. 3: a – is the amplitude spectrum, $S(\omega)$, of channel A (solid curve) and the

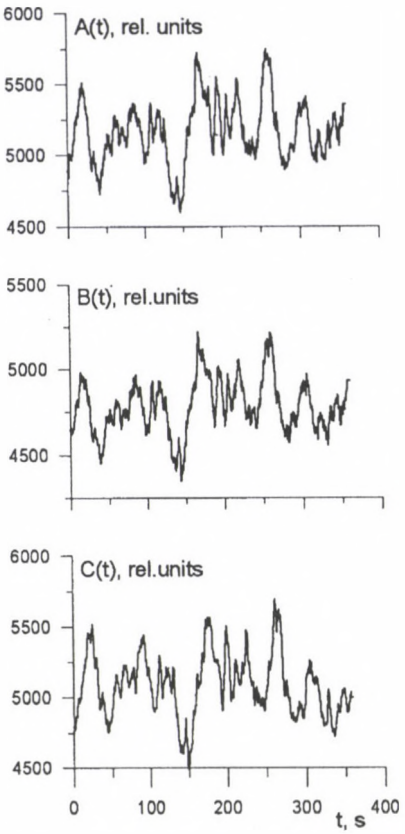


Fig. 3. An example of a typical 6-minute recording of the ETS-2 signal scintillations from the outputs of three antennas A, B, C (sequentially from top to bottom) recorded on October 6, 1990, beginning at 0749 LT. Only varying parts of signals are shown. Linear scale; amplitude of variations is about 0.1 of mean value

phase time derivative, $\phi'_t(\omega)$, the broken curve; b – spatial phase derivatives $\phi'_x(\omega)$ and $\phi'_y(\omega)$; d and c – are the azimuth, $\psi(\omega)$, and velocity $v(\omega)$ moduli, of the phase front propagation. As expected, with a displacement of the "frozen" interference pattern, like Fig. 3, $\phi'_x(\omega)$ and $\phi'_y(\omega)$ are almost proportional to the frequency, and all components of the scintillation spectrum travel in the predominant direction with an almost constant phase velocity, v. In this case, the results obtained are consistent with data from classical correlation and dispersion analysis (Briggs 1968).

Distributions (corresponding to Fig. 4) of the azimuths $P(\psi)$ – (a) and velocities $P(v)$ – (b) are given in Fig. 5. $P(\psi)$ shows a well-defined maximum at ψ of about 220° and $P(v)$ shows a maximum at v of about 30 m/s. Where necessary, one can construct "weighted" amplitude distributions in directions, $P_s(\psi)$, and velocities, $P_s(v)$, which make it possible not only to take into account the number of cases corresponding to a given interval of distributions, but also the weight of each velocity

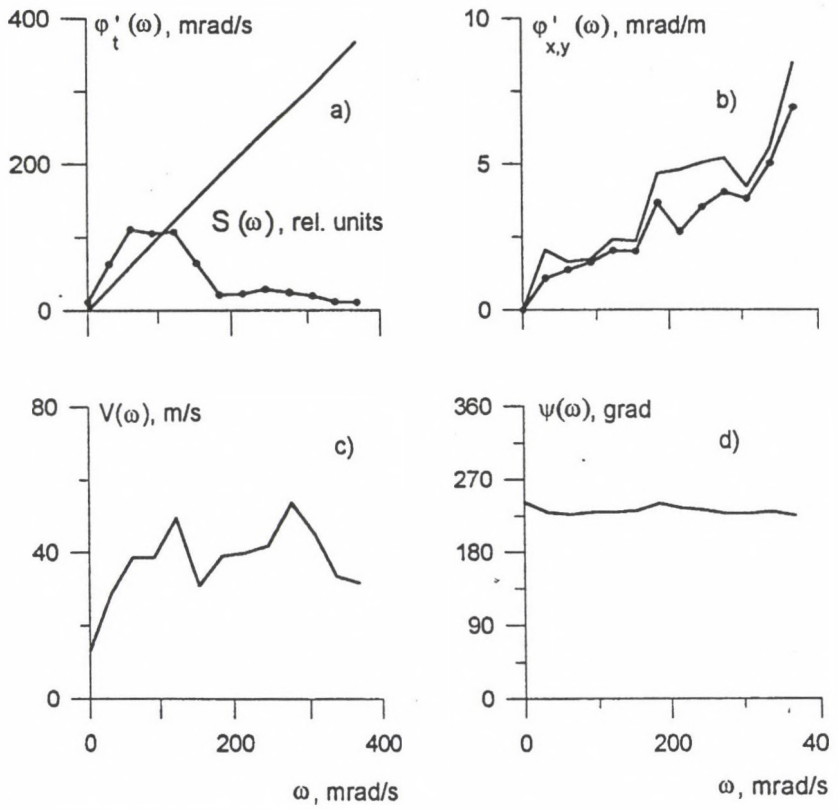


Fig. 4. Frequency dependences of space-time phase characteristics of the scintillation spectrum, corresponding to the record in Fig. 3: a – is the amplitude spectrum, $S(\omega)$, of channel A (solid curve) and the phase time derivative, $\phi'_t(\omega)$, is the broken curve; b – is the spatial phase derivatives $\phi'_x(\omega)$ – solid curve; $\phi'_y(\omega)$ – broken curve; c and d – are the velocity $v(\omega)$ moduli and azimuth, $\psi(\omega)$

vector using the amplitude spectrum, $S(\omega)$. Relevant distributions are presented in Fig. 5 ($P_s(\psi)$ – (c) and $P_s(v)$ – (d)). As would be expected, the "simple" and "weighted" distributions are almost the same for the "frozen" interference pattern.

