PN-ACM-514
79556
C5-174

TEL AVIV UNIVERSITY

FACULTY OF ENGINEERING
Department of Fluid Mechanics and Heat Transfer
P.O.Box 39040, Ramat Aviv, Tel Aviv 69978, ISRAEL
Telephone:(03)420930 Telex:342171 VERSY IL. Fax:972-3-419513

AID PROJECT DPE-G-SS-6006-00
"IDENTIFICATION OF AQUIFER HYDRAULIC PROPERTIES UNDER UNCERTAINTY"

FIFTH PROGRESS REPORT
FOR THE PERIOD JULY 1 - DECEMBER 31, 1989

1. Activity of the Israeli team.

During this period the folowing tasks, along the project plan, were carried out:

(i) The Israeli team has applied together with the Portuguese group the technique of stochastic identification of recharge and transmissivity to the Rio Maior aquifer in Portugal. Thus, one of the major aims of the project has been fulfilled. The draft of a paper incorporating the results of this activity is attached herein in Appendix 1.

(ii) Further theoretical progress has been made on the influence of aquifer boundaries upon the statistical moments of the head and transmissivity. The results are to be summarized and published soon in two papers in Water Resources Research.

2. Activity of the Portugaese team (reported by Drs. Lobo Ferreira and J. Rodriquez).

During the 6-month period starting July 1, 1988, the Portuguese team worked in the following areas:

(i) Selection of a fractured aquifer in the northern region of Portugal to analyze the possibility of applying the theory developed by the Israely team to this kind of media. To achieve this, contacts were carried out with the Portuguese A.Cavaco firm, that drilled over 300 bore holes in the region, and also with the Israeli Tahal firm that worked as a consultant. This was a project with agriculture purposes. Only a small number of the analyzed data points could be considered as being part of a unique aquifer.

(ii) The Bagueixe fractured aquifer was considered to have a reasonable set of characteristics suitable for the project. It includes 8 wells with transmissivities evaluated throught pumping tests and also 8 piezometric head points. The Portuguese team reworked the transmissivity analysis to assess the precision of the

FEB 27 1989

available results. A map of the region, with the interpolated piezometric surface, was drawn.

(iii) Another fractured aquifer was also selected in the River Ave Basin. A relative large amount of data on transmissivities is available for this aquifer. Unfortunately these were evaluated with the measurements recorded during air-lift drilling, by semi-empirical formulae.

(iv) Some new trials with the Aveiro quaternary aquifer were not successful to evaluate a non-zero integral scale of the media. This means that the variance of the computed transmissivities is of the order of magnitude of the measured transmissivity values. New piezometric data in areas not yet surveyed would be necessary to overcome this drawback.

(v) The Portuguese team cooperated on the version of the paper on the Rio Maior aquifer (Appendix 1).

3. Joint activity.

Reported under items 1 (i) and 2 (v) above.

Gedeon Dagan
Gedeon Dagan
Professor of Fluid Mechanics

APPENDIX 1

ESTIMATION OF THE HYDRAULIC PARAMETERS OF THE RIO-MAIOR AQUIFER IN PORTUGAL BY USING STOCHASTIC INVERSE MODELING

by

Y. Rubin[1], J.P. Lobo Ferreira[2], J.D. Rodrigues[2] and G. Dagan[1]

ABSTRACT

The paper describes in detail the preliminary study of a section of the Rio-Maior aquifer in Portugal. The aim of the study is to identify the transmissivity and head regional distributions and the natural recharge as a function of time. A small number of transmissivity measurements on one hand and a relative abundance of head measurements (taken over a long period of time) on the other led to the casting of the identification problem in the framework of stochastic inverse modeling. The basic assumptions of the approach which was developed previously (Dagan, 1985, Rubin and Dagan, 1987a,b) are that the logtransmissivity Y is a normal and stationary random space function, the aquifer is unbounded, and a first-order approximation of the flow equation is adopted. The expected value of the piezometric head H as well as the Y unconditional autocovariance function are supposed to have analytical expressions which depend on a parameters vector θ. The proposed solution of the inverse problem consists of identifying θ based on the model and the measurements of Y and H by employing a maximum likelihood procedure and subsequently computing the statistical moments of Y and H conditioned on the same data. The paper contains a general description of the theoretical approach as well as a detailed description of its application to the Rio-Maior aquifer

---

[1]Department of Fluid Mechanics and Heat Transfer, Faculty of Engineering, Tel-Aviv University, Ramat Aviv, Tel Aviv, 69978, Israel

[2]LNEC, National Laboratory of Civil Engineering, Avenida do Brasil 101, 1799, Lisbon, Portugal

starting from screening of the data and identification of outliers among measurements, through unconditional identification of θ and the recharge, and finally drawing of the regional maps of the conditional Y and H with their error of estimation.

1. INTRODUCTION

Modern management of aquifers, either for exploitation or for pollution control purposes, is carried out by using numerical models of flow and transport. The application of such models for predictive aims requires the knowledge of the spatial distribution of the aquifer hydraulic properties (transmissivity, storativity) as well as that of the natural distributed recharge. Then, numerical codes are able to determine the head distribution under existing or new conditions, depending on boundary conditions and wells operation. Transport models use both head and properties distributions in order to evalute velocities and solute motion throughout the formation.

The aforementioned information is obtained by field measurements, the common method being that of pumping tests. Besides, the head is measured by piezometers at different points of the aquifer. Since drilling of pumping wells and carrying out pumping tests are costly, quite often these are performed only at a few locations. In contrast, head measurement by piezometrs is much cheaper and such measurements may be available at a relatively large number of points. The determination of the spatial distribution of the aquifer properties based on this mixed type of measurements is known as the "inverse problem" (for an extensive review, see Yeh, 1986).

The usual scarcity of measurements is compounded by an additional difficulty, namely that aquifers are generally heterogeneous, i.e. their properties vary in space and often in an irregular manner. Hence, straightforward interpolation between a few measured values in order to obtain the transmissivity planar distribution is not possible. Instead, probabilistic procedures have been developed in the last few years to tackle heterogeneity. In these procedures the transmissivity is regarded as a random space function (or a regionalized variable in the geostatistical terminology), which is subjected to uncertainty. Thus, usual interpolation is replaced by kriging, which predicts the expected value and estimation variance of a regionalized variable, based on the its measured values at a number of points. As we have mentioned above, the application of kriging is generally precluded because of the scarcity of transmissivity measurements, and one has to determine the spatial distribution of the expected value and variance by solving the inverse problem. This is done, how-

ever, in a stochastic context, i.e. with both properties and head fields regarded as random and related through the equations of flow.

We have developed a procedure to solve the stochastic inverse or identification problem in the last few years. The theoretical basis has been presented in Dagan (1985), Rubin and Dagan (1987 a,b) and Dagan and Rubin (in press). The method is underlain by a few simplifying assumptions and its future application depends on its succesful implementation in a few test cases. Such implementations have been carried out for the Avra Valley aquifer (Rubin and Dagan, 1987 a,b), which was unpumped and under steady flow conditions, and for a portion of the Israeli Coastal Aquifer (Dagan and Rubin, in press), which is heavily pumped and in transient flow. In both cases measurements were relatively abundant and the method has not been used at its utmost advantage. In the present article we apply the identification method to the Rio Maior aquifer in Portugal, which is characterized by the situation described above: a small number, namely 3, of transmissivity measurements, and a much larger, namely 14, piezometers, in which heads were measured monthly, are available. Application of the method yields the following final products of the identification: evaluation of the the expected value of the natural recharge as function of time (Fig.6), maps of the expected value (Fig.8) and estimation variance (Fig.10) of logtransmissivity and maps of the expected value (Fig.11,12) and variance (Fig.13) of the head in a few months. While the maps of expected values are comparable with the usual ones obtained by deterministic procedures, those of variance reflect the uncertainty of prediction.

The aim of the study is to illustrate the implementation of a method which is believed to be of interest to further hydrological applications and to cast some light on the problems encountered when translating theory to real life cases.

2. GEOLOGY AND HYDROGEOLOGY OF THE RIO-MAIOR AQUIFER

The Rio Maior basin is located approximately 80 km north of Lisbon, on the border of the lower Tagus Cenozoic basin. It is a small graben-like depression of a rough rectangular shape running northwest to southeast. The depression is filled up with Pliocen and Quaternary sediments, with a thick layer of sands at the bottom which, in the deepest part of the basin is overlaid by a lens-shaped and very anisotropic sequence of lignites, clays, sands and diatomites, commonly referred to as the lignite-diatomite complex. These lignites have a certain economic value and they are the main justification for the studies that have been carried out at the site. The underlaying layer is a quite homogeneous body and is composed of medium to fine-grained siliceous, white sands containing around 10 percent of kaolinite. This layer outcrops in a dew places at the basin border where it is exploited for ceramic and glass industry. It also constitutes the main aquifer of the basin from where the data used in this study were obtained.

Some boreholes were drilled down to the bottom of this sand aquifer showed it to reach a thickness of about 100 m in its deepest part, and to terminate abruptly against the graben faults.

The boundary formations range from Jurassic to Miocene and they are composed mainly of siltstones, claystones, sandstones, marls and limestones. Fig. 1 shows a sketch geological map of the region and Fig. 2 presents a transversal cross-section of the basin.

Beneath lignite-diatomite complex, the aquifer is of the leaky type, as inferred from its behavior during the pumping tests. Outside this zone it is unconfined although no pumping test was carried in this area. Further information on the hydrogeology of this basin is provided in the paper by Delgado Rodrigues [1983].

