Calibration and Reliability in Groundwater Modelling (Proceedings of the ModelCARE 99 Conference held at Zürich, Switzerland, September 1999). IAHS Publ. no. 265, 2000.

Simultaneous estimation of transmissivity (or conductivity), storage coefficient (or porosity) and effective recharge, in a stochastic framework

M. M. PACHECO FIGUEIREDO, L. RIBEIRO & J. M. P. FERREIRA LEMOS

Faculdade de Engenharia da Universidade do Porto (FEUP), Rua dos Bragas, 4099 Porto, Portugal, and

CVRM, Instituto Superior Técnico (IST), Av. Rovisco Pais, 1096 Lisboa, Portugal

e-mail: mpfiguei@fe.up.pt

Abstract Solving differential equations, which rule transient groundwater flow in confined or unconfined porous aquifers, implies the previous knowledge of hydrogeological parameters and effective recharge. In a stochastic framework the hydrogeological parameters are seen as random functions and their log-transforms can be represented by stationary isotropic multivariate normal random functions characterized by a small number of parameters. Recharge from rainfall can be assessed as a percentage of total precipitation. An algorithm, based on an inverse method approach, which simultaneously estimates all these parameters is described. This algorithm is complemented by a co-kriging technique which produces the first and second conditional moments of those hydrogeological parameters. A short description, as well as an application of this methodology to a synthetic case, is presented, including comparisons between "real" and expected conductivity, porosity and piezometric fields. The variance of estimation of the expected fields is also given.

INTRODUCTION

In a stochastic framework, hydrogeological parameters are seen as random functions and, under certain conditions, it can be assumed that a logarithm transform of these parameters behaves as a stationary isotropic multivariate normal random function. Transmissivity (or conductivity) is represented by the log-transform $Y = \ln T$ (or $Y = \ln K$), which is defined by the expected value, m_Y, nugget effect, w_Y, variance, σ^2_Y, integral scale, l_Y, and the covariance type. The same happens with the storage coefficient (or porosity), with its log-transform $G = \ln S$ (or $G = \ln \omega_d$) also defined by the same four parameters (m_G, w_G, σ^2_G, l_G) and covariance type. It is assumed that these random functions, Y and G, are not correlated. In those regional (unconfined) aquifers whose main recharge comes from rainfall, infiltration can be assessed as a percentage of total precipitation ($R = rP$). Then, if precipitation P is known, the effective recharge R can be quantified by parameter r.

THE STOCHASTIC DIRECT PROBLEM

The algorithm described here was developed for both types of porous regional aquifers (confined and unconfined). However, only the unconfined aquifer will be considered, for which the differential equation is:

$$\nabla\left|\left(e^{Y} D\nabla H\right)-e^{G}\frac{\partial H}{\partial t}-r\,P-Q-\sum_{i=1}^{N_w}Q_{w,i}\delta(\mathbf{x}-\mathbf{x}_{w,i})=0\right. \tag{1}$$

where H is the piezometric head, Q and $Q_{w,i}$ are, respectively, diffuse and point abstractions (or recharges) and the depth of the unconfined aquifer is represented by $D = H - Z_{\text{bottom}}$.

In a stochastic framework, a random function can be understood as the sum of an expected value and a zero mean perturbation. In that context, let $H = \langle H\rangle + h$, $Y = \langle Y\rangle + f$ and $G = \langle G\rangle + g$. Substituting these expressions in equation (1), replacing e^{f} by $(1 + f)$ and e^{g} by $(1 + g)$ and ignoring all second order terms (e.g. residual flux), a new equation is derived. Taking the expectation of this equation, a differential equation is obtained similar to (1) which relates expected values:

$$\nabla\left|\left(e^{\langle Y\rangle}\langle D\rangle\nabla\langle H\rangle\right)-e^{\langle G\rangle}\frac{\partial\langle H\rangle}{\partial t}-r\,P-Q-\sum_{i=1}^{N_w}Q_{w,i}\delta(\mathbf{x}-\mathbf{x}_{w,i})=0\right. \tag{2}$$

where $\langle D\rangle = \langle H\rangle - Z_{\text{bottom}}$. Subtracting this equation from the existing one, before expectation has been taken, a differential equation relating perturbations h, f and g is obtained:

$$\begin{aligned}&\nabla\left|\left(e^{\langle Y\rangle}\langle D\rangle\nabla h\right)+\nabla\right|\left(e^{\langle Y\rangle}h\nabla\langle H\rangle\right)-e^{\langle G\rangle}\frac{\partial h}{\partial t}+e^{\langle Y\rangle}\langle D\rangle\nabla f\left|\nabla\langle H\rangle+\right.\\&+f\,e^{\langle G\rangle}\frac{\partial\langle H\rangle}{\partial t}-g\,e^{\langle G\rangle}\frac{\partial\langle H\rangle}{\partial t}+f\,r\,P+f\,Q+f\sum_{i=1}^{N_w}Q_{w,i}\delta(\mathbf{x}-\mathbf{x}_{w,i})=0\end{aligned} \tag{3}$$

The integration of this differential perturbation equation, by the finite element method, gives place to a matricial expression which relates nodal head perturbations **h**, at different instants, with nodal perturbations **f** and **g**:

$$\mathbf{h}_k=\mathbf{A}_k\cdot\mathbf{f}+\mathbf{B}_k\cdot\mathbf{g}\,,\qquad k=1,\ldots,t_{\text{final}} \tag{4}$$

The matrices, $\mathbf{A}_k$ and $\mathbf{B}_k$, allow the evaluation of the covariance matrix of piezometric head $\mathbf{C}^{HH}$ and the cross covariance matrices $\mathbf{C}^{HY}$ and $\mathbf{C}^{HG}$, whose submatrices are defined by equations (5), (6) and (7):

$$\mathbf{C}^{H_kH_l}=E\left[\mathbf{h}_k\cdot\mathbf{h}_l^T\right]=\mathbf{A}_k\cdot\mathbf{C}^{YY}\cdot\mathbf{A}_l^T+\mathbf{B}_k\cdot\mathbf{C}^{GG}\cdot\mathbf{B}_l^T\quad,\quad k,l=1,\ldots,t_{\text{final}} \tag{5}$$

$$\mathbf{C}^{H_kY}=E\left[\mathbf{h}_k\cdot\mathbf{f}^T\right]=\mathbf{A}_k\cdot\mathbf{C}^{YY}\,,\qquad k=1,\ldots,t_{\text{final}} \tag{6}$$

$$\mathbf{C}^{H_kG}=E\left[\mathbf{h}_k\cdot\mathbf{g}^T\right]=\mathbf{B}_k\cdot\mathbf{C}^{GG}\,,\qquad k=1,\ldots,t_{\text{final}} \tag{7}$$

where $\mathbf{C}^{YY}$ and $\mathbf{C}^{GG}$ are the unconditional covariance matrices of random functions Y and G characterized by the statistical parameters w_Y, $\sigma^2{}_Y$, l_Y, w_G, $\sigma^2{}_G$ and l_G.

THE STOCHASTIC INVERSE PROBLEM

The problem remains as how to estimate model parameters $\boldsymbol{\beta} = (m_Y, m_G, r)$ and statistical parameters $\psi = (w_H, w_Y, w_G, \sigma^2{}_Y, \sigma^2{}_G, l_Y, l_G)$ (w_H is the variance of the head measurement errors, without any spatial correlation), taking into account field data, $K^*(\mathbf{x}_D)$ and $\omega_d^*(\mathbf{x}_D)$, and improving the estimation with piezometric head data $H^*(\mathbf{x}_D,t)$. To solve this problem, a stochastic inverse problem approach was developed, based on the works of Hoeksema & Kitanidis (1984) and Sun & Yeh (1992a). This inverse problem is stated through the well-known maximum likelihood estimator, which produces an objective function to be minimized:

$$S(\beta,\psi)=\ln\left|\mathbf{C}_D(\beta,\psi)\right|+\mathbf{e}_D^T\cdot\mathbf{C}_D^{-1}(\beta,\psi)\cdot\mathbf{e}_D \tag{8}$$