4. Conclusions

1. A statistical method is proposed to determine the characteristics of the dynamics of the radio signal interference pattern in the plane of the receiving antenna system by measuring variations of the phase derivatives with respect to time and spatial coordinates as a function of time or frequency. These data are used to calculate corresponding values of the velocity vector as a function of time or frequency.

2. Subsequently, velocity and direction distributions are constructed and ana-

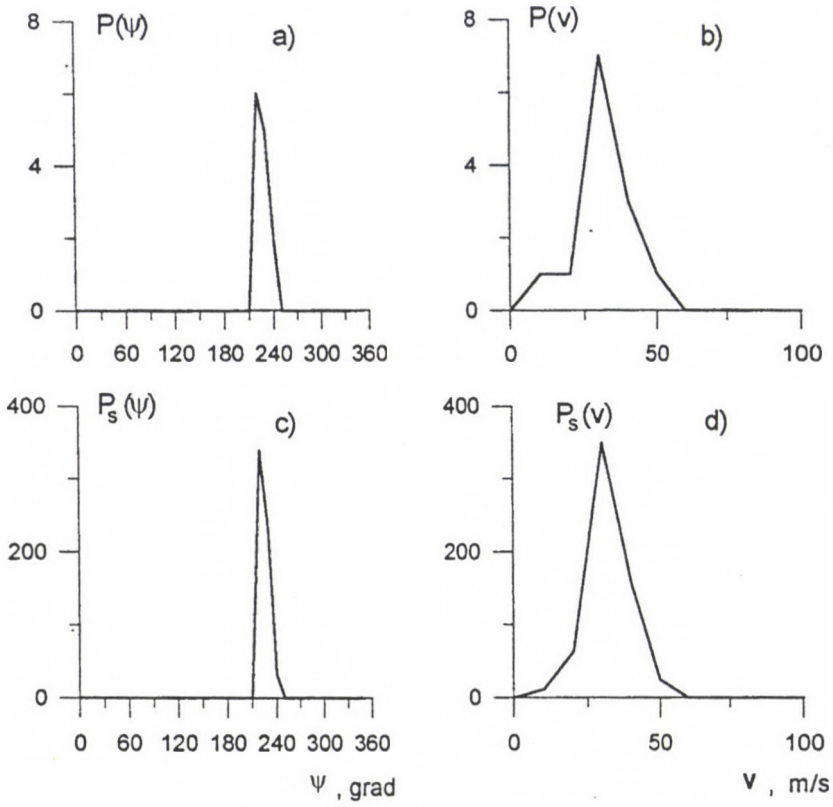


Fig. 5. Distributions of azimuths $P(\psi)$ – (a) and velocities $P(v)$ – (b); weighted distributions of azimuths $P_s(\psi)$ – (c) and velocities $P_s(v)$ – (d), corresponding to the dependences $\psi(\omega)$ and $v(\omega)$ in Fig. 4

lyzed to verify the hypothesis of the presence of a predominant displacement. If it exists, then the pattern can be considered to be travelling, and the mean travel velocity can be determined from the velocity distribution.When the signal is processed in the time range, this means that a consistent displacement of the "frosen" interference pattern is proceeding throughout the entire analysis interval; when analyzing in the frequency domain, this means that all components of the signal spectrum travel in the predominant direction. In the latter case, results using this method are in agreement with data from classical correlation and dispersion analysis.

3. Method may be useful both in experiments with ionosphere-reflected radio waves and in transionospheric radio sounding operations. The new avenues for remote diagnostics of TIDs are based on exploiting standard measurements of transionospheric radio signal characteristics and coordinate-time measurements using multichannel receivers of the Global Positioning System (GPS).

Acknowledgments

The author is greatly indebted to V G Mikhalkovsky and K S Palamartchouck for their assistance in preparing the English version of the manuscript. This work was supported by the Russian Foundation for Fundamental Research under Grants 96–05–64162 and .97–02–96060. We are also indebted to the Illinois University Administration (USA) for support of this work under the Agreement on Scientific Cooperation Between Illinois University and ISTP SD RAS.

References

Afraimovich E L, Zvezdin V N, Minko N P, Shapovalov A N 1991: *Radio Sci.*, 26, 1177–1198.

Briggs B H 1968: *J. Atmos. Terr. Phys.*, 30, 1777–1794.

Mercier C 1986: *J. Atmos. Terr. Phys.*, 48, 605–624.

Acta Geod. Geoph. Hung., Vol. 32(3-4), pp. 469-479 (1997)

GPS DIRECT AND INVERSE RADIOINTERFEROMETRY — NEW METHODS OF INVESTIGATING TRAVELLING IONOSPHERIC DISTURBANCES

E L AFRAIMOVICH[1], K S PALAMARTCHOUK[1], N P PEREVALOVA[1]

The new ways for remote diagnostics of Travelling Ionospheric Disturbances (TIDs) are based on exploiting standard measurements of transionospheric radio signal characteristics and coordinate-time measurements using dual-frequency multichannel receivers of the Global Positioning System (GPS). Signals from GPS satellites can be used following a classical radiointerferometry scheme at a network of GPS receivers located at the desired distance from each other. However, implementing a direct scheme can turn out to be an expensive and not always convenient method (when installed aboard ships, for example). We offer a new method of detecting TIDs based on the GPS, inverse radiointerferometry. The essence of this technique is that, unlike a direct scheme, all measurements are taken at a single point where the GPS receiver is located. Primary data of phase delay measurements from different satellites residing in the receiver's field of view are used to reconstruct TIDs characteristics.