The area is crossed by the small and intermitent Caniceira creek, which joins the main water course, the Maior river, in the southwestern border of the Graben basin. The relation between the Caniceira creek and the aquifer is not clearly established but no major contributions are expected to

come from this source due to the small extension of the creek over the aquifer and to its intermitent character. On the other hand, it is clearly established that the Maior river is the main discharge boundary of the aquifer and permanent flow of not less than 140 liter/sec is estimated to discharge through it. The mean annual precipitation in Rio Maior region is about 800 mm and it occurs mainly between October and April.

Possible recharge sources of the aquifer are precipitation over the unconfined areas, flow from the lateral boundaries and leakage from the upper formations. The water flowing in the overlaying complex is likely to move preferentially in horizontal directions discharging into the underlaying aquifer mostly at the border of the lens-shaped complex where it merges into the underlaying sands.

3. DEVELOPMENT OF THE THEORETICAL APPROACH

The aim of this section is to give a general, though brief, review of the major theoretical considerations taken towards the formulation of the approach. For a more comprehensive and detailed review, the reader is refered to the study of Rubin and Dagan (1987) and Dagan and Rubin (in press).

In this study we are concerned with groundwater saturated flow at the regional scale, i.e. over a medium with horizontal extent much larger than its thickness. This flow is generally described by the model equation satisfied by the piezometric head H(x,t) as follows:

$$\frac{\partial}{\partial x}\left(T\,\frac{\partial H_t}{\partial x}\right) + \frac{\partial}{\partial y}\left(T\,\frac{\partial H_t}{\partial y}\right) = -R + S\,\frac{\partial H_t}{\partial t} = -R_{ef} \qquad (1)$$

where **x**(x,y) are Cartesian coordinates in the flow domain D, t is the time, T(x,y) is the transmissivity, S is storativity, and R is a distribute source. Eq.1 pertains to untapped aquifers, whereas the presence of pumping wells requires introduction of additional term to (1).

Eq. (1) is generally supplemented by boundary and initial conditions for H, e.g. H or its derivatives are given on the boundaries of D and H is given at t=0. The solution of (1) for H, for given T, S, and R as well as given boundary and initial conditions, is known as the direct problem and it is well documented in the literature.

In the present study, dealing with the inverse problem, the given information is contained in : (i) values of measured transmissivities $T_j = T(x_j)$ at j=1,...,M points within D ; (ii) records of the head $H_{t,\ell k} = H_t(x_\ell, t_k)$ at ℓ=M+1,...,N points and at various k=1,...,P times ; (iii) boundary and initial conditions for H.

Thus we arrive at the definition of the inverse problem, which, stated simply, is the determination of the spatial distributions of T(x), R and S based on the above information, and in accordance with (1). A general review of existing solutions of the inverse problem is beyond the scope of this study, and the reader is refered to Yeh (1986).

The derivation of the model inverse solution here is based on some basic assumptions, given herein:

(i) T is regarded as a random space function of lognormal distribution, i.e. the values of $Y_j = \ln T_j$ at an arbitrary set of points x_j, form a multivariate normal vector. The stationary unconditional joint p.d.f. is characterized completely by the expected value $\langle Y \rangle$=const and the auto-covariance $C_Y(x_m, x_n) = \langle Y'(x_m)\, Y'(x_n) \rangle = C_Y(r'_{mn})$, where $r' = |r'|$ is the distance between two points. In particular we shall adopt here the exponential covariance

$$C_Y(r') = w\,[1 - H(r')] + \sigma_Y^2 \exp(-r'/I) \qquad (2)$$

where w represents an uncorrelated component of Y (a nugget), H(r') is here the Heaviside step function, σ_Y^2 is the variance of spatially correlated Y values and I is the correlation scale. Hence, Y statistical structure is represented in terms of four parameters.

(ii) We regard R_{ef} as a function of time only. These assumption is justified by the lesser space variability of R and S, generally observed in the field (cf. Freeze, 1975 for S). A scheme for the simultaneous identification of T, R and S is outlined in Dagan and Rubin (in press). The identification of S is possible for pumped aquifers, and it requires the additional knowledge of discharges of wells.

(iii) A first-order approximation is adopted for the flow equation (1). With $h(x,t) = H - \langle H \rangle$ and $Y' = Y - \langle Y \rangle$, the head and logtransmissivity residuals, respectively, a first-order approximation is adopted for H, resulting in the following two equations which replace (1)

$$\nabla^2 \langle H \rangle = - \frac{R_{ef}}{T_G} \tag{3}$$

$$\nabla^2 h = - \nabla.(Y' \nabla \langle H \rangle) \tag{4}$$

where $T_G = \exp(m_Y)$ is the geometric mean. The first-order approximation is employed since it simplifies considerably the flow equation with variable, random coefficients (1), leading to the simpler, linear forms (3) and (4). The expansion in Y' is valid in theory for small σ_Y^2 and Dagan (1985) showed indeed the first-order approximation to be valid for σ_Y^2 smaller than unity. However, in a recent study (Rubin and Dagan, inpress), the first order approximation was compared with results of the Monte-Carlo simulations of Smith and Freeze (1979), showing a good agreement even for σ_Y^2 larger than unity.

(iv) The covariances of Y and h at any two points are evaluated under the assumption of an unbounded domain. There are some justifications for this assumption. First, the joint Y,h correlation structure is dictated by the integral scale I (2), which is assumed to be much smaller than the areal extent of the flow domain. Second, in many cases boundaries locations and boundary conditions are poorly defined, and the assumption of unboundedness is as good as any. Most important, however, is that a rigorous study of the effect of boundaries on the Y,h correlation structure (Rubin and Dagan, in press) have shown it to be limited to a zone of a finite extent, of the order of an integral scale normal to the boundary, and even in this zone it is rather small.

(v) The effects of the presence of the boundaries of the flow domain on the flow regime is accounted for in an indirect, approximate manner. $\langle H \rangle$ is represented by a quadratic polynomial trend, which is a function of the space coordinates, in the form:

$$\langle H(x,t) \rangle = H_0(t) - J_0(t) \cdot x - \frac{1}{2} \sum_{i=1}^{2} \sum_{j=1}^{2} c_{ij} x_i x_j \tag{5}$$

where H_0 is the head H at the origin, J_0 is its gradient there, c_{ij} are three independent coefficients ($c_{11}, c_{12}=c_{21}, c_{22}$). Eq. (5) does not satisfy exactly eq.(1) or (3) unless R, J_0 and c_{ij} are time independent and H_0 is a linear function of time. This leads to the inherent assumptions that the head surfaces at different times are essentially parallel. Substitution of (5) into (3) results in the simple relationship:

$$c_{11}(t)+c_{22}(t) = \frac{R(t)-S(dH_0/dt)}{T_G} \tag{6}$$

generalizing the similar one for steady flow (Rubin and Dagan, 1987, eq.14). Thus, $c_{11}+c_{22}$ account for the effects of any distributed source. The additional freedom offered by the two remaining independent parameters of the quadratic trend is presumably taking into account the departure of the average flow from a uniform one, due to boundary conditions. We shall adopt the quasi-steady representation (9) even if dH_0/dt is not constant and if R is time dependent. This is accurate if these functions are slowly time-varying in time (criteria for the validity of the quasi-steady approach has been derived by Dagan, 1982)

4. A SCHEME FOR PARAMETERS IDENTIFICATION PROCEDURE.

The aim of the present section is to outline the procedure of identifying θ, the vector of parameters describing the spatial variability of Y and of the mean head surface.

Following the geostatistical approach, the aquifer is regarded as a realization out of an ensemble of statistical structure characterized by $\langle Y \rangle$ and C_Y (2). H and Y are thus regarded as space random functions, which are defined completely by their statistical moments. According to the findings of some field studies (e.g. Hoeksema and Kitanidis, 1985), Y is assumed normal, and the same is true in the first-order approximation for the residual h, which satisfies the linear equation in Y' (4). Hence, the entire statistical structure of the bivariate normal variables $Y(x)$ and $H(x',t)$ is expressed by the expected values $\langle Y \rangle$, $\langle H \rangle$ and by the unconditional covariances $C_Y(x,x') = \langle Y'(x)Y'(x') \rangle$, $C_{YH}(x,x',t) = \langle Y(x)h(x',t) \rangle$ and $\Gamma_H(x,x',t,t') = (1/2)\langle [h(x,t)-h(x',t')]^2 \rangle$, where x and x' are the vector coordinates of two arbitrary points. By unconditional moments we refer to the ensemble of formations in which the constraints imposed by measurements at the measurement points are not taken into account. Furthermore, all the above moments are for given values of S and R(t), which are regarded as independent of the space coordinates.

Under the assumptions of Section 3, we are able to write the above moments in terms of a series of parameters, by following closely the procedure of Rubin and Dagan (1987). Thus, $\langle Y \rangle$ and the given $C_Y(r')$ (2) are expressed in terms of the four parameters

$$\theta_1 = \langle Y \rangle \; ; \; \theta_2 = \sigma_Y^2 \; ; \; \theta_3 = w \; ; \; \theta_4 = I \tag{8}$$

It is emphasized that other forms of covariances rather than (2), e.g. semi-spherical, could be selected, with a somewhat different parametrization.

The expected value $\langle H \rangle$ is given in (5) in terms of the additional parameters

$$\theta_5 = J_{0,x}(t) \; ; \; \theta_6 = J_{0,y}(t) \; ; \; \theta_7 = c_{11}(t) \; ; \; \theta_8 = c_{12}(t) \; ; \; \theta_9 = c_{22}(t) \; ; \; \theta_{10} = H_0(t) \tag{9}$$

The remaining two covariances C_{YH} and Γ_H are derived by using the equation of flow (4) for h. The close form analytical solutions are given in Eqs. (16)-(21) of Rubin and Dagan (1987a) and are not reproduced here. These covariances are expressed in terms of the parameters vector θ (8,9).

