



HAL
open science

Determination of Geostatistical Parameters Using Well Test Data

Y. Gautier, B. Noetinger

► **To cite this version:**

Y. Gautier, B. Noetinger. Determination of Geostatistical Parameters Using Well Test Data. Oil & Gas Science and Technology - Revue d'IFP Energies nouvelles, 2004, 59 (2), pp.167-183. 10.2516/ogst:2004013 . hal-02017295

HAL Id: hal-02017295

<https://ifp.hal.science/hal-02017295>

Submitted on 13 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Geostatistical Parameters Estimation Using Well Test Data

Y. Gautier^{1*} and B. Noetinger¹

¹ Institut français du pétrole, 1 et 4, avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex – France

e-mail: yann.gautier@ifp.fr – benoit.noetinger@ifp.fr

* Corresponding author

Résumé — Détermination de paramètres géostatistiques par l'utilisation des données d'essais de puits — Des études de réservoirs de plus en plus nombreuses font appel à une description probabiliste du réservoir. Ceci permet de quantifier les incertitudes sur les prévisions de production, à condition de disposer d'un modèle géostatistique convenablement paramétré. Les paramètres les plus courants sont la (ou les) longueur(s) de corrélation et la variance de la perméabilité. Ces paramètres sont déterminés à l'aide de considérations géologiques, auxquelles on ajoute éventuellement des mesures pétrophysiques sur carottes à l'échelle du laboratoire. Il serait toutefois intéressant de disposer d'une technique de mesure directe de ces paramètres, en utilisant des données investigant des échelles plus importantes du réservoir considéré. Dans cet article, nous examinons l'apport des essais de puits, qui ont l'avantage d'être disponibles au début de l'exploitation du gisement, et dont on peut montrer qu'ils "moyennent" la perméabilité du réservoir sur une zone correspondant à leur rayon d'investigation. La méthode proposée consiste à chercher les paramètres géostatistiques rendant le test de puits effectivement obtenu sur le terrain le plus probable possible. La technique utilisée repose sur des techniques de changement d'échelle, combinées à des approches de calage bayésien et des méthodes d'optimisation. Les résultats obtenus sur cas synthétiques montrent une bonne détermination des ordres de grandeur des paramètres estimés.

Abstract — Determination of Geostatistical Parameters Using Well Test Data — In this paper we describe a new method to obtain estimations of the geostatistical parameters (GPs) such as the correlation length, l_c and the permeability variance, σ_{ln}^2 from well test data. In practical studies, the GPs are estimated using geological and petrophysical data, but often, these data are too scarce to give precise results. The proposed method uses the Bayesian inversion theory, in conjunction with a fast evaluation of well tests that implies upscaling techniques. The method was tested using synthetic well-test data performed on some training images, and estimations of the underlying correlation length, l_c and permeability variance, σ_{ln}^2 were recovered. These estimations give a correct order of magnitude of the actual values, but as noticed in similar methods, the uncertainties are high. Once the GPs are estimated, other well established techniques can be used to get well-test matched reservoir images consistent with the geostatistical model. We will see that excellent well test data are needed, and that the method could be improved using multiple well test data.

NOMENCLATURE

Mathematical notations

$\delta(\mathbf{r})$	Dirac delta-function
$Y(t)$	Heaviside function $Y(t) = 0$ if $t < 0$ and $Y(t) = 1$ if $t \geq 0$
\mathbf{I}_N	$N \times N$ Identity matrix
\mathbf{A}^t	Transpose of the matrix \mathbf{A}
$ \mathbf{A} $	Determinant of the matrix \mathbf{A}
$\text{Tr } \mathbf{A}$	Trace of the matrix \mathbf{A}
$[x] = \bar{x}$	Mean of the random variable x
$p(H_1)$	Probability density function of H_1
$p(H_1 H_2)$	Probability density function of H_1 if H_2 is known