Keywords: GPS; radiointerferometry; TEC perturbations; travelling ionospheric disturbances

1. Introduction

TIDs with a typical size Λ ranging from a few hundred km to a thousand km are of great interest as a matter for scientific inquiry in the ionosphere physics and as a factor that limits the accuracy of modern-day radio engineering systems used in navigation and radio interferometry; a rich variety of publications addressed the issue of TIDs. The central problem of TIDs research is to look for main geophysical factors governing the structure and dynamics of these disturbances.

One such factor is the neutral wind. Its action manifests itself in coupling of acoustic-gravity waves, with TIDs being their ionospheric response, along the propagation direction. The last two decades witnessed no decrease of interest in the study of this phenomenon. Until the present time, however, experimental research and modeling of the coupling effect has not yet reached a point where the process is trustworthily established and its origin adequately understood. At the same time the resolution of this issue would be highly instrumental in improving substantially the diagnostic capabilities of remote radio probing of the atmosphere.

Another important factor that determines the structure and dynamics of ionospheric irregularities, is the Earth's magnetic field. It is thought of as being securely

[1]Institute of Solar-Terrestrial Physics, POB 4026, Irkutsk, 664033, Russia, Fax: +7 3952 462 557, e-mail: afra@iszf.irk.ru

1217-8977/97/$ 5.00

established that the magnetic field effect is most clearly manifested in the structure and dynamics of small-scale ionospheric irregularities (SSIs) of a size typically between 0.1 and 1.0 km.

Traditionally, experimental research of the origin of two irregular features so different in size (TIDs and SSIs) has been done without considering simultaneously both types. Recently there appeared publications, showing the need to examine a wide spectrum of irregularities as a unity. New data were obtained through enhanced activities of radio astronomers investigating ionospheric irregularities that limit drastically the resolving power of modern-day high-precision radio astronomy systems. The advent of long-baseline multiantenna radio interferometers made it possible to improve markedly the sensitivity of these systems when detecting ionospheric irregularities, as well as enhancing information capabilities to reconstruct their space-time characteristics (Jacobson et al. 1995).

Of particular interest are publications by Jacobson and Erickson (1992, 1993), Jacobson et al. (1996) which reported intriguing results on magnetically-oriented irregularities (MDIs) of an intermediate size ($\Lambda \sim 10 - 100$ km).

This paper is devoted to a brief analysis of the capabilities enabled by GPS radiointerferometry techniques, including the method of inverse GPS radiointerferometry suggested by these authors. Results derived by modeling and processing data from schemes of direct and inverse radiointerferometry of TIDs at a network of GPS receivers with baselines of the order of 10–100 km will be presented in succeeding papers.

2. Classical radiointerferometry of TIDs

Figure 1a presents, according to a simplified two-dimensional scheme, the height distribution of electron density $N(h)$; TIDs in the neighborhood of the $H_{\max}$-maximum of $N(h)$ travelling with the velocity V, as well as MDI localized (according to data reported by Jacobson et al. 1996) in the plasmasphere.

A set of experimental data on the TIDs dynamics was obtained using methods of transionospheric sounding by VHF signals from geostationary satellites by measurements of the Faraday rotation of the plane of polarization proportional to the value of the total electron content (TEC) along the path between the receiver and the satellite — see a review by Davies (1980). A spaced-antenna reception scheme with a long baseline, comparable with the irregularity size (of the order of 10–100 km), was used in this case.

Further evidences of TIDs were obtained using radiointerferometers relatively recently. In classical (direct) radiointerferometry, TIDs characteristics are reconstructed based on measuring the phase difference $\Delta\phi_{m,n}(t)$ of transionospheric radio signals (in VHF or UHF frequency ranges) from cosmic point radio sources or geostationary satellites at spaced (on the Earth's surface) antennas with numbers m and n.

In Fig. 1b, on a two-dimensional scheme of classical radiointerferometry the position of the radio source and two spaced-antenna reception sites are indicated, respectively, as T, A and B. The phase values proportional to the value of the TEC

along the path between AT and BT, are marked by ϕ_A and ϕ_B. Dashed line represents the front on ground level, arrow represents the movement of the satellite along its orbit. Thus, the phase difference $\Delta\phi_{A,B}(t)$ is proportional to the component of the TEC gradient along the path between the radio source and receiving antennas A and B; therefore, the spatio-temporal properties of the phase distribution on the terrestrial surface (the anisotropy, and the propagation velocity and direction of the phase front) are interpreted as corresponding TEC characteristics.

Such experiments utilize powerful modern radiotelescopes (Mercier 1986, Jacobson and Erickson 1992, 1993), or special-purpose antenna arrays are devised (Afraimovich et al. 1991, Afraimovich 1995, Carlos and Massey 1994). In both cases these are sophisticated, expensive and unique projects, and much effort has to be directed toward implementing them, the first consideration being their time consuming nature. The main disadvantage of the analysis of VHF signals from geostationary satellites is that satellites of this class are limited in number and are nonuniformly distributed in longitude.

3. Methods of analyzing TIDs radiointerferometry data

We now consider some methods of reconstructing TIDs characteristics which were used in classical radiointerferometry and can be useful when processing GPS data. These methods are used following the required procedure of band filtering separating the signal phase variations caused by the influence of TIDs, from slow variations as a consequence of the movement of the radio source and diurnal TEC variations.