In the stage of parameter identification, the objective is to identify the matrix of parameters θ_{jk} (9) where k stands for the time $t=t_k$. This is achieved by the procedure suggested by Kitanidis and Vomvoris (1983) and Hoeksema and Kitanidis (1984) and followed by Rubin and Dagan (1987), namely by a maximum likelihood procedure (MLP) applied simultaneously to the multivariate normal vector Y_j and H_{jk}, whose dimension is M + (N-M)·P. This results in large matrices to be handled in the ML process. Instead, the ML can be applied sequentially, i.e., identifying the vector θ_k (k=1,..,P) at each t_k with the aid of the head measurements at the same time only. Besides simplicity, this procedure is supported by the assumption that the flow is slowly varying in time, and the sequantial application of a quasi-steady analysis is applicable. Furthermore, when the time intervals $t_k - t_{k-1}$ are large enough, and considering measurements error, the head data at each t_k can be considered as a different and independent realization of the same stochastic process. Carrera and Neuman (1986) suggested an autoregressive autocorrelation model to account for the time dependence, leading to the same conclusion.

The estimates $\hat{\theta}$ are obtained from the minimization of the likelihood function applied to the measurement vector, leading also to the error covariance matrix Σ of $\hat{\theta}$. The matrix Σ, the inverse of the Fisher information matrix, is sometimes referred to as the Cramer-Rao lower bound. Although it constitutes the lower bound of the parameters covariances, it was found in many cases (cf., Schweppe, 1973) to be asymptotically (in N) close to the true covariances.

With given Σ, the confidence limits of the parameters can be evaluated. Small variances and intervals of confidence of the parameters are partially indicative of a correct model parametrization, and hence of the suitability of the suggested approach to the problem at hand (an indepth discussion about establishing optimal model parametrization is given in Carrera and Neuman, 1986).

At the completion of the stage of model validation, the recharge can be calculated, by employing eq.(6), after replacing dH_0/dt by its finite difference approximation $[H_{0,k+1} - H_{0,k}]/\Delta t$, where $\Delta t = t_{k+1} - t_k$, and by using prior knowledge of S, if pumping wells of known discharges are absent. By the procedure outlined in Rubin and Dagan (1987a), which follows the developments in Feinerman et al (1986), the impact of the estimation errors of θ are accounted in evaluating the estimation variance of R.

The time sequence of $\hat{R}_k$ (k=1,..,P) can supply useful information. For example, multi-seasonal water balances offer the possibility to estimate the relation between precipitation and natural recharge. If one is interested in the estimates of R, of the log-transmissivity and of the head statistical moments only, the computation may be terminated at this stage.

5. THE CONDITIONING STAGE

The unconditional moments derived previously define stochastically the ensemble of heterogeneous formations. Only in a subset of realizations do Y and H assume the measured values at the measurements points. The conditional p.d.f. of Y and H is defined as the one based on this subset, the conditioned Y and H being multivariate normal as well.

The method for computing the conditional moments of Y(x), i.e. $\langle Y^c(x)\rangle$ and $\sigma^{2,c}_Y(x)$ which relies on Gaussian conditional probability, is given in the study of Rubin and Dagan (Section 3, 1987). It is similar to cokriging and it consists of expressing Y as a linear function of the head and transmissivity residuals (with respect to the unconditional mean) at the measurement points x_i (i=1,...,N). The coefficient for each residual is a function of its relative position with regard to x_i. The coefficients are computed by solving a linear system of equations, which require the knowledge of the unconditional second moments. In general, conditioning must be carried out simultaneously for Z, the vector of all measurements at all times. However, consistent with our separation of the identification process sequentially in time, conditioning is also carried out on Z_k, measurements at

time t_k, rather then Z. The conditional mean of Y at different times t_k should be ideally the same, since Y is supposed to be time independent. Since they might differ only slightly if the model is correct, their value can be averaged over time.

Finally, the computation of the conditional variance of Y is carried out according to the method outlined in the previuos study of Rubin and Dagan (1987), and it includes the effect of parameters variances of estimation.

The computation of the conditional moments of H is very similar to that of Y. The general relationships between the conditional and unconditional moments are given in Dagan (1985) and further developed to account for parameters uncertainty in Rubin and Dagan (1987a). They can be applied directly by using the expressions derived here for the unconditional moments.

It is worthwhile to recal here the main features of conditioning or kriging : at the measurement points the expected value of Y is equal to the measured one and the variance is zero, the conditional variance is smaller or equal than the unconditional one and the conditioning by a measurement affects an area of radius of the order I around the point.

6. ON THE IDENTIFICATION OF OUTLIERS

When analyzing and processing raw data such as head measurements, one always faces the possibility that some of them might be largely erroneous. While all data are subject to measurement errors, some of them may be disinformative and misleading from a variety of reasons. An example for that might be a pieozometric head measured through a piezometer that has penetrated a zone outside the system it was primarily intended to measure, or a large error in the determination of the topographic altitude of the well head. Such measurements are called outliers. This situation is exacerbated by the fact that usually data collection and data analysis are performed by different persons, and sometimes even by different agencies. Consequently, the hydrologist usually faces the difficult task of outlining the data which are in large error, and hence might lead to erroneous in-

terpretations.

While the effect of outliers may be insignificant at the stage of identification if a large number of measurements is available, it is precisely opposite at the stage of conditioning. Indeed, a measurement has a large impact on the values of Y and H in its neighborhood.

The elimination of outliers may be facilitated by having some 'expert' opinion (gained for example through experience with similar geological formations) about how the results, in our case the areal distribution of the transmissivity, should look. Journel (1986) refers to this kind of knowledge as 'soft data', distinguishable from the 'hard data' which is comprised of direct measurements, and he also suggests a method to account for it.

In the present case study, it was suspected a-priori that some head measurements are outliers. This results mainly from the fact that the Rio-Maior aquifer is known in some parts to be comprised of several subaquifers. It was also believed that the occurence of land subsidence at some locations might have led to erroneous data reading. Furthermore, it is believed that the transmissivity variations stem mainly from changes of thickness, which have been measured independently and are not incorporated in the identification process. Two approaches were adopted, both aimed at locating and removing outliers.

In the first approach, which is intuitive in nature, the algorithm described here has been applied repetitively, eliminating each time a few head measurements. The process has been stopped and a few measurements were declared outliers when the map of expected values of Y looked acceptable in the expert's view.

The other approach, quantitative in nature, relies on the statistical structure which is attributed to the data. Consider the vector of the measurements minus their expected values, $z \simeq N(0,Q)$, which can be transformed into a vector of independent standard normal variates by following the linear transformation $z'=U^{-1}z$, where $Q=UU^T$. The covariance matrix Q can be calculated based on the identified parameters in (8) and (9). Since z'_i are independent N(0,1), it follows that

$$\sum_{i=1}^{N} z'^2_i \simeq \chi^2(N-p) \tag{10}$$

where N is the dimension of z' and p is the number of parameters fitted from the data. This test is a goodness-of-fit test which can be combined with the requirement $-2 \leq z'_i \leq 2$, for z'_i normal variates at the 95 significance level. These two tests combined, employed also by Kitanidis and Vomvoris (1984), serve as a tool for detecting outliers. The suspected outliers are expected to outcrop in the form of z'_i with large modulus. Furthermore, it is expected that upon their removal, the χ^2 test (10) will enhance the acceptability of the identified model on safer grounds. Not very surprisigly, the outliers identified through this method were those head measurements whose distance from the fitted quadratic curve (5) was the largest. While the identified stochastic model was in the acceptance region of α=0.5% significance level before removal of outliers, it was accepted at the 10% acceptance region after removal of the outliers. In this way, the probability of Type II error, i.e., the probability of accepting mistakenly the identified model was reduced considerably.

Application of both methods, carried out independently, led to the same results, i.e. they led to the identification of the same head measurements as outliers. It is emphasized, though, that there is always the possibility that measurements that are considered as outliers are indeed significant information, and their removal may lead to lose of information. Thus, although the second approach is automatic, it is by no means a replacement for a judicious consideration.

7. RESULTS

7.1 General Comments

The previously described methodology was applied to the northern region of the Rio Maior white-sand aquifer, which is, in its majority, a confined aquifer (see Section 2). For an overview we have inserted the region under study in a square which has an area of 9 km².

The aims of the study are the following: (i) evaluation of the vector of parameters Θ which models the random Y and H joint field, based on the limited transmissivity data and the sequence of head measurements taken periodically; (ii) Evaluation of the effective recharge of the aquifer; (iii) Drawing of the regional maps of $\langle Y^c \rangle$, $\sigma^{2,c}_Y$, $\langle H^c \rangle$ and $\sigma^{2,c}_H$, based on the Y and H measurement values.

7.2 The Data

The available data for this study consists of a 19 months sequential set of 18 piezometric head data measurements and also a set of (only) 3 transmissivity values.

The above data, which consists of measurements, constitutes our 'hard-core' data. In addition, there is also the so-called 'soft data' (see Journel, 1986), based on the experience gathered by the authors about this aquifer. This refers to the flat piezometric surface of this aquifer (indicating small recharge coming from the top confining aquitard), and above all to the high level of homogeneity of the hydraulic conductivity, the main heterogeneity source being the thickness variation. The incoorporation of this data in the evaluation process is described in Section 6, and summarized below.

From the initial set of 18 head measurement points a first set of 14 head points was selected. This was done because some of the initial points are located very near to each other, forming a cluster of points which do not provide additional head information. Finally, from the set of 14 points a set of 9 points was selected, by eliminating 5 supposedly outliers, as indicated in Section 6.

The final selected information is formed by the sequence of 9 head measurements on a monthly basis covering the period from January 1985 to July 1986 (19 months), and 3 transmissivity measurements. The limited information on the experimental evaluated transmissivities of the aquifer does not allow for a reliable evaluation of the regional log-transmissivity field based on the usual kriging procedures, and suggests employing the inverse method which can interpret also head measurements.