where $\mathbf{e}_D=\{H^*(\mathbf{x}_D,t)-H_M(\mathbf{x}_D,t);\ Y^*(\mathbf{x}_D)-m_Y;\ G^*(\mathbf{x}_D)-m_G\}$ is the vector of the residuals at the measurement points and $\mathbf{C}_D(\beta,\psi)$ is the generalized covariance matrix of the residuals at the same points. The formulation of this objective function, equation (8), and the evaluation of its derivatives with respect to the ten parameters $\theta=(\beta,\psi)$ implies solving, for the time period involved in the inverse problem, the mean flow equation (2), the sensitivity equations of H to the model parameters β and the perturbation equation (3). It also implies the calculation of the derivatives of matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ with respect to the model parameters, and the build up of $\mathbf{C}_D(\beta,\psi)$ and its derivatives with respect to all parameters θ. The optimization problem is solved with the help of a quasi-Newton method, the BFGS algorithm, complemented by a parabolic line-search. All differential equations are integrated by the finite element method (isoparametric elements of eight nodes).

ESTIMATION OF HYDROGEOLOGICAL PARAMETERS

Once the optimization problem is solved, the estimation of the spatial distribution of Y and G is performed by a co-kriging technique. For this purpose, submatrices of $\mathbf{C}_D(\beta,\psi)$ are used to build the system of equations which gives the co-kriging coefficients. Co-kriging of $\hat{f}(\mathbf{x}_0)=Y(\mathbf{x}_0)-m_Y$ and its covariance matrix is performed by:

$$\left[\begin{array}{c|c}\mathbf{C}^{H_DH_D} & \mathbf{C}^{H_DY_D}\\ \hline \mathbf{C}^{Y_DH_D} & \mathbf{C}^{Y_DY_D}\end{array}\right]\times\left[\begin{array}{c}\mu^Y_{\bar{\mathbf{x}}_0}\\ \hline \lambda^Y_{\bar{\mathbf{x}}_0}\end{array}\right]=\left[\begin{array}{c}\mathbf{C}^{H_DY_0}\\ \hline \mathbf{C}^{Y_DY_0}\end{array}\right] \tag{9}$$

$$\hat{f}(\mathbf{x}_0)=\sum_{i=1}^{n_Y}\lambda^Y_{i,\bar{\mathbf{x}}_0}f(\mathbf{x}_{i,Y})+\sum_{j=1}^{n_H}\mu^Y_{j,\bar{\mathbf{x}}_0}h(\mathbf{x}_{j,H}) \tag{10}$$

$$C^{YY}_{CoK}(\mathbf{x}_a,\mathbf{x}_b)=C^{YY}(\mathbf{x}_a,\mathbf{x}_b)-\sum_{i=1}^{n_Y}\lambda^Y_{i,\bar{\mathbf{x}}_a}C^{YY}(\mathbf{x}_{i,Y},\mathbf{x}_b)-\sum_{j=1}^{n_H}\mu^Y_{j,\bar{\mathbf{x}}_a}C^{HY}(\mathbf{x}_{j,H},\mathbf{x}_b) \tag{11}$$

Co-kriging of $\hat{g}(\mathbf{x}_0)=G(\mathbf{x}_0)-m_G$ and its covariance is realized by identical expressions, which are omitted for reasons of brevity.

SYNTHETIC CASE

A synthetic case was set up (Fig. 1) in order to demonstrate an application of this methodology. The "real" log conductivity field Y, was produced, with the help of a simulated annealing algorithm, in order to present, approximately, the following statistical parameters ($m_Y = \ln(20)$, $w_Y = 0.06$, $\sigma^2_Y = 0.24$, $l_Y = 1.0$ km). The same procedure was used to get the "real" log porosity field, G ($m_G = \ln(0.25)$, $w_G = 0.02$, $\sigma^2_G = 0.08$, $l_G = 1.0$ km). The corresponding K and ω_d fields are presented in Fig. 2. The direct simulation with the "real" hydrogeological parameters was performed with the effective recharge due to rainfall infiltration defined by $r = 0.40$. Hydrogeological and piezometric data were "collected" at, intentionally few, random located points (Fig. 1).