Jacobson and Erickson (1992) suggested a method which they used in experiments with cosmic radio sources at the VLA radio telescope (Jacobson and Erickson 1993) and an appropriate antenna array (Jacobson et al. 1995, 1996). These interferometers consist of a set of antennas separated by distances comparable with the size of the irregularities under investigation (10–100 km).

The method of Jacobson and Erickson (1992) is based upon solving a system of equations for complex Fourier-components $\Phi(\omega)$ of the phase differences $\Delta\phi_{m,n}(t)$ for each of m,n-pairs of the interferometer antennas. The least-squares technique is used to determine the amplitude $A(\omega)$ and the horizontal wave vector $\mathbf{k}(\omega)$ in terms of the sum of travelling plane waves in a phase perturbation

$$R(\omega) = A(\omega)\exp(-i\mathbf{kr})\,, \tag{1}$$

where $\mathbf{r}$ is a vector of the baseline between the antennas. The objective of the solution is to find such a set of main TIDs parameters (the size, and the propagation velocity and direction) which gives the best fit to measured values of $\Delta\phi_{m,n}(t)$.

In this case it is tested, to what extent the model (1) is adequate to a particular data set. It was shown experimentally that this condition is not always satisfied, especially in the nighttime (Jacobson et al. 1995). An important limitation is also the fact that the TIDs wavelength must exceed the maximum baseline size twice as a minimum.

Mercier (1986) suggested a statistical method to analyze the phase distribution obtained at the radio astronomical interferometer with a small baseline ("short" compared with Λ of TIDs). Primary data comprise time dependencies of the spatial phase derivatives $\phi'_y(t)$ and $\phi'_x(t)$ along the directions y (to the north N) and x (to the east E) determined when measuring $\Delta\phi(t)$. The method of Mercier (1986) involves the determination of a series of instantaneous values in the direction $\psi(t)$, counted from the north in a clockwise direction

$$\psi(t) = \mathrm{atan}(\phi'_x(t)/\phi'_y(t)) \tag{2}$$

and constructing subsequently, in a chosen time interval, the distribution function of azimuth $P(\psi)$. The central value of ψ is used by Mercier as estimated the azimuth of prevailing propagation of TIDs (modulus 180°).

The other method is based on analyzing the phase distribution anisotropy in the antenna array plane by determining the interference pattern contrast. In this case the ratio $C_{x,y}$ is calculated

$$\begin{array}{lll} C_{x,y} = \sigma_X/\sigma_Y, & \text{if} & \sigma_X > \sigma_Y \\ C_{x,y} = \sigma_Y/\sigma_X, & \text{if} & \sigma_Y > \sigma_X\,, \end{array} \tag{3}$$

where X and Y are series of the transformed values of $\phi'_x(t)$ and $\phi'_y(t)$, obtained by rotating the original coordinate system (x,y) by the angle β

$$\begin{array}{rcl} X(t) & = & \phi'_x(t)\sin\beta + \phi'_y(t)\cos\beta \\ Y(t) & = & -\phi'_x(t)\cos\beta + \phi'_y(t)\sin\beta \end{array} \tag{4}$$

and σ_X and σ_Y are r.m.s. of the corresponding series.

Mercier (1986) showed that it is possible to find such a value of the rotation angle β_0, at which the ratio $C_{x,y}$ will be a maximum and equal to the value of contrast C. This parameter characterizes the degree of anisotropy of the phase distribution. The angle β_0 in this case indicates the direction of elongation, and the angle $\alpha_0 = \beta_0 + \pi/2$ indicates the direction of the wave vector $\mathbf{K}$ coincident (modulus 180°) with the propagation direction of the phase front.

The method of Mercier (1986) essentially makes it possible to determine only the anisotropy and the direction of irregularity elongation of the phase distribution. The TIDs direction is not determined directly but is chosen from geophysical considerations. The modulus of the TIDs velocity is also not measured.

A statistical, angle-of-arrival and Doppler method (SADM) is proposed by Afraimovich (1995) for determining characteristics of the dynamics of the radio signal interference pattern in the plane of the receiving antenna system by measuring variations of phase derivatives with respect not only to spatial coordinates $\phi'_y(t)$ and $\phi'_x(t)$, but additionally to time $\phi'_t(t)$ — the frequency Doppler shift. This permits to ascertain the unambiguous orientation of the wave-vector's $\psi(t)$ of phase disturbance in the range $0-360°$ and determine the velocity modulus $v(t)$ of phase

front propagation at each specific instant of time by the formulas

$$\begin{aligned} W_y(t) &= \phi'_y(t)/\phi'_t(t) = |W|\cos\psi \\ W_x(t) &= \phi'_x(t)/\phi'_t(t) = |W|\sin\psi \\ v(t) &= |W|^{-1} = (W_x^2(t) + W_y^2(t))^{-1/2}. \end{aligned} \tag{5}$$

Instantaneous values of $v(t)$ and $\psi(t)$ were used in our case to construct in a selected time interval — not only distribution functions of the azimuth $P(\psi)$ (as suggested by Mercier (1986)) but also the velocity $P(v)$. The main difference of SADM from existing techniques is that its area of application is not limited to some generally accepted model for phase or amplitude perturbation of the interference pattern. On the contrary, analyzing the distributions facilitates a choice between different models and makes it possible to assess the degree of their validity. Besides, the employment of SADM is not limited by the size of the baseline between the antennas and can also be used when the baseline is larger than Λ (unlike the method of Jacobson and Erickson 1992).