7.3 Unconditional Stage

The unconditional stage of the study was divided in the following steps: (i) Identification of the parameters of the random Y field, namely $\langle Y\rangle_i$, w_i, $\sigma^2_{Y_i}$, I_i, (i=1,..,19) and the covariance matrix, for each of the 19 months; (ii) Averaging the parameters of Y of the 19 months, namely:

$$\langle Y\rangle=\sum_{i=1}^{19}\frac{\langle Y\rangle_i}{19}\ ;\ \sigma^2_Y=\sum_{i=1}^{19}\frac{\sigma^2_{Y_i}}{19}\ ;$$
$$I=\sum_{i=1}^{19}\frac{I_i}{19}\ ;\ w=\sum_{i=1}^{19}\frac{w_i}{19}\ ; \qquad (11)$$
$$\Sigma_{11}=\sum_{i=1}^{19}\frac{\Sigma_{11i}}{19^2}\ ;$$

etc., all other $\Sigma_{k\ell}$ being $(\Sigma_{k\ell})=\sum_{i=1}^{19}\frac{(\Sigma_{k\ell})_i}{19}$. Here $\Sigma_{k\ell}$ stands for the covariance of estimation of parameters θ_k and θ_ℓ, determined by the ML procedure outlined in Section 4. The results of step (i) are showed in Figs.3, 4 and 5, which show the time series of the maximum likelihood estimators of $\langle Y\rangle$, σ^2_Y and I, respectively. It is evident that the differences between months are very small, as should be for these structural parameters. Such differences are due not only to variations of the piezometric surface of the aquifer, but also to possible head measurement errors. The small variations with time presented before strengthens the confidence in the model which renders the averaged variances of $\langle Y\rangle$, σ^2_Y, I and w.

The final results of the unconditional parameters estimates and the covariance matrix of the averaged estimators are given in Table 1. It is emphasized that the smallness of the variances of estimation reflects the accuracy of the ML identification procedure on one hand, and the effect of averaging over 19 months in Eq.11.

7.4 Recharge Evaluation

According to Table 1, the net distributed effective recharge R_{ef} from the top confining aquitard is almost zero. This is due to the fact that the parameters related to the recharge, i.e. c_{11} amd c_{22}, have small or zero values. The effective recharge R_{ef} of the aquifer is evaluated through the relationship (see Eq.6):

$$R_{ef} = R - S \frac{dH_0}{dt} = T_G \, (c_{11}+c_{22}) \tag{12}$$

where H_0 is the expected value of the piezometric head of the aquifer at the origin and S is the storativity. In our case, since c_{11} and c_{22} are very small, this expression may be reduced to:

$$R = S \frac{dH_0}{dt} \tag{13}$$

which indicates that the recharge through leakage is counterbalanced by the increase in storage. The value of S was obtained by averaging the 3 available storativity measurements (located at the same points where the transmissivities were measured). Those values of S are very close to each other, yielding the average value $S = 1.1 \times 10^{-3}$. The values H_0 were calculated for each one of the 19 months, based on the head measurements for each month.

The recharge was computed by using (13), in which dH_0/dt has been replaced by $\Delta H_0/\Delta t$, $\Delta t = t_k - t_k - 1 = 1$ month, and $H_{0,k}$ has been identified in each month by the ML procedure. Fig.6 shows the time series of the evaluated recharge for the Rio Maior aquifer, according to Eq.12. Fig.7 shows the time series of the precipitation in Rio Maior. Comparing both figures one observes a good agreement between the shapes of the recharge and of the precipitation, with a delay of about two months for the recharge.

7.5 Conditioning of the logtransmissivity Y

The conditioning of Y was done on the measurements of Y and H for each one of the 19 months (for details of the procedure see Dagan, 1985). For this purpose, use was made of the parameters of Y and of the covariance matrix averaged over 19 months, shown in Table 1. In contrast, for the head H use was made of the estimates of H_0, J_x, J_y, c_{11}, c_{22} and c_{12} from each of the 19 months, since these parameters naturally depend on head measurements. Finally, the evaluated monthly results for Y were averaged for the 19 months.

Fig.8 shows the final regional log-transmissivity map of the aquifer, based on the final averaged parameters. The similarity between this map and the map of Rio Maior's bed rock isolines, Fig.9, is emphasized. Notice that the top of the aquifer is almost horizontal and, as referred to previously, the hydraulic conductivity of the aquifer is considered to be a very homogeneous one. Thus, the transmissivity of the aquifer is strongly corelated with the thickness of the aquifer.

Fig.10 show the regional maps of $\sigma^{2,c}_{Y}$, i.e. the conditional generalized variance of Y (i.e., including the parameters estimation errors in the conditioning procedure) following the procedure of Rubin and Dagan (1987).

7.6 Conditioning of the head H

To complete the study, the conditioning of the piezometric heads of the aquifer was carried out for two different months, one in the winter time (January 1986) and the other in the summer time (July 1986).

This conditioning was done with the averaged values of Y parameters, shown in Table 1, and the values of H parameters evaluated for each one of the two referred months.

Figs.11 and 12 show, respectively, the maps of the expected values of the piezometric heads obtained for January 1986 and those obtained for July 1986. As expected the piezometric surface of the aquifer, in both cases, is rather flat, showing also little differences in the piezometric heads between

summer time and winter time.

Finally, Fig.13 presents the variances of the generalized conditional H for the month of July, 1986. It is seen that the values of $\sigma^{2,c}_{H}$ are negligible in the majority of the region under study, showing a good knowledge on the piezometric heads of this aquifer.

8. SUMMARY AND CONCLUSIONS.

The presents study demonstrates the applicability of the stochastic identification procedure developed in the past to an aquifer which is characterized by scarcity of transmissivity measurements and a larger number of piezometers which monitor the water head. It is believed that this state of affairs is typical for many real life cases, for which the use of the inverse method is definitely justified.

The results of Table 1 indicate that the logtransmissivity variance is relatively small, justifying the use of our methodology. Furthermore, it was inferred from some field observations that the hydraulic conductivity is relatively homogeneous and the spatial variation of the transmissivity (Fig.8) is related mainly to the change of thickness (Fig.9). The agreement between the two strengthens the confidence in the present methodology, which confirms this finding.

Another result of interest, again supporting the assumptions adopted from the outset, is the constancy of the parameters characterizing transmissivity which were determined independently each month (Figs.3,4,5). Large time variations of these parametrs would have indicated inadequacy of the model.

One of the results of the analysis, of considerable interest in hydrological applications, is the identification of the expected value of the distributed recharge (Fig.6). The smallnes of its value in the present case, confirms the qualitative statements about the existence of a confinement in this part of the aquifer.

The main outcome of the identification procedure, namely the map of the expected values of the logtransmissivity (Fig.8), is the basis for modeling of flow in the Rio Maior aquifer under new exploitation conditions. The map of logtransmissivity variance (Fig.10) may be used in order to assess the uncertainty and risk associated with the prediction. Similarly, the spatial distributions of the transmissivity and heads (Figs.11,12) and their variances are the starting point for evaluating the paths and spread of pollutants discharged in the aquifer under the present flow conditions.

ACKNOWLEDGEMENT Partial support by the Agency of International Development is gratefully acknowledged.

References

Carrera J., and S.P. Neuman, 1986, Estimation of aquifers parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, Water Resour. Res., 22(2), 199-210.

Dagan G.,, 1982, Analysis of flow through homogeneous random aquifers: 2. Unsteady flow in confined formations, water Resour. Res., 18(5), 1571-1585.

Dagan G., 1985, Stochastic modeling of groundwater by unconditional and conditional probabilities: The inverse problem, Water Resour. Res., 21(1), 65-72.

Dagan G., and Y. Rubin, in press, Stochastic identification of recharge, transmissivity and storativity in aquifer transient flow: a quasi-steady approach, Water Resour. Res.

Delgado-Rodrigues J., 1983, The Noordbergum effect and characterization of aquitards at the Rio-Maior mining project. Groundwater, 21(2).

Hoeksema R.J., and P.K. Kitanidis, 1985, Analysis of the spatial structure of properties of selected aquifers, Water Resour. Res., 21(4), 563-572.

Journel, A.G., 1986, Constrained interpolation and qualitative information - the soft Kriging

approach, Math. Geol., V.18, p.269-286.

Kitanidis P.K., and E.G. Vomvoris, 1983, A geostatistical approach to the inverse problem in groundwater modeling (steady-state) and one-dimensional simulations, Water Resour. Res., 19(3), 677-690.

Smith, L., and R.A. Freeze, 1979, Stochastic analysis of steady state groundwater flow in a bounded domain: 2. Two dimensional simulations, Water Resour. Res., 15(6), 1543-1559.

Rubin Y., and G. Dagan, 1987a, Stocahstic identification of transmissivity and effective recharge in steady groundwater flow: 1. Theory, Water Resour. Res., 23(7), 1185-1192.

Rubin Y., and G. Dagan, 1987b, Stocahstic identification of transmissivity and effective recharge in steady groundwater flow: 2. Case study, Water Resour. Res., 23(7), 1193-1200.

Rubin, Y., and G. Dagan, in press, Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers: 1. Constant head boundary, Water Resour. Res.

Yeh, W. W-G., 1986, Review of parameter identification procedures in groundwater hydrology: The inverse problem, Water Resour. Res., 22(2), 95-108.

Figure Captions

Figure 1 - Geologic map of the Rio-Maior lignite basin.

Figure 2 - Simplified geologic cross section of the Rio-Maior basin along line AB in Fig.1.

Figure 3 - The sequential MLP estimators of $\theta_1=\langle Y\rangle = \langle LnT\rangle$, (T in m²/day).

Figure 4 - The sequential MLP estimators of $\theta_2=\sigma^2_Y$.

Figure 5 - The sequential MLP estimators of $\theta_3=I$ (in Km).