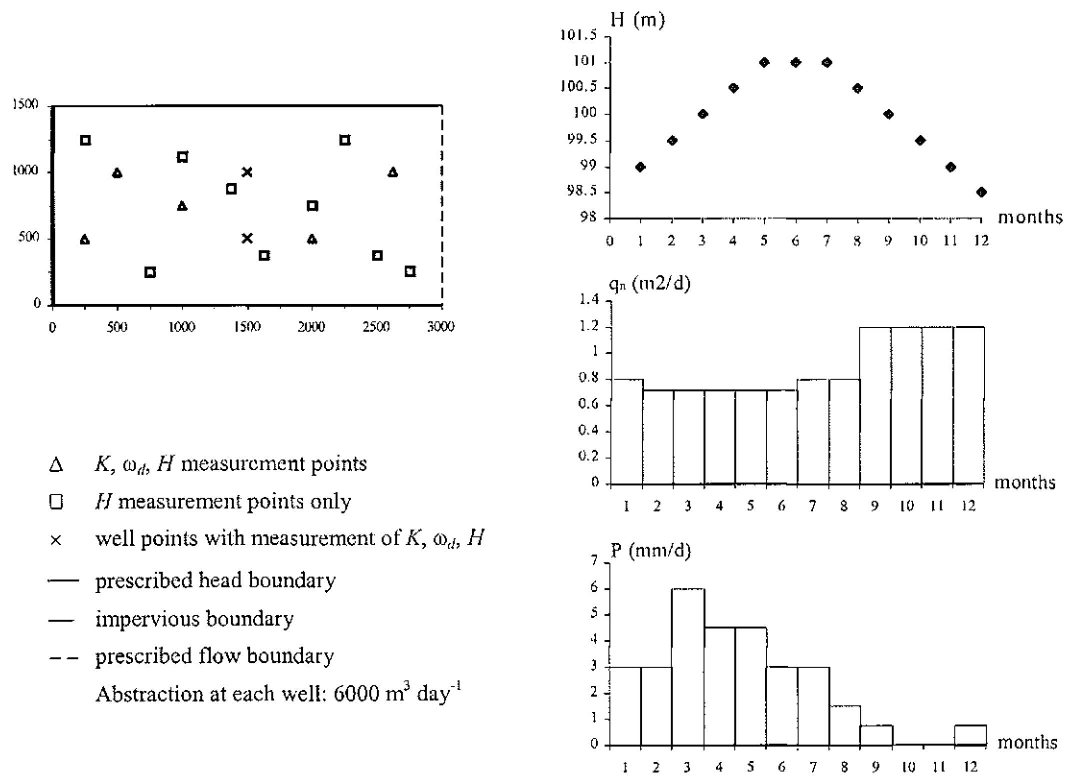

Fig. 1 Aquifer definition: well points; data measurement points; boundary conditions: prescribed head (H) and flow (q_n); precipitation (P).

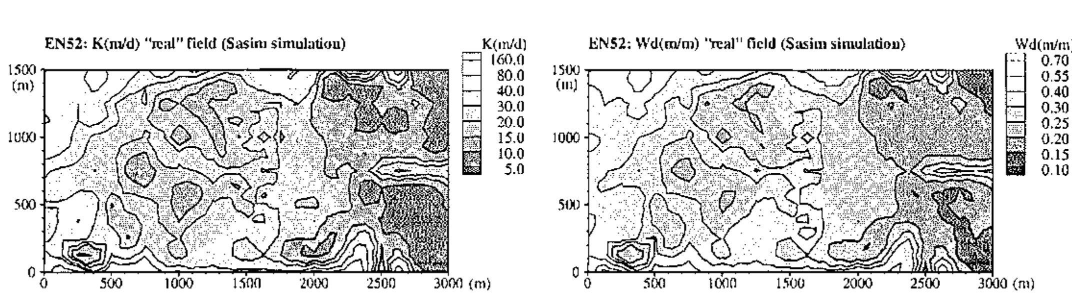

Fig. 2 "Real" fields of K (m day^{-1}) and ω_d.