4. A dual-frequency GPS receiver as a new facility and high technology for ionospheric measurements

The new ways for remote diagnostics of TIDs are based on exploiting standard measurements of transionospheric radio signal characteristics and coordinate-time measurements using dual-frequency multichannel receivers of the GPS. Using these receivers almost at any site of the globe and at any time simultaneously at two coherently-connected frequencies $f_1 = 1575.42$ MHz and $f_2 = 1227.6$ MHz high-precision measurements are made of the group and phase delay between the receiver on the ground and the transmitters on the system's satellites residing in the reception zone (usually as many as 6–8 satellites at the same time for TurboRogue SNR-8000 receiver).

Until very recently, main efforts of researchers were directed toward exploring the possibilities and developing methods of measuring and modeling the space-time distribution of TEC. Primarily, this was dictated by the necessity of a careful analysis and the need to develop methods of improving the GPS effectiveness. In addition, TEC measurement is a standard procedure in case of GPS receivers, and an appropriate data treatment does not present any serious technological difficulty.

A substantially less extensive literature is devoted to the use of data from the GPS for the investigation of ionospheric irregularities, in particular TIDs (Calais and Minster 1995, 1996). However, the sensitivity of phase measurements in the GPS allows the detection of irregularities with an amplitude of up to 10^{13} m^{-2}, or 10^{-4} of the diurnal TEC variation. However, the problem here is not the technology of obtaining data but the development of methods for the use of them.

5. Direct GPS radiointerferometry of TIDs

Signals from GPS satellites can be used following a classical radiointerferometry scheme at a network of GPS receivers located at the desired distance from each other, similarly to the Carlos and Massey (1994) radiointerferometer (Fig. 1b). In this case it is necessary to overcome a number of nonfundamental technological difficulties of data handling which are caused by taking into account the motion of GPS satellites and, in particular, by a relatively short observing time interval for a particular satellite at the reception point. But on the other hand, the advantage of direct GPS radiointerferometry of TIDs is that the reliability of determining characteristics of TIDs dynamics is improved by simultaneously using two coherent frequencies f_1 and f_2 of up to 8 or more GPS satellites, the directions toward which cover more-or-less uniformly the ionospheric volume probed.

The sensitivity of GPS measurements of TIDs is not worse than that realized by Carlos and Massey (1994), despite the fact that GPS frequencies exceed geostationary satellite frequencies by an order in the radiointerferometer setup by Carlos and Massey (1994). This is achieved in the GPS through better technology of phase measurements.

Processing data following a direct scheme of radiointerferometry, a network of GPS receivers can exploit all methods to reconstruct TIDs characteristics as mentioned in Section 3.

The method of Jacobson and Erickson (1992) is realized with slight modifications

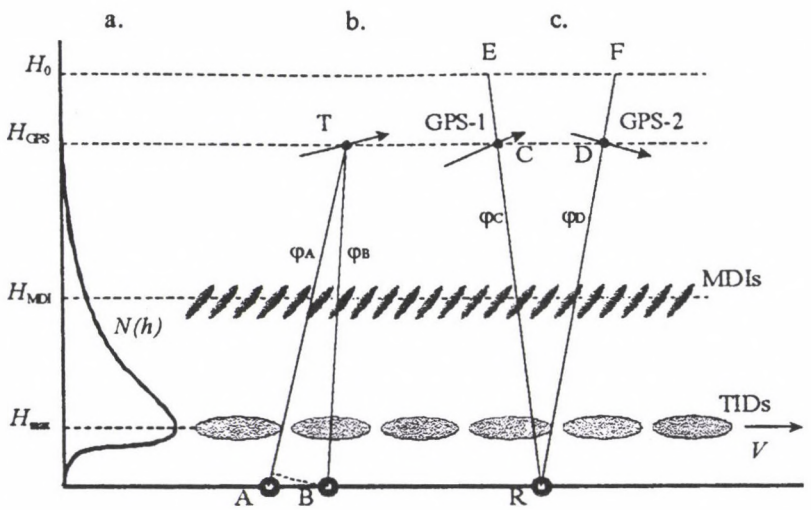


Fig. 1. a – The height distribution of electron density $N(h)$; TIDs occurring in the neighborhood of the H_{max}–maximum of $N(h)$ and travelling with the velocity V, as well as MDI located (according to data reported by Jacobson et al. 1996) in the plasmasphere. b – Simplified two-dimensional scheme for the position of two receivers A and B, and one GPS satellite T, in direct radiointerferometry scheme. c – Simplified two-dimensional scheme for the position of the receiver and two GPS satellites GPS-1 and GPS-2 labeled R, C and D, respectively, in inverse radiointerferometry scheme

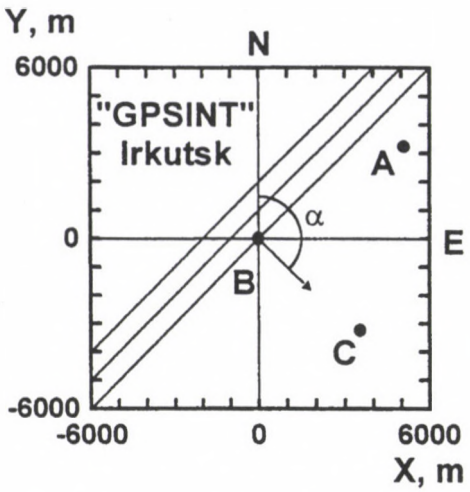


Fig. 2. Geometry of the GPS-interferometer GPSINT. A, B, C – reception points spaced at about 3–5 km where dual-frequency multichannel GPS-receivers are installed. Parallel lines represent the lines of equal TEC according to simple model; the arrow shows the direction of a normal to these lines; and N and E are the directions to the north (N axis) and to the east (E axis)

(taking into account the motion of satellites). Applying the method of Mercier (1986) and Afraimovich (1995), in Eqs (2) and (5) the derivatives $\phi'_y(t)$ and $\phi'_x(t)$, are determined from a set of the phase differences $\Delta\phi_{m,n}(t)$ for each of the m, n-pairs of receivers, which makes it possible to increase the statistics of measurements and improve their reliability.