Figure 6 - The time series of recharge (in mm/month) in Rio-Maior North as obtained by Eq.12.

Figure 7 - The time series of average precipitation (in mm/month) in Rio-Maior North.

Figure 8 - Regional contour map of the conditional logtransmissivity $Y^c(x)$ (T in m²/day).

Figure 9 - Mean annual piezometric surface and bedrock contour lines of Rio-Maior North.

Figure 10 - Regional contour map of the conditional logtransmissivity variance $\sigma^{2,c}_Y(x)$.

Figure 11 - Regional contour map of the conditional head $H^c(x)$ (in m) for January 1986 in Rio-Maior North.

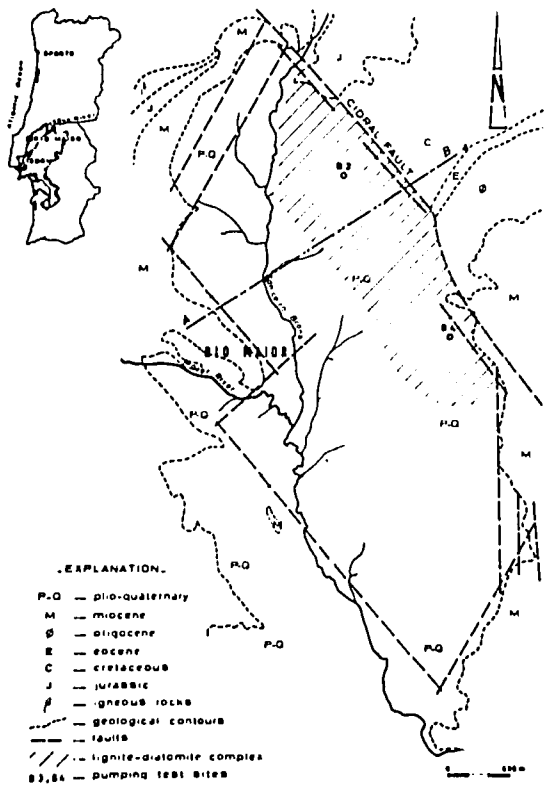
Figure 12 - Regional contour map of the conditional head $H^c(x)$ (in m) for July 1986 in Rio-Maior North.

Figure 13 - Regional contour map of the conditional head variance $\sigma^{2,c}_H(x)$ (in m²) for July 1986 in Rio-Maior North.

Parameter	$\theta_1=\langle Y \rangle$	$\theta_2=\sigma_Y^2$	$\theta_3=I$	$\theta_4=w$	$\theta_5=J_x$	$\theta_6=J_y$
	$Y=\ell n\ T$ T in m²/day		in km		m/km	m/km
Expected Value	6.2046	0.2251	0.3145	0.	0.9883	-1.0905
Variance of Estimation	0.004	0.0012	0.0035	0.0008	0.0014	0.0012

Parameter	$\theta_7=c_{11}$	$\theta_8=2c_{12}$	$\theta_9=c_{22}$
	in m/km²	in m/km²	in m/km²
Expected Value	0.	0.2243	0.0026
Variance of Estimation	0.0069	0.0037	0.0075

Table 1: Unconditional average parameters and their variances of estimation.


CIDRAL FAULT
BIO MAJOR
N
.EXPLANATION.
P-Q — plio-quaternary
M — miocene
Ø — oligocene
E — eocene
C — cretaceous
J — jurassic
β — igneous rocks
— geological contours
— faults
— lignite-diatomite complex
B3,B4 — pumping test sites

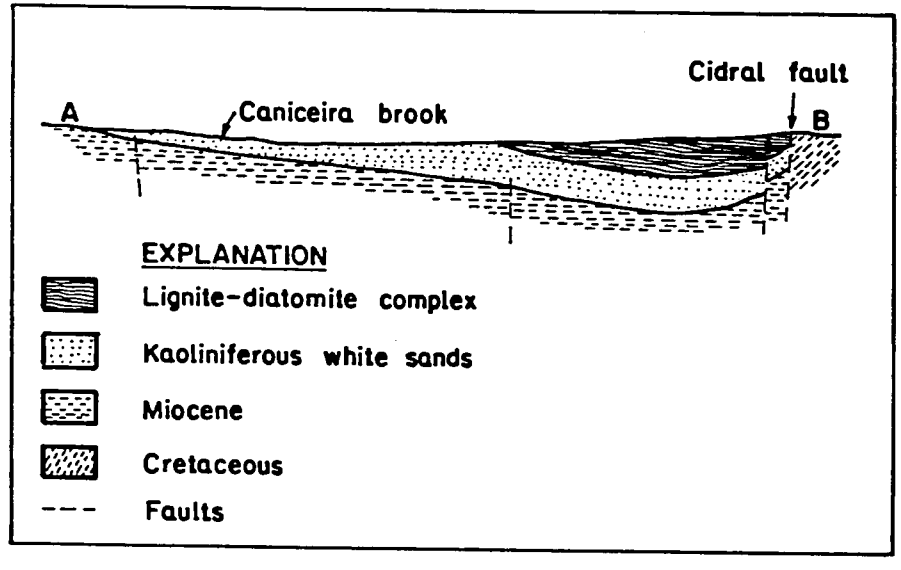

Cidral fault
A
Caniceira brook
B
EXPLANATION
Lignite-diatomite complex
Kaoliniferous white sands
Miocene
Cretaceous
Faults

Fig. 2

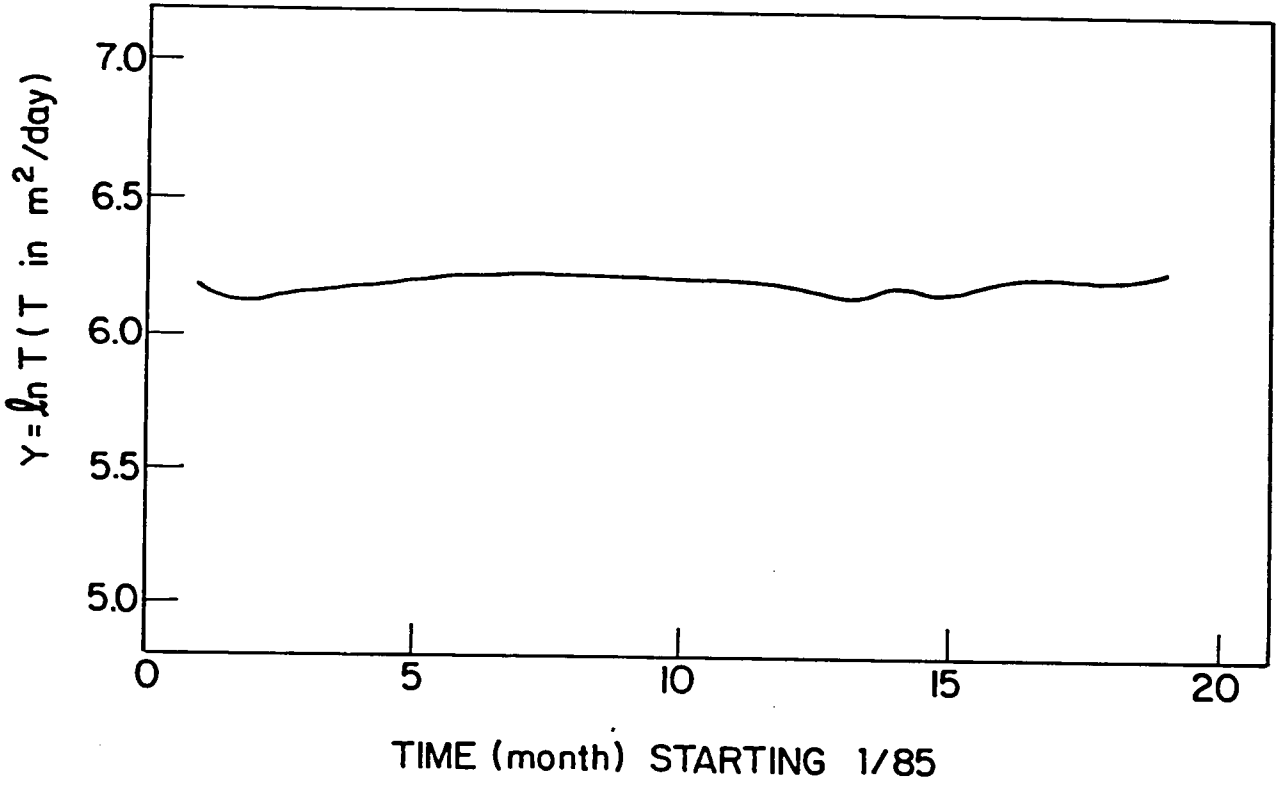

7.0
6.5
6.0
5.5
5.0
Y = ln T (T in m²/day)
0
5
10
15
20
TIME (month) STARTING 1/85