Other notations

c_t	Total compressibility of rock + fluids (bar^{-1})
\mathbf{C}	Covariance matrix of the parameters
\mathbf{d}	Observation vector ($d_i = 1/k_{app}(t_i)$) (mD^{-1})
$G(r, t)$	Weighting kernel for ring permeabilities
\mathbf{G}^{2D} and \mathbf{G}	Represent the 2D well test simulator and his fast approximation
h	Depth of the reservoir (m)
\mathbf{J}	Jacobian matrix of \mathbf{G} (mD^{-1})
$(k_j)_{j=1 \dots N_r}$	Ring permeabilities (mD)
k_{eff}	Effective permeability (mD)
k_g	Geometric mean of the permeabilities (mD)
k_w	Wellbore permeability (mD)
$\hat{k}(t)$	Estimation of the apparent permeability at time t (mD)
$k_{app}(t)$	Apparent permeability at time t (mD)
$L(\Theta)$	Negative of the log-likelihood function
l_c	Correlation length (m)
M	Fisher information matrix (approximation of the Hessian of L)
N_t	Number of observations (ie. number of time steps)
N_r	Number of parameters (ie. number of rings)
N_x, N_y	Number of gridblocks in x and y directions
$P(\mathbf{r}, t)$	Pressure at the point \mathbf{r} and at time t (bar)
Q	Well rate (m^3/s)
$(r_j)_{j=1 \dots N_r}$	Ring radius (m)
\mathbf{R}	Covariance matrix of the observations (mD^{-2})
$R_{app}(t)$	Radius of investigation at time t (m)
R_{min}, R_{max}	Radius control parameter to choose the number of observations used in the optimization procedure (m)
\mathbf{y}	Parameter vector ($y_j = \log(k_j)$)
Δ_x, Δ_y	Gridblock dimensions in x and y dimensions (m)
λ	Parameter computed to improve the convergence rate of the optimization algorithm
ϕ	Rock porosity

μ	Fluid viscosity (cp)
$\rho(h)$	Correlation function related with the variogram γ_{ln} in the considered cases by $\gamma_{ln}(h) = \sigma_{ln}^2 - \rho(h)$
σ_{ln}^2	Variance of the log-permeability
σ_c^2	Control parameter which represent our knowledge of k_w
σ_w^2	Control parameter for the observation matrix ($\mathbf{R} = \sigma_w^2 \mathbf{I}_{N_t}$)
Σ	Covariance matrix for the observation pdf $p(\mathbf{d} \Theta)$
Θ	Geostatistical parameter vector ($\Theta = [\sigma_{ln}^2 \ l_c]^t$)
Ξ	Hessian of $L(\Theta)$

INTRODUCTION

The use of geostatistical models to improve production forecasts is becoming more and more popular among reservoir engineers. In order to reduce the forecasting uncertainties, an important challenge is to history match production data, using the information provided by geostatistics in a consistent manner. This problem was investigated by many authors, using various optimization techniques and the Bayesian paradigm [1–8]. In general, the underlying geostatistical model, as well as the geostatistical parameters (GPs) such as the correlation length, l_c and the log-permeability variance, σ_{ln}^2 are supposed to be fixed.

Most often, in practical studies, the GPs are estimated using geological information and statistical analysis of the available petrophysical data [9,10]. In a lot of cases, these data are too scarce to give precise results, and more particularly if the typical interwell spacing is very large compared to the correlation length. This problem received some attention in the oil-oriented technical literature. Yadavalli et al. [11, 12] using Sagar [6] and Feitosa's [13] approach were able to get estimations of the GPs from a single well test. Nevertheless, in their approach, they derived the variogram of a two-dimensional permeability field from the computation of the variogram of a radial one-dimensional equivalent permeability field that leads to the correct well test. Their main assumption is that the two variograms are equivalent. We will show that this is not the case in general and that the relation between the two variograms is rather complex (Eq. (25)). It is therefore not obvious to give an interpretation of their results. Yortsos [14] proposed an alternative technique that works when a lot of independent well test data are available.

The determination of the GPs from head measurement data received considerable attention in the hydrogeology literature (see [15–17] and references therein). The authors, in these studies, used a Bayesian framework that formalized the history matching process using least square optimization techniques. In the present paper, we will adapt this

general framework to the particular case of well test data. The main idea is to evaluate the joint probability density of having any given reservoir image and any given set of dynamic data measurements, for fixed GPs. The GPs estimation is done using the maximum likelihood principle which states that among the GPs, the probability density to get the actual dynamic data must be maximum. The method was tested using synthetic well-test data on some training images, and estimations of the underlying correlation length, l_c and permeability variance, σ_{ln}^2 were found. As noticed in the literature dealing with the geostatistical approach of inverse problems [18, 17, 19, 20], we will find that the estimates give the good order of magnitude of the actual values but the uncertainties are high. Once the GPs are estimated, one can thus obtain well-test matched reservoir images consistent with the geostatistical model using standard inversion methods.

1 WELL TESTS AND APPARENT PERMEABILITY

Well tests are commonly used in reservoir engineering for better understanding the structure of the reservoir. Typically the investigated scale is hectometric, and important properties such as the average permeability, the reservoir boundary positions, skin and wellbore storage effects can be estimated from a well test interpretation.

Thanks to the improvement of pressure gauges, modern well test interpretation techniques make an intensive use of the pressure derivative curve. This tool, as well as the associated concept of instantaneous apparent permeability are very powerful techniques to interpret well tests in a geostatistical context. In a first part, we present our main assumptions and introduce the apparent permeability as well as its main properties.

1.1 The Diffusivity Equation

Hereinafter, we will work with the following assumptions:

- the reservoir is 2D and of infinite dimension
- the flow is single-phase
- the porosity, ϕ and the viscosity μ are constant
- the fluids and the rock are considered slightly compressible and the total compressibility c_t is constant
- skin and wellbore storage effects are negligible

Within this framework, considering that a well located at the origin is pumping at a constant rate Q from $t = 0$, the pressure variation $P(\mathbf{r}, t)$ and the permeability map, $k(r)$ are related by the diffusivity Equation [21–23]:

$$\phi \mu c_t \frac{\partial}{\partial t} P(\mathbf{r}, t) = \nabla \bullet (k(\mathbf{r}) \nabla P(\mathbf{r}, t)) + \frac{\mu Q}{h} \delta(r) Y(t) \quad (1)$$

h denotes the reservoir thickness and the second term of the right hand side is a source term. It is the product of a spatial Dirac function $\delta(r)$ by a Heaviside function $Y(t)$ (which is

equal to 0 if $t < 0$ and 1 if $t \geq 0$). This approximation (known as the line source approximation) holds if the well radius is infinitesimally small with respect to the typical scale of the reservoir.

To get a well defined problem, we must fix some boundary conditions. Initially, the pressure variation is uniformly equal to 0 (*i.e.* $\forall \mathbf{r}, P(\mathbf{r}, t = 0) = 0$) and the pressure variation vanishes far from the well (*i.e.* $\forall t, P(\mathbf{r} \rightarrow \infty, t = 0) = 0$).

1.2 The Apparent Permeability

This diffusivity Equation can be solved analytically for a constant permeability, k_0 [22, 23]. Typically, the product $k_0 h$ can be derived from the bottomhole pressure $P_0(t) = P_0(r \rightarrow 0, t)$:

$$k_0 h = \frac{\mu Q}{4\pi \frac{dP_0(t)}{d \log t}} \quad (2)$$

This relation is the basis of the derivative technique in well test interpretation. Additional physical phenomena such as skin, wellbore storage, and practical constraints transform well test interpretation into a full speciality. An important technical literature exists on the subject, with intensive use of type-curve analysis and numerical optimization methods [21, 22, 24].

Equation (2) can be generalized in the heterogeneous case defining the apparent permeability concept:

$$k_{app}(t) = \frac{\mu Q}{4\pi h \frac{dP(t)}{d \log t}} \quad (3)$$

Of course, for a homogeneous reservoir, $k_{app}(t)$ is independent of time and is equal to the constant permeability, k_0 . We need also an estimation of the radius of investigation which corresponds to the range investigated by the well test:

$$R_{app}(t) = C \sqrt{\frac{k_{app}(t) t}{\phi \mu c_t}} \quad (4)$$

Here, the dimensionless constant C , the order of magnitude of which is unity, depends on which definition is used [24]. For the sake of simplicity, we will set $C = 1$ throughout the paper.

In a geostatistical context, the relation between $k_{app}(t)$ and the permeability field $k(\mathbf{r})$ is very complex and a gridded reservoir simulator is needed. Hereinafter we will consider that $k(\mathbf{r})$ is a random variable⁽¹⁾, and that the actual permeability field is only a particular realization of $k(\mathbf{r})$.

In this statistical framework, some important asymptotic results can be derived using perturbation methods [26] that can be extended in a systematic manner using a Feynman

(1) Gelhar [25], in the introduction of his book (§ 1.3), mentions some key points for working with random variables and stochastic processes. This approach can formally handle the uncertainties due to the lack of measurements of porous media characteristic parameters.

diagram approach [27]. For short times, the apparent permeability tends to the wellbore area permeability, k_w :

$$\lim_{t \rightarrow 0} k_{app}(t) = k_w \tag{5}$$

and for long times, the apparent permeability tends to the effective permeability k_{eff} corresponding to the upscaled reservoir permeability :

$$\lim_{t \rightarrow \infty} k_{app}(t) = k_{eff} \tag{6}$$

It can be shown that the typical time to reach this last limit is such that $R_{app}(t) = l_c$, when the radius of investigation is in the order of magnitude of the permeability correlation length. This means that the well test is performing a homogenization of the reservoir permeability. This is the physical basis of the proposed method where the geostatistical parameters will be estimated from well tests.

2 FAST EVALUATION OF WELL TESTS

Since we will use an optimization method to evaluate the geostatistical parameters and, therefore, an important amount of simulations will be required, a fast approximation of well tests is needed. The approximation we will use was first proposed by Oliver [4] for low permeability contrasts where a first order expansion of the diffusivity Equation (1) was considered. This approach was further extended by Sagar [6] and Feitosa [13] for high permeability contrasts. They found that the apparent permeability can be considered as an averaging process over volumes of increasing size with respect to time. More precisely, the approximation of the apparent permeability $\hat{k}(t)$ is a weighted sum of a “homogenized permeability field” that contains N_r ring permeabilities $k_j, j = 1 \dots N_r$ (Fig. 1):

$$\frac{1}{\hat{k}(t_i)} = \sum_{j=0}^{N_r} G_{ij} \cdot \frac{1}{k_j} \tag{7}$$

The weights G_{ij} depend on a kernel function $G(r, t)$ (2):

$$G_{ij} = \frac{\int_{\rho_{i,j}}^{\rho_{i,j+1}} z dz G\left(\frac{z^2}{2}\right)}{\int_0^\infty z dz G\left(\frac{z^2}{2}\right)} \tag{8}$$

with

$$\rho_{i,j} = \frac{r_j}{R_{app}(t_i)} \tag{9}$$

(2) $G(r, t) = \frac{1}{2} z e^{-z} \left(2K_1(z) \left(1 - \frac{1}{z} \right) + K_1(z) + K_2(z) \right)$ where K_0, K_1, K_2 represent the zero, one and two order modified Bessel functions and $z = \frac{1}{2} \left(\frac{r}{R_{app}(t)} \right)^2$

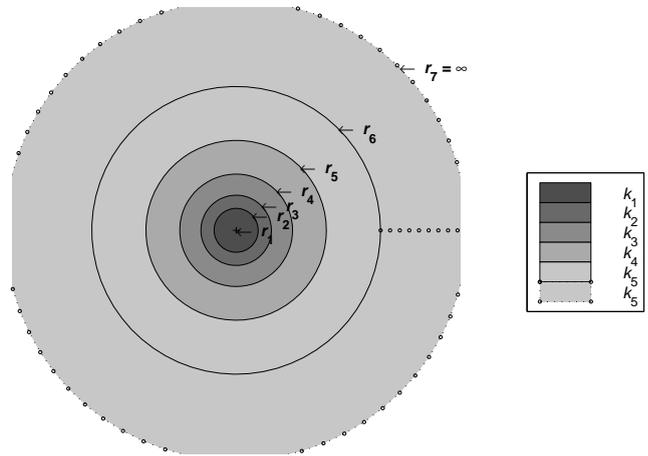


Figure 1

Definition of ring permeabilities and ring radii for a homogenized field containing 6 rings.

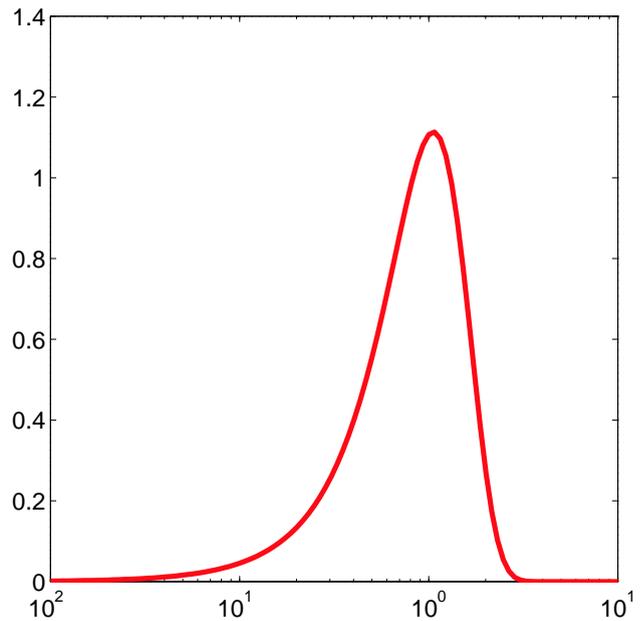


Figure 2

Permeability weighting kernel $\sqrt{z}G(z)$ vs. $z = r^2/t$.

We have plotted in Figure 2 the function $\sqrt{z}G(z)$ vs. $z = r^2/t$ which appears in Equation (9). The maximum of this function is for $r = 0.92R_{app}(t)$ and 99% of the whole contribution comes from the area contained between the two radii $r_{min} = 0.12R_{app}(t)$ and $r_{max} = 2.34R_{app}(t)$. Furthermore, we have a normalization condition:

$$\int_0^\infty dz G(z) = 1 \tag{10}$$

and, assuming that $r_1 = 0$ and $r_{N_r} = \infty$:

$$\sum_{j=1}^{N_r} G_{ij} = 1 \quad (11)$$

The next step is to relate the ring permeabilities k_j to the original 2D permeability map. Feitosa [13] used different averaging formulae (arithmetic, harmonic and geometric) to compute k_j from a 2D lognormal permeability field. He compared the resulting estimations with some reference results obtained using a complete finite difference well test simulation. He found that the best match of the apparent permeability is obtained when the geometric mean is used. This result is consistent with the general properties of the apparent permeability given by Equation (6). For short times, independently of the averaging rule, $\hat{k}(t)$ will tend to the wellbore permeability. For long times, using (6), $\hat{k}(t)$ has to converge to the effective permeability which is precisely the permeability geometric mean in the 2D lognormal case [28]. Since, in the following, we will only consider lognormal permeability fields, the ring permeabilities are the geometric mean of the fine scale permeabilities contained in the ring:

$$Y_j = \log(k_j) = \sum_{l \in \Omega_j} \omega_l \log(K_l) \quad (12)$$

where K_l denotes the Cartesian permeabilities, Ω_j are the indices of the gridblocks overlapping the ring j and ω_l is a factor that represents the fraction of the gridblock's area contained in the ring.

3 GEOSTATISTICAL PARAMETER ESTIMATION USING WELL TEST DATA

Kitanidis and Vomvoris [19] estimate geostatistical parameters using steady-state pressure and permeability measurements for a 1D flow problem. Our present goal is to generalize the method for transient flow. Now, the data considered are measurements of apparent permeability values. Nevertheless, the same formalism can be applied. The interested reader can refer to different papers [19, 18, 17] and to the references therein.

Let us now present the geostatistical approach and some notations. We will consider that we have N_t apparent permeability measurements and a permeability map described by the value of the permeability at the centers of N_m gridblocks.

We define an observation space and a parameter space where:

- \mathbf{Y} is a N_m parameter vector containing the parameters of the problems, which are, in our case, the permeability logarithm map.

$$Y_j = \log(K_j), \quad j = 1 \cdots N_m \quad (13)$$

The parameter mean $\langle \mathbf{Y} \rangle = \bar{\mathbf{Y}}$ is considered as known. The $(N_m \times N_m)$ covariance matrix

$$\mathbf{C}^{2D}(\Theta) = \langle (\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})^t \rangle \quad (14)$$

is a known function of a few geostatistical parameters contained in a vector Θ (variance, correlation length, nugget effect, etc). The variogram shape is assumed to be known.

- \mathbf{d} is a N_t vector containing all the observations which, in this case, are the inverse of the apparent permeability values for a set of time steps.

$$d_i = \frac{1}{k_{app}(t_i)}, \quad i = 1 \cdots N_t \quad (15)$$

Calling \mathbf{G}^{2D} the transfer function (the simulator) that maps the parameter space into the observation space, we obtain the following relation:

$$\mathbf{d} = \mathbf{G}^{2D}(\mathbf{Y}) + \mathbf{v} \quad (16)$$

\mathbf{v} is a random error vector. We suppose that $\langle \mathbf{v} \rangle = \mathbf{0}$ (the estimation is non-biased) and that the error covariance matrix $\mathbf{R}^{2D} = \langle \mathbf{v}\mathbf{v}^t \rangle$ does not depend on Θ .

We denote $p(\mathbf{d}, \mathbf{y}|\Theta)$, the probability density function PDF of the model parameters and of the observations, $p(\mathbf{d}|\mathbf{y}, \Theta)$, the PDF of the observations assuming that the model parameters are known and, $p(\mathbf{y}|\Theta)$, the PDF of the model parameters assuming that the geostatistical parameters are known. Using Bayes theorem, we obtain

$$p(\mathbf{d}, \mathbf{Y}|\Theta) = p(\mathbf{d}|\mathbf{Y}, \Theta) p(\mathbf{Y}|\Theta) \quad (17)$$

We define the marginal PDF of the observations as the sum over all possible configurations of the model parameters of $p(\mathbf{d}, \mathbf{y}|\Theta)$:

$$p(\mathbf{d}|\Theta) = \int d^{N_m} \mathbf{Y} p(\mathbf{d}, \mathbf{Y}|\Theta) \quad (18)$$

The sum is taken over N_m dimensions. Kitanidis [17] shows that geostatistical parameter estimation can be derived from two different methods:

- the Maximum A Posteriori (MAP) method: Θ estimation is obtained by maximizing $p(\mathbf{d}, \mathbf{Y}|\Theta)$ with respect to both \mathbf{Y} and Θ .
- the Geostatistical Approach (GA) method: the estimation of Θ is obtained by maximizing $p(\mathbf{d}|\Theta)$ with respect to Θ . This method was widely used in the literature [19, 18, 20, 29, 30].

Therefore two different estimations are obtained: $\tilde{\Theta}_{MAP}$ and $\tilde{\Theta}_{GA}$. Kitanidis [17], by studying asymptotic results, showed that $\tilde{\Theta}_{MAP}$ is a biased estimation whereas $\tilde{\Theta}_{GA}$ is not. Moreover, he demonstrates, using an illustrative example, that the bias is highly dependent on the model discretization.

Therefore, we shall focus on the GA method. We assume that the two PDFs $p(\mathbf{d}|\mathbf{Y}, \Theta)$ and $p(\mathbf{Y}|\Theta)$ are both Gaussian:

$$p(\mathbf{d}|\mathbf{Y}, \Theta) = \frac{1}{(2\pi)^{n/2} |\mathbf{R}^{2D}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{d} - \mathbf{G}^{2D}(\mathbf{Y}))^t \times (\mathbf{R}^{2D})^{-1} (\mathbf{d} - \mathbf{G}^{2D}(\mathbf{Y}))\right) \quad (19)$$

$$p(\mathbf{Y}|\Theta) = \frac{1}{(2\pi)^{m/2} |\mathbf{C}^{2D}|^{1/2}} \times \exp\left(-\frac{1}{2} (\mathbf{Y} - \bar{\mathbf{Y}})^t (\mathbf{C}^{2D})^{-1} (\mathbf{Y} - \bar{\mathbf{Y}})\right) \quad (20)$$

The second assumption is true since we consider lognormal permeability fields. In order to make notations easier, we assume implicitly the dependence of \mathbf{C} on Θ . Equation (18) becomes:

$$p(\mathbf{d}|\Theta) \propto \frac{1}{|\mathbf{R}^{2D}|^{1/2} |\mathbf{C}^{2D}|^{1/2}} \int d\mathbf{Y}^{N_m} e^{I(\mathbf{Y})} \quad (21)$$

with

$$I(\mathbf{Y}) = -\frac{1}{2} \left[(\mathbf{Y} - \bar{\mathbf{Y}})^t (\mathbf{C}^{2D})^{-1} (\mathbf{Y} - \bar{\mathbf{Y}}) + (\mathbf{d} - \mathbf{G}^{2D}(\mathbf{Y}))^t (\mathbf{R}^{2D})^{-1} (\mathbf{d} - \mathbf{G}^{2D}(\mathbf{Y})) \right] \quad (22)$$

The GA approach involves the maximization of (21) to estimate the GPs. Here we have to point out that the sum is taken over N_m dimensions which implies typically 10^6 gridblocks. In practice, it is impossible to evaluate such an integral. Furthermore, the GA method involves an optimization step, and consequently many flow simulations which are very CPU time consuming. Therefore, we will have to simplify the problem: we will use the fast evaluation of the apparent permeability described in the previous section. Considering that the estimation is close to the simulation results, Gautier [31] shows that we have:

$$p(\mathbf{d}|\Theta) = \int d\mathbf{Y}^{N_m} p(\mathbf{d}, \mathbf{Y}|\Theta) = \int d\mathbf{y}^{N_r} p(\mathbf{d}, \mathbf{y}|\Theta) \quad (23)$$

\mathbf{y} is the vector containing N_r radial permeabilities defined by (12). Its mean and covariance matrix are derived from the 2D permeability field:

$$\bar{y}_j = \sum_{l \in \Omega_j} \omega_l \langle Y_l \rangle \quad (24)$$

$$\begin{aligned} C_{ij} &= ((y_i - \bar{y}_i) (y_j - \bar{y}_j)) \\ &= \sum_{k \in \Omega_i} \sum_{l \in \Omega_j} \omega_l \omega_k \langle (Y_k - \bar{Y}_k) (Y_l - \bar{Y}_l) \rangle \\ &= \sum_{k \in \Omega_i} \sum_{l \in \Omega_j} \omega_l \omega_k \rho(d_{kl}) \end{aligned} \quad (25)$$

ρ represents the correlation function. The computation of the matrix \mathbf{C} is expensive since it involves a large number of products. Nevertheless, if matrices are stored, it has to be done only once for a configuration of ring and for a particular variogram shape. From (7), the observations and the parameters are now related by:

$$\mathbf{d} = \mathbf{G}(\mathbf{y}) = \mathbf{G} e^{-\mathbf{y}} \quad (26)$$

\mathbf{G} is the matrix containing the weights of the kernel function G_{ij} defined by (8). The simplified problem involves the maximization of $p(\mathbf{d}|\Theta)$ that can be written in the new parameter space:

$$p(\mathbf{d}|\Theta) \propto \frac{1}{|\mathbf{R}|^{1/2} |\mathbf{C}|^{1/2}} \int d\mathbf{y}^{N_r} e^{I(\mathbf{y})} \quad (27)$$

with

$$I(\mathbf{y}) = -\frac{1}{2} \left[(\mathbf{d} - \mathbf{G}(\mathbf{y}))^t \mathbf{R}^{-1} (\mathbf{d} - \mathbf{G}(\mathbf{y})) + (\mathbf{y} - \bar{\mathbf{y}})^t \mathbf{C}^{-1} (\mathbf{y} - \bar{\mathbf{y}}) \right] \quad (28)$$

Usually, instead of maximizing the observation PDF with respect to the GPs, we minimize the negative log-likelihood function $L(\Theta)$, which is defined by:

$$L(\Theta) = -\log(p(\mathbf{d}|\Theta)) \quad (29)$$

If the relation $\mathbf{G}(\mathbf{y})$ between the model parameters and the observation is linear, the summation in (27) can be handled analytically using Gaussian integration formula. When the relation is non-linear, we use an approximate method known as the Laplace method or in hydrology as the quasi linear geostatistical approximation [18]: it assumes that $I(\mathbf{y})$ is a peaked function, and that the near peak area represents the main contribution to the integral. The method can be divided into four steps:

1. Choose a first guess for the GPs, $\Theta_{k=0}$
2. First optimization.
 Θ_k is fixed. Find $\tilde{\mathbf{y}}$ which maximizes $I(\mathbf{y})$ using a Gauss-Newton method:

$$I(\tilde{\mathbf{y}}) = \max_{\mathbf{y}} (I(\mathbf{y})) \quad (30)$$

3. Computation of the negative log-likelihood for Θ_k .
 Compute $L(\Theta_k)$ analytically by linearizing $I(\mathbf{y})$ near $\tilde{\mathbf{y}}$ (Using Laplace Method)
4. Second optimization.
 Computation of the gradients $(\partial L(\Theta)/\partial \Theta)$ and of the Hessian matrix Ξ evaluated at Θ_k :

$$\Theta_{k+1} = \Theta_k - \Xi_k^{-1} \left(\frac{\partial L(\Theta)}{\partial \Theta} \right)_{\Theta_k} \quad (31)$$

Go back to step 2 until convergence is reached. More mathematical details can be found in Appendix A.

4 NUMERICAL SIMULATION

4.1 Well Test Example

In this part, we will briefly describe some numerical results. Reference well test simulations are performed using SIMTEST, a 3D finite volume simulator developed at the IFP [32]. First the diffusivity Equation (1) is solved. Then, using the bottomhole pressure measurements, we can compute the apparent permeability for a set of times.

$$k_{app}(t_i) = \frac{\mu Q}{4\pi h \left[\frac{dP(t)}{d \log t} \right]_i} \quad (32)$$

Since time increases geometrically the pressure log-derivative is computed using [22]:

$$\begin{aligned} \left[\frac{dP(t)}{d \log t} \right]_i &\approx \frac{\log(t_i/t_{i-1}) P_{i+1}}{\log(t_{i+1}t_{i-1}) \log(t_{i+1}/t_i)} \\ &+ \frac{\log(t_{i+1}t_{i-1}/t_i^2) P_i}{\log(t_{i+1}/t_{i-1}) \log(t_i/t_i)} \\ &- \frac{\log(t_{i+1}/t_i) P_{i-1}}{\log(t_i/t_{i-1}) \log(t_{i+1}/t_{i-1})} \end{aligned} \quad (33)$$

Figure 3 displays a zoom of a permeability field realization around the well (50 gridblocks in each direction).