However, using a direct scheme can turn out to be an expensive and not always convenient method. This applies specifically for territories where the number of GPS receivers, which can be used for research purposes is still small, and if they are installed on moving platforms (vessels, aircraft or low-orbit satellites).

The use of Mercier's method (1986) is exemplified by an investigation of travelling ionospheric disturbances from first measurements of TEC and its gradients during the solar eclipse of March 9, 1997, obtained with the GPS-radio interferometer "GPSINT" at Irkutsk. The interferometer consists of three receivers (one TurboRogue SNR-8000, and two Ashtech Z-12) located at the vertices of a triangle and spaced at about 3–5 km.

The geometry of the GPS-interferometer "GPSINT" at Irkutsk is presented in Fig. 2. Here A, B, and C are the reception sites. Parallel lines show arbitrarily the lines of equal TEC accordingly to simple model; the arrow shows the direction of a normal to these lines; and N and E are the directions to the north (y axis) and to the east (x axis).

To implement the method of Mercier (1986), described in Section 3, the continuous series $\Delta\phi_{AB}(t)$ and $\Delta\phi_{CB}(t)$, that are proportional to the horizontal component of the TEC gradients, were filtered out in order to eliminate diurnal trends in the measured quantities, including those introduced by changes in the satellite position

and variations in the regular ionosphere. Investigating TIDs use was made of the eighth-order Butterworth filter with a bandwidth from 10 to 30 min that specifies the range of medium scale TIDs periods of our interest. This filtering range was chosen such that it is within the commonly accepted range of TIDs periods determined by acoustic-gravity waves (AGW).

The filtered-out series $\Delta\phi_{AB}(t)$, and $\Delta\phi_{CB}(t)$ were used to determine the variations induced by the TIDs in the phase derivatives $\phi'_x(t)$, and $\phi'_y(t)$ from spatial coordinates directed, respectively, eastward E and northward N.

In practice, instantaneous values of ψ, determined every 30 s by (2), were used in our case to construct — in an about 2–4 hours time intervals — the distribution function of the azimuth $P(\psi)$, as suggested by Mercier (1986).

Figure 3 presents variations of direction $\psi(t)$ obtained from PRN14 and PRN25 data for March 9, 1997 (a — from 0824 to 1116 LT; b — from 0646 to 1154 LT).

Distributions of azimuths $P(\psi)$, corresponding to the dependencies $\psi(t)$ from Fig. 3, are shown in Fig. 4 (c — PRN14; d — PRN25). $P(\psi)$ is characterized by a pronounced maximum at a mean value of $< \psi >$ of about 110° and standard deviations of about 30° for PRN14 and of about 93° and standard deviations of about 36° for PRN25.

Figure 4 presents also the azimuthal dependence of the ratio of dispersions $C_{x,y}(\psi) = (\sigma_x/\sigma_y$ for the series $\phi'_x(t)$, and $\phi'_y(t)$, corresponding to a — PRN14 and b — PRN25. Full lines denote the values of contrast C and the azimuth α_0 corresponding to it. Maximum values of C are, respectively, 3.5 and 2.9. The directions α_0 are, respectively, 120° and 100° and almost coincide with the directions ψ that are determined by the $P(\psi)$-distributions Fig. 4c and Fig. 4d.

6. Inverse GPS radiointerferometry of TIDs

We offer a new method of detecting TIDs based on the GPS, inverse radiointerferometry. The essence of this technique is that, unlike a direct scheme, all measurements are taken at a single point where the GPS receiver is located.

In the simplified two-dimensional scheme in Fig. 1c the position of the receiver and two GPS satellites GPS-1 and GPS-2 are labeled R, C and D, respectively. To reconstruct TIDs characteristics, we use primary data of measurements of the phase delay $\phi_C(t)$ and $\phi_D(t)$ for each of frequencies f_1 and f_2 of the satellites GPS-1 and GPS-2 residing in the receiver's field of view. Despite the fact that, unlike the direct scheme, emission frequencies f_1 and f_2 of different satellites are not connected coherently, the high stability of reference oscillators of the satellites and their phasing technology in the GPS permit them to be considered quasicoherent, with a sufficient accuracy for our purposes. Therefore, when reconstructing TIDs characteristics, it is possible to use variations in phase differences $\Delta\phi_{m,n}(t)$ measured simultaneously from all n- and m-satellites observed by the receiver, and filtered out slowly varying components caused by the movement of the satellites, and diurnal TEC variations.