Fig. 3

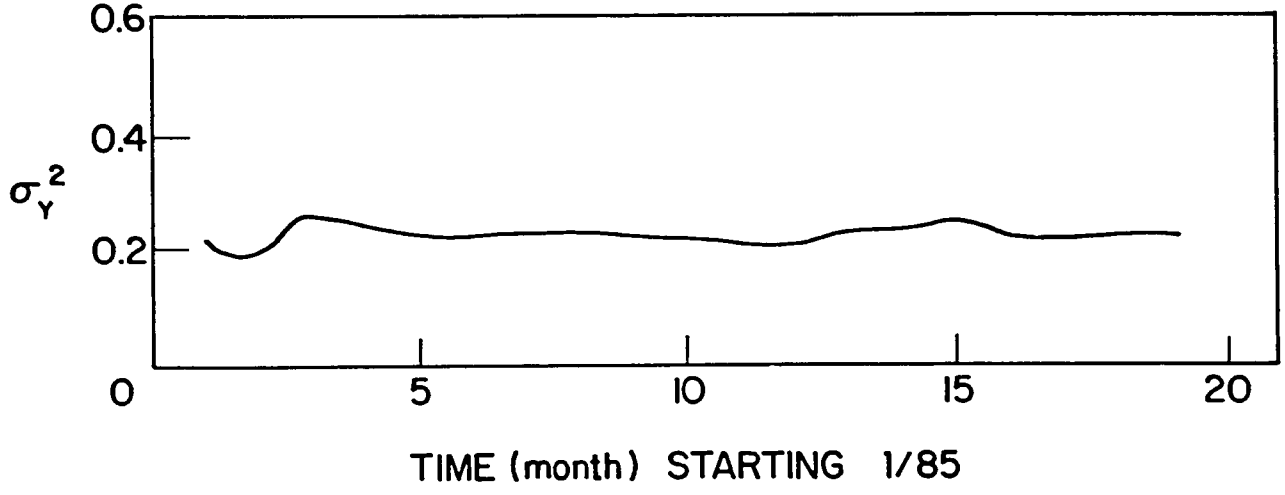


Fig. 4

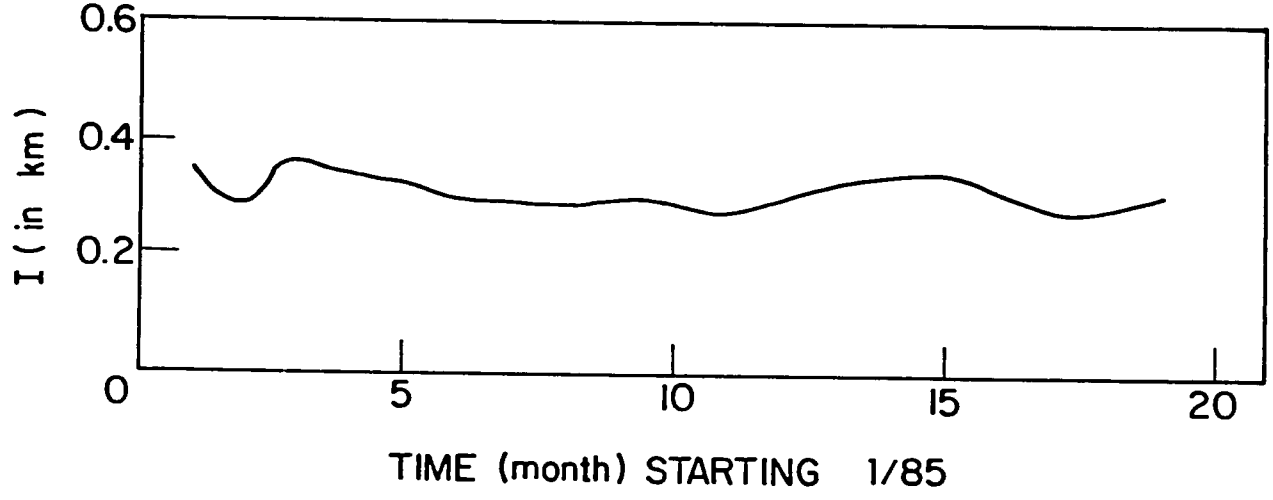


Fig. 5

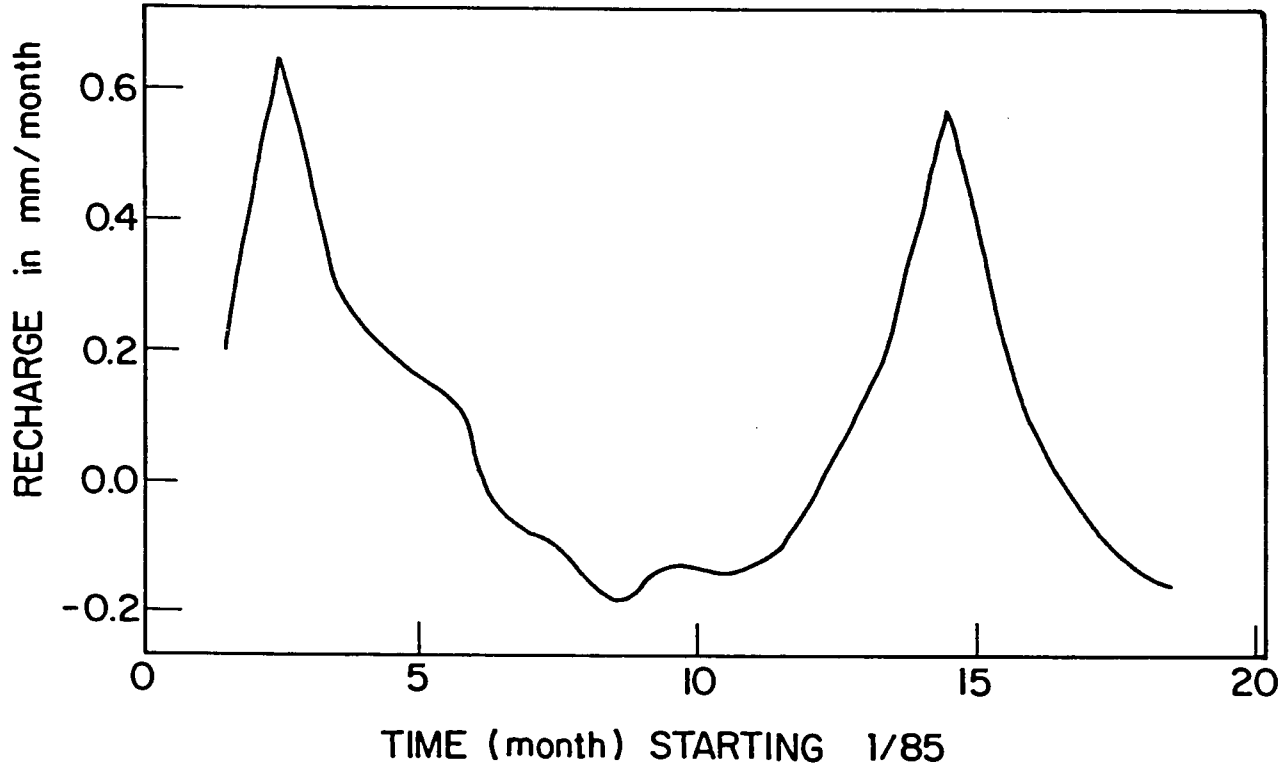

RECHARGE in mm/month
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
20
TIME (month) STARTING 1/85