The inverse modelling algorithm gave, for the ten parameters described above, the values presented in Table 1. The estimated fields, produced by the inverse modelling and co-kriging methodology, are shown in Fig. 3 and their variance of estimation in Fig. 4. In order to obtain the expected piezometric fields, the simulation model (equation (2)) was run for the hydrogeological estimated fields. A visual comparison between the "real" and expected piezometric fields, at month 6, can be drawn from Fig. 5. The variance of the expected piezometric field is the main diagonal of the matrices given by equation (5) solved for the conditional (co-kriged) covariance of Y

Table 1 Model and statistical parameters given by the inverse modelling algorithm.

m_Y	m_G	r	w_H (mm^2)	w_Y	w_G	σ^2_Y	σ^2_G	l_Y (km)	l_G (km)
3.11332	−1.26781	0.39998	1.37672	0.06545	0.39920	0.54538	0.11806	0.70033	4.84504

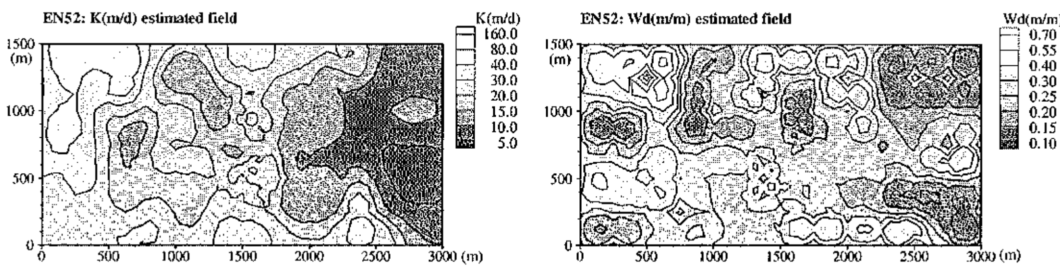

Fig. 3 Co-kriged fields of K (m day^{-1}) and ω_d.

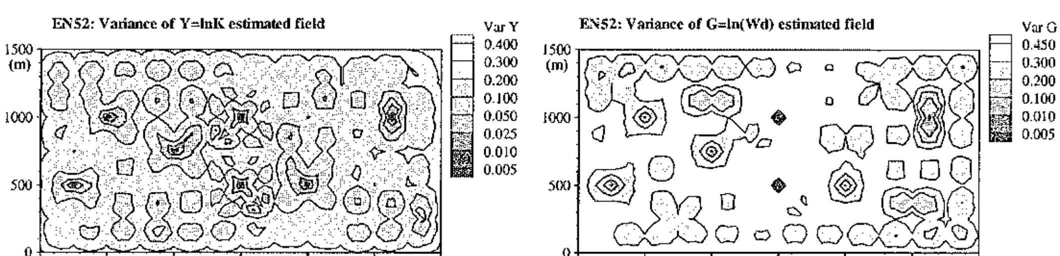

Fig. 4 Variance of co-kriging estimation of log-transformed fields $Y = \ln K$ and $G = \ln\omega_d$.

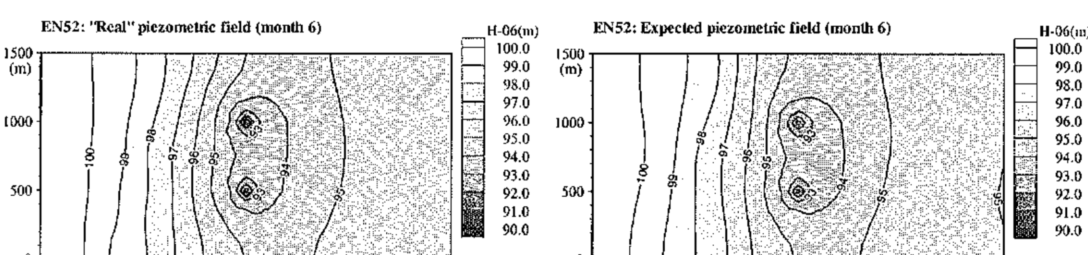

Fig. 5 "Real" and estimated piezometric fields at month 6.

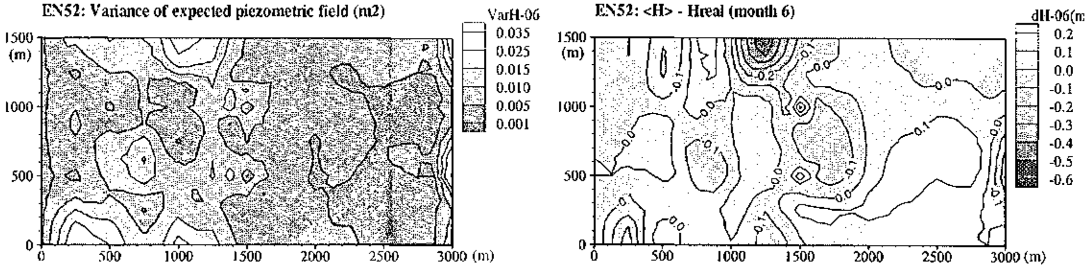

Fig. 6 Variance of expected piezometric field and difference between expected and "real" piezometric fields, at month 6.

and G, with matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ obtained from perturbation equation (3) integrated for the estimated hydrogeological fields and effective recharge. Finally, this conditional variance and the difference between the expected and "real" piezometric fields, at the same month, is presented in Fig. 6.

CONCLUSIONS

Parameter r reaches, almost exactly, the value chosen for the "real" direct simulation. A good degree of accuracy has been registered in every synthetic case tested. The two other model parameters do not behave so well. The same applies to the other statistical parameters. The most inaccurate approximation typically occurs with the log-transform of porosity, although some tests conducted for synthetic confined aquifers revealed a significantly better approximation of the log-transform of storage coefficient, S. The estimated spatial distributions of K and ω_d reflect these facts. Furthermore, it has been verified that the statistical parameters which characterize the random functions are very sensitive to small changes in measurement point positioning, even when it only involves changes of one or two "piezometers". However, the expected piezometric field presents a remarkably better approximation to the "real" field. And the real advantage of estimating K and ω_d, or T and S, with an inverse modelling and co-kriging procedure is well understood when a comparison is made with piezometric fields based on hydrogeological parameters estimated exclusively by geostatistical means. The differences are particularly evident, for example, in the vicinity of well points, where the flow is strongly determined by the hydraulic conductivity.

REFERENCES

Carrera, J. & Neuman, S. P. (1986a) Estimation of aquifer parameters under transient and steady state conditions. 1. Maximum likelihood method incorporating prior information. *Wat. Resour. Res.* **22**(2), 199–210.

Carrera, J. & Neuman, S. P. (1986b) Estimation of aquifer parameters under transient and steady state conditions. 2. Uniqueness, stability and solution algorithms. *Wat. Resour. Res.* **22**(2), 211–227.

Dagan, G. (1985) Stochastic modeling of groundwater flow by unconditional and conditional probabilities: the inverse problem. *Wat. Resour. Res.* **21**(1), 65–72.

Dagan, G. (1989) *Flow and Transport in Porous Formations*. Springer-Verlag, Berlin, Germany.

Dagan, G. & Rubin, Y. (1988) Stochastic identification of recharge, transmissivity and storativity in aquifer transient flow: a quasi-steady approach. *Wat. Resour. Res.* **24**(10), 1698–1710.

Deutsch, C. V. & Journel, A. G. (1992) *GSLIB. Geostatistical Software Library and User's Guide*. Oxford University Press, New York.

Fletcher, R. (1987) *Practical Methods of Optimization*. John Wiley & Sons, New York.

Hoeksema, R. J. & Kitanidis, P. K. (1984) An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modelling. *Wat. Resour. Res.* **20**(7), 1003–1020.

Rubin, Y. & Dagan, G. (1987a) Stochastic identification of transmissivity and effective recharge in steady groundwater flow. 1. Theory. *Wat. Resour. Res.* **23**(7), 1185–1192.

Rubin, Y. & Dagan, G. (1987b) Stochastic identification of transmissivity and effective recharge in steady groundwater flow. 2. Case study. *Wat. Resour. Res.* **23**(7), 1193–1200.

Sun, N.-Z. & Yeh, W. W.-G. (1992a) A stochastic inverse solution for transient groundwater flow: parameter identification and reliability analysis. *Wat. Resour. Res.* **28**(12), 3269–3280.

Sun, N.-Z. (1994) *Inverse Problems in Groundwater Modeling*. Kluwer Academic Publishers, Dordrecht, The Netherlands.