In the proposed method of inverse interferometry, the scheme of transionospheric sounding is reversible. One may assume that the transmitter is located at the point R on the ground, and the spatially separated receivers are at the points C

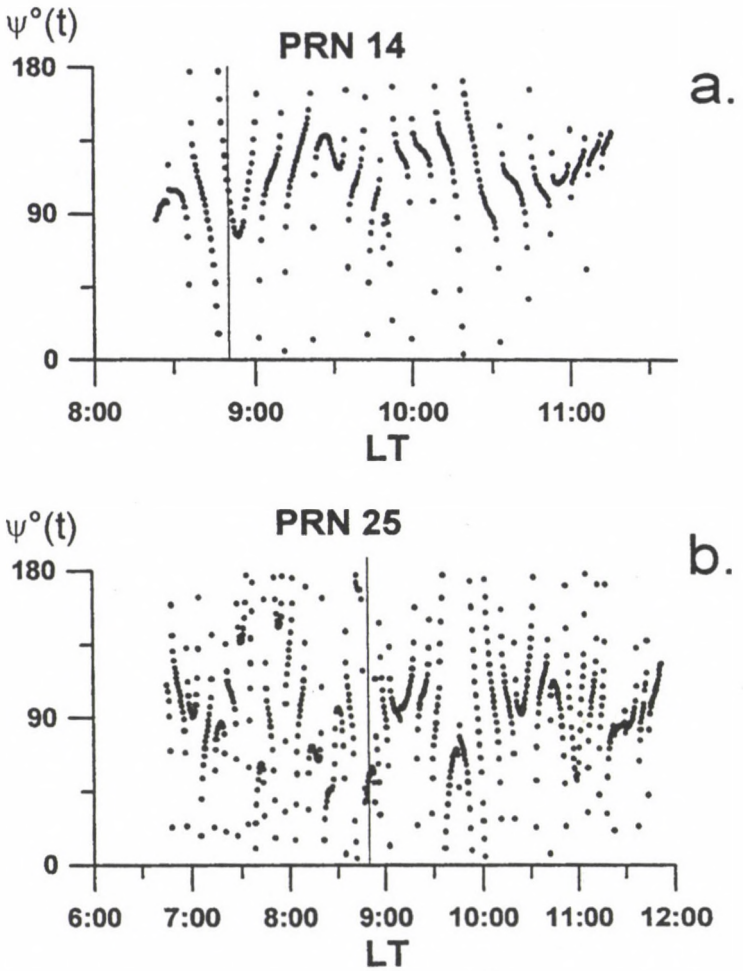


Fig. 3. Variations of direction $\psi(t)$ obtained from PRN14 and PRN25 data for March 9, 1997 (a – from 0824 to 1116 LT; b – from 0646 to 1154 LT). Vertical bars represent moment of maximal phase of exlipse in Irkutsk

and D. The phase differences $\phi_C(t) - \phi_D(t)$ are easy to convert into the phase differences $\Phi_E(t) - \Phi_F(t)$ between the points E and F on a plane where an imaginary radiointerferometer with the baseline EF is located "behind the ionosphere" at altitude H_0 higher than the GPS satellite orbit height. For both frequencies f_1 and f_2

$$\Phi_E(t) - \Phi_F(t) = \phi_C(t) - \phi_D(t) + \delta\phi_C - \delta\phi_D + \phi_{0,C} - \phi_{0,D}\,, \tag{6}$$

where $\delta\phi_C = CE \times f/c$; $\delta\phi_D = DF \times f/c$; and $\phi_{0,C}$ and $\phi_{0,D}$ are unknown initial values of the phases, c – light velocity. The distances RC, CE, RD and DF are determined directly from navigation information from the GPS receiver output.