Fig. 6

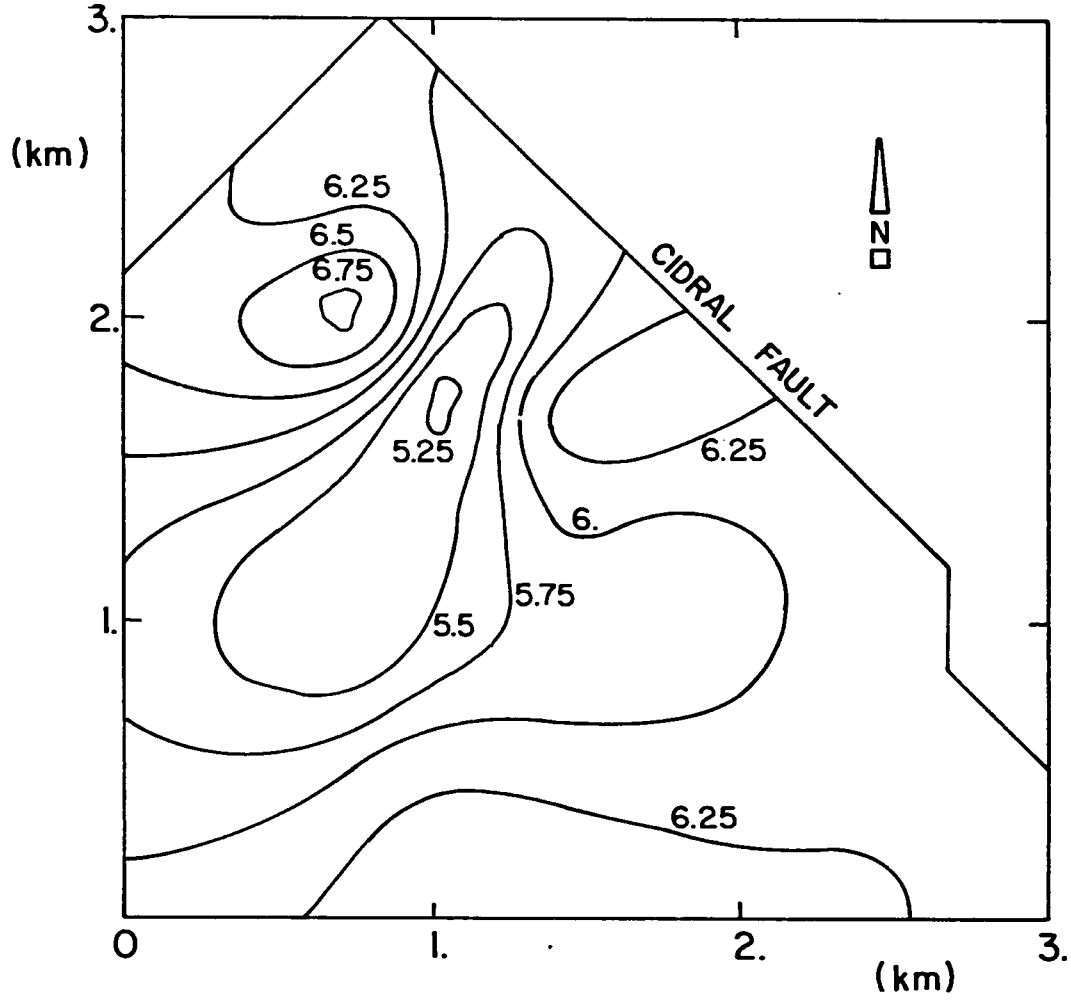


Fig. 8

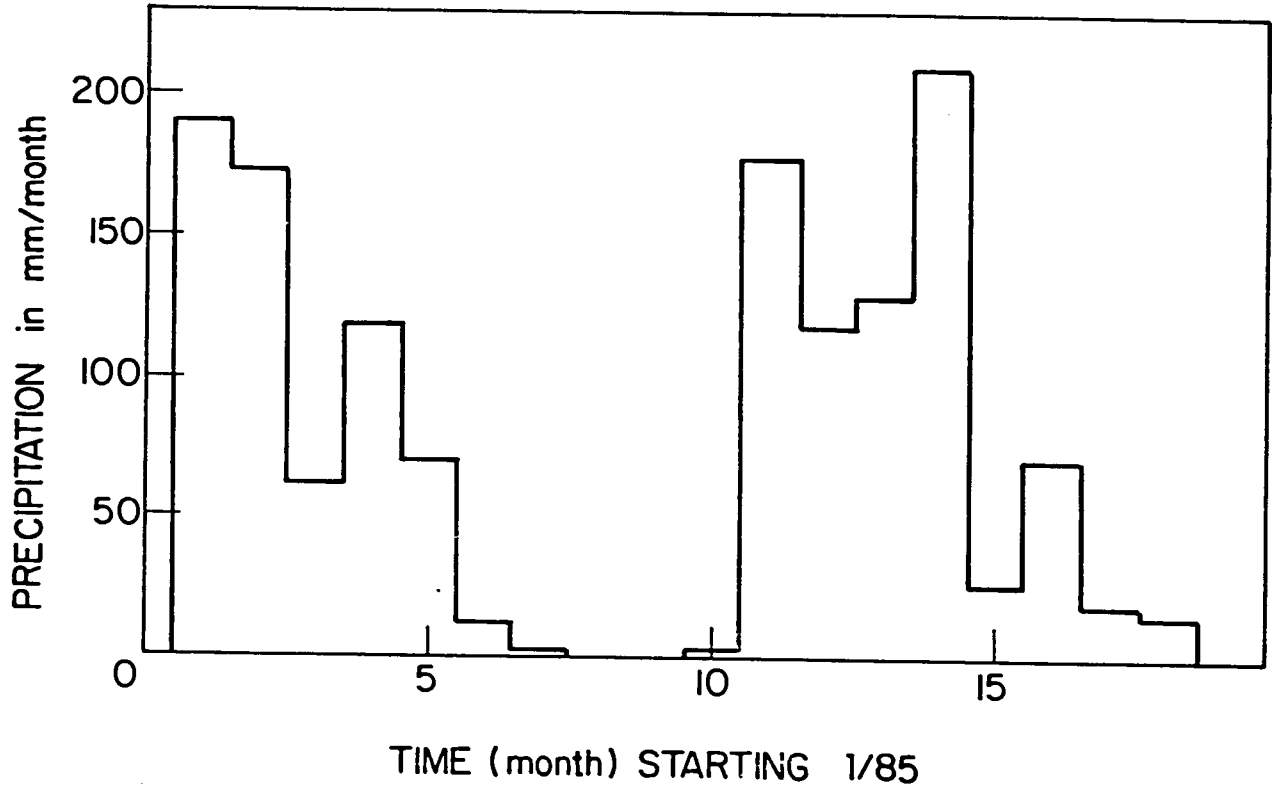

PRECIPITATION in mm/month
200
150
100
50
0
5
10
15
TIME (month) STARTING 1/85

Fig. 7

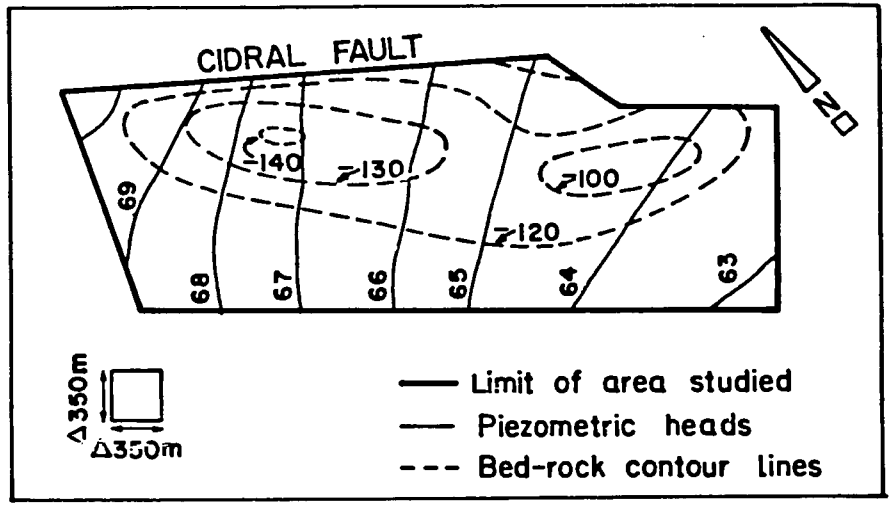

CIDRAL FAULT
N
140
130
100
120
69
68
67
66
65
64
63
Δ350m
Δ350m
Limit of area studied
Piezometric heads
Bed-rock contour lines

Fig. 9

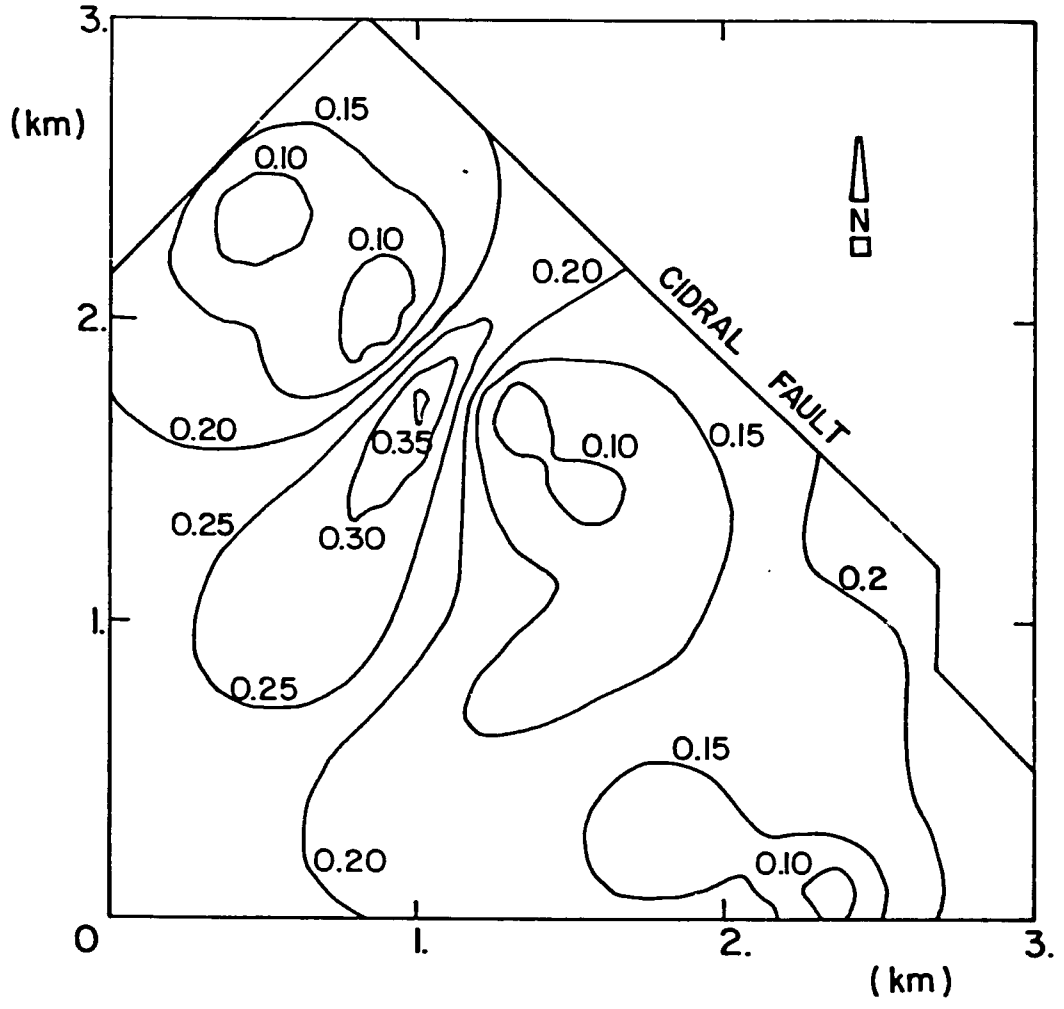

3.
(km)
2.
1.
0
1.
2.
3.
(km)
N
CIDRAL FAULT
0.15
0.10
0.10
0.20
0.20
0.35
0.10
0.15
0.25
0.30
0.2
0.25
0.15
0.20
0.10

Fig. 10

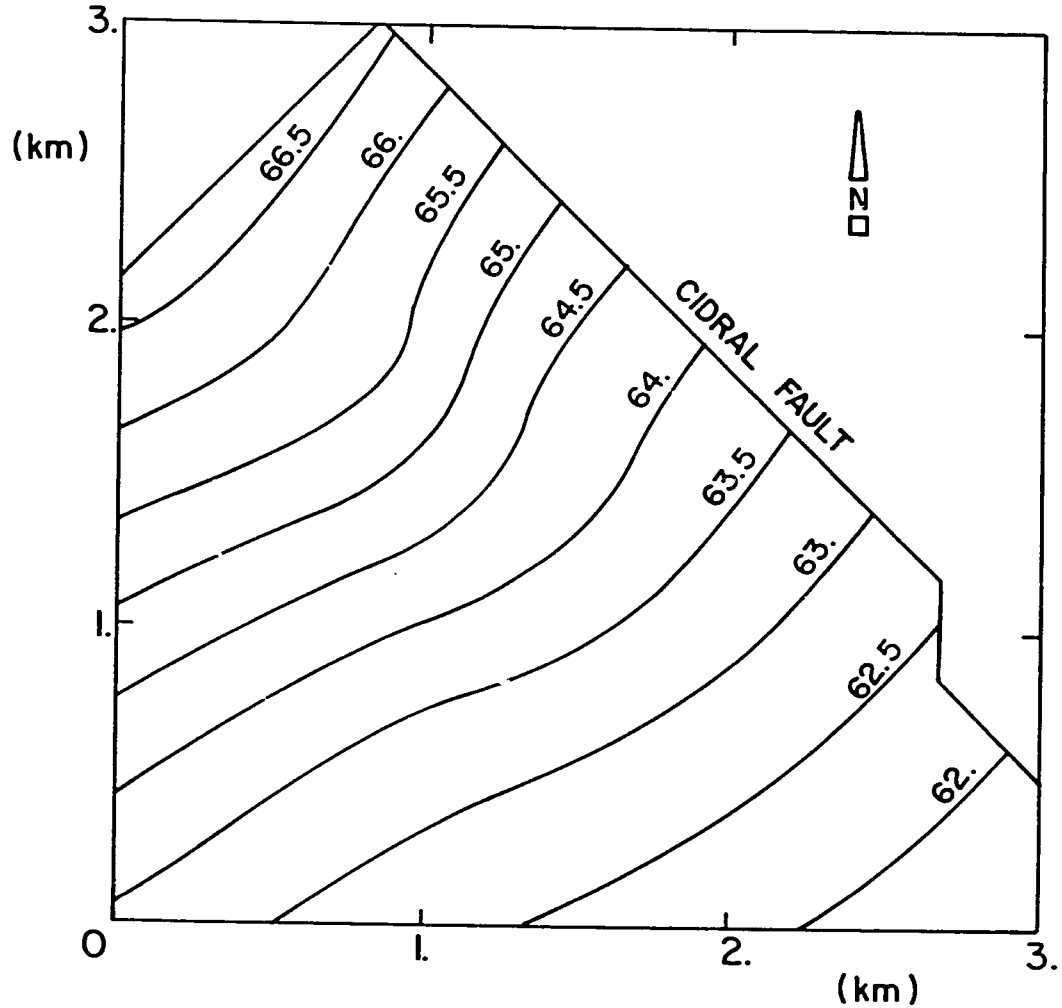


Fig. 11

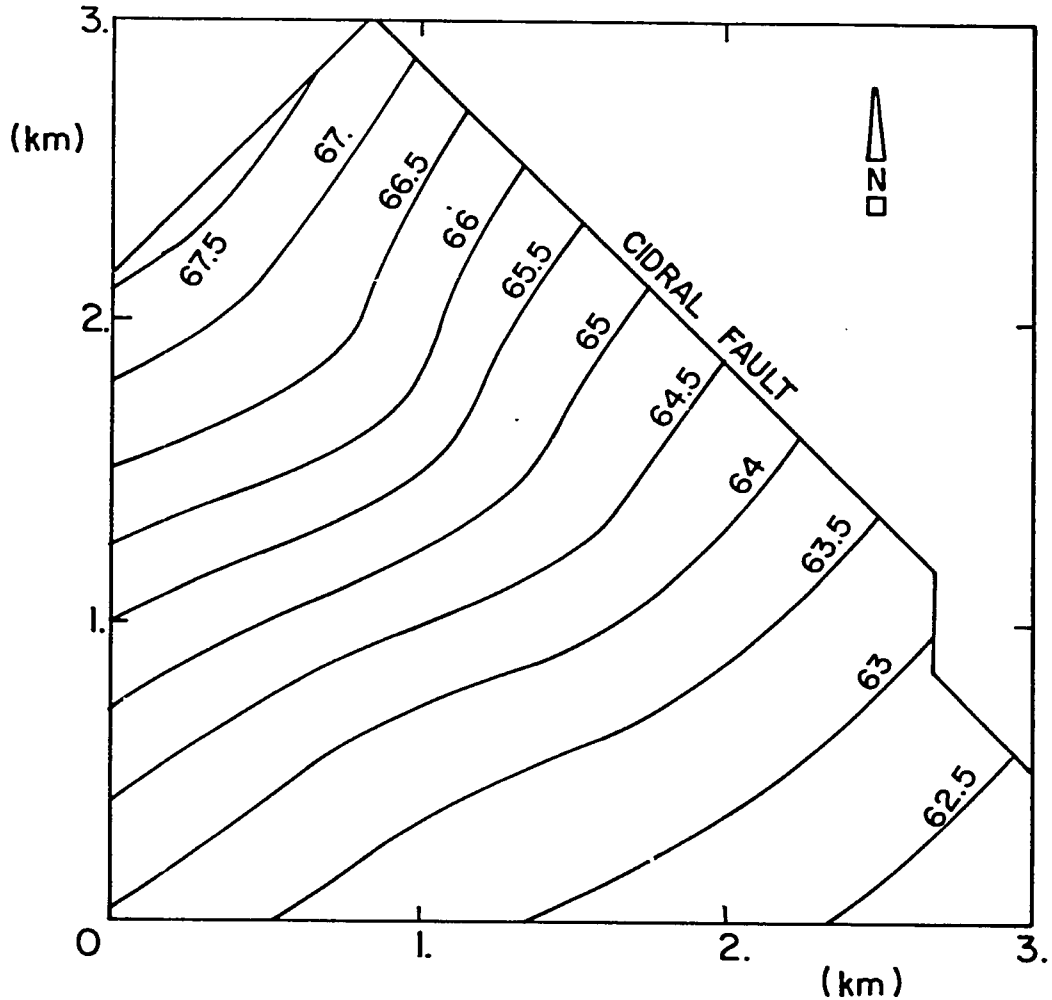

3.
(km)
2.
1.
0
1.
2.
3.
(km)
N
CIDRAL FAULT
67.5
67.
66.5
66
65.5
65
64.5
64
63.5
63
62.5

Fig. 12

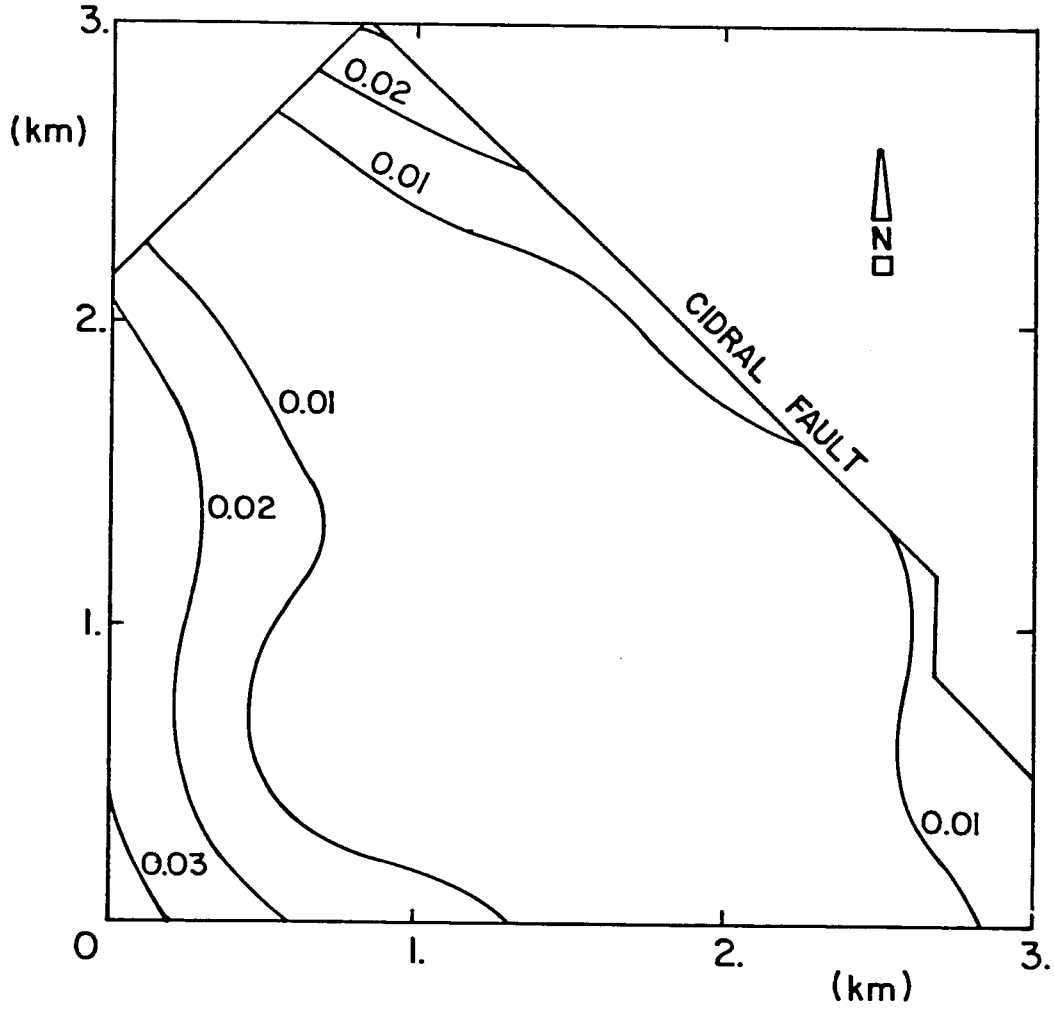


Fig. 13