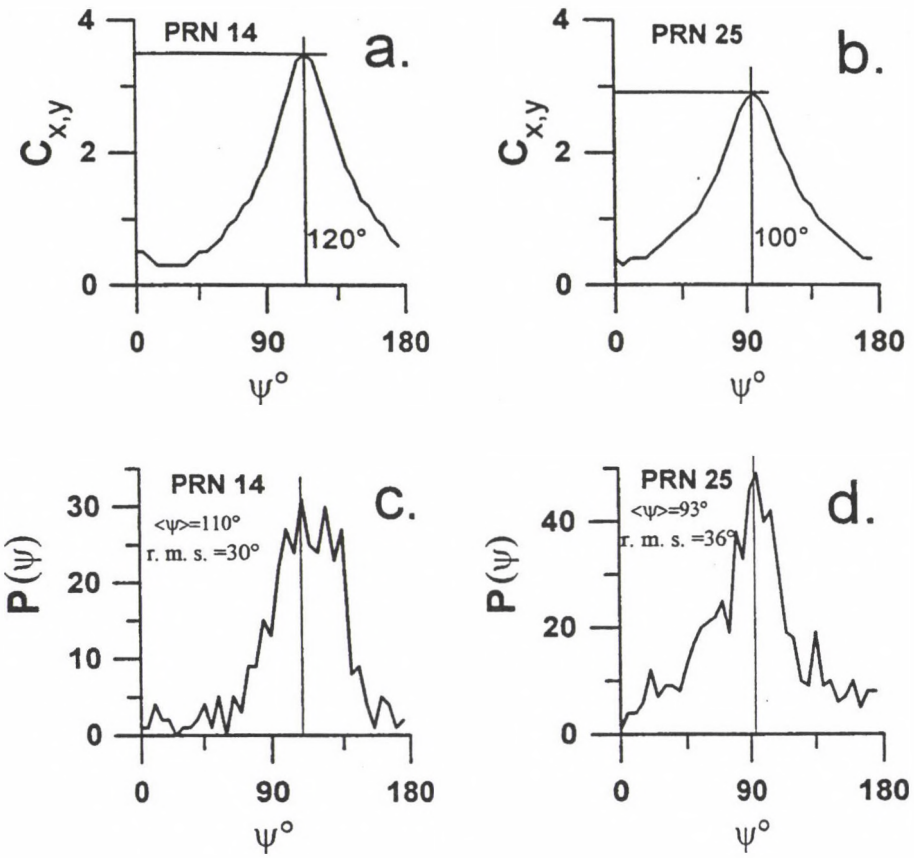


Fig. 4. Azimuthal dependencies of the ratio of dispersions $C_{x,y}(\psi) = \sigma_x/\sigma_y$ for the series $\phi'_x(t)$ and $\phi'_y(t)$, corresponding to (a) – PRN14 and (b) – PRN25. Lines denote the values of contrast C and the azimuth α_0 corresponding to it. Distributions of azimuths $P(\psi)$, corresponding to the dependencies $\psi(t)$ from Fig. 3 (c – PRN14; d – PRN25). Vertical lines denote te values of $< \psi >$ for PRN14 and PRN25

Thus, when implementing the methods outlined in Section 3, we suggest that, instead of a set of phase differences $\Delta\phi_{m,n}(t)$ for each of m, n-pairs of antennas in the inverse interferometry scheme, a set of phase differences $\Delta\phi_{m,n}(t)$ for each of m, n-pairs of GPS satellites should be used.

In the direct method, the distance to the satellite exceeds greatly the height of TIDs; therefore, the TEC perturbation "shadow" moves on the ground virtually with the same velocity and in the same direction as the disturbances themselves. In the inverse interferometry scheme, the TEC perturbation "shadow" moves on the surface where the imaginary interferometer is located, though in the same direction as the TEC perturbations, but with a larger velocity. In this case, when estimating the velocity modulus, it is necessary to take into account the geometry of transionospheric sounding, which presents no serious problem.

Acknowledgements

Authors thanks V G Mikhalkovsky for his assistance in preparing the English version of the manuscript. We are also indebted to the Illinois University Administration (USA) for support of this work under the Agreement on Scientific Cooperation between Illinois University and Institute of Solar-Terrestrial Physics, Siberian Division of Russian Academy of Sciences. This work was supported by the Russian Foundation for Fundamental Research under Grants 96-05-64162 and 97-02-96060.

Rerefences

Afraimovich E L, Zvezdin V N, Min'ko N P, Shapovalov A N 1991: *Radio Sci.*, 26, 1177–1198.

Afraimovich E L 1995: Dynamics and anisotropy of travelling ionospheric disturbances as deduced from transionospheric sounding data. I – Statistical angle-of-arrival and Doppler method (SADM). Preprint ISTP SD RAN, No. 5–95.

Carlos R C, Massey R S 1994: *IEEE Trans. Geosci. Remote Sens.*, 32, 954–958.

Calais E, Minster J B 1995: *Geophys. Res. Lett.*, 22, 1045–1048.

Calais E, Minster J B 1996: *Geophys. Res. Lett.*, 23, 1897–1900.

Davies K 1980: *Space Sci. Rev.*, 25, 357–430.

Jacobson A R, Erickson W C 1992: *Astron. Asprophys.*, 257, 401–409.

Jacobson A R, Erickson W C 1993: *Ann. Geophys.*, 11, 869–888.

Jacobson A R, Carlos R C, Massey R S, Wu G 1995: *J. Geophys. Res.*, 100, 1653–1665.

Jacobson A R, Hoogeveen G, Carlos R C, Wu G, Fejer B J, Kelley M C 1996: *J. Geophys. Res.*, 101, 19665–19682.

Mercier C 1986: *J. Atmos. Terr. Phys.*, 48, 605–624.

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

VOLUME 32

AKADÉMIAI KIADÓ, BUDAPEST
1997

CONTENTS

Numbers 1–2

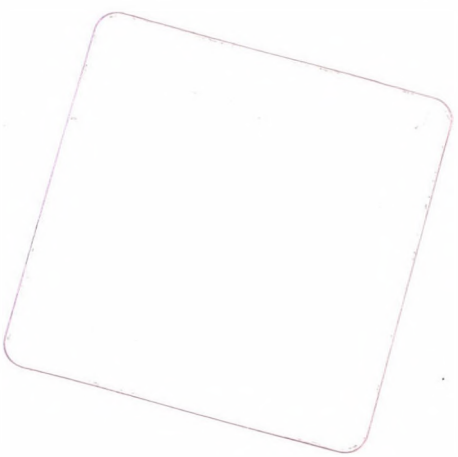
Book reviews

Numbers 3–4

New methods, experiments and ideas

AUTHOR INDEX

KEYWORD INDEX



PRINTED IN HUNGARY
Akadémiai Nyomda, Martonvásár

INSTRUCTIONS TO AUTHORS

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

Manuscripts should be typewritten in duplicate or on disk (LATEX*.tex files or Word for Windows Documents or text-only ASCII files) with a written copy, or via E-mail (LATEX*.tex files or text-only ASCII files).

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

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

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

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

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

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

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

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

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

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

5. *Legends* should be short and clear. The place of 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.

7. All the *illustrations* should contain the figure number and author's name in pencil on the reverse. The most important point with figures is clearness. Photographs and half-tone illustrations should be sharp and well contrasted. Colour photographs will be accepted, but the extra cost of reproduction in colour must be borne by the authors (approx. US$ 260 per page).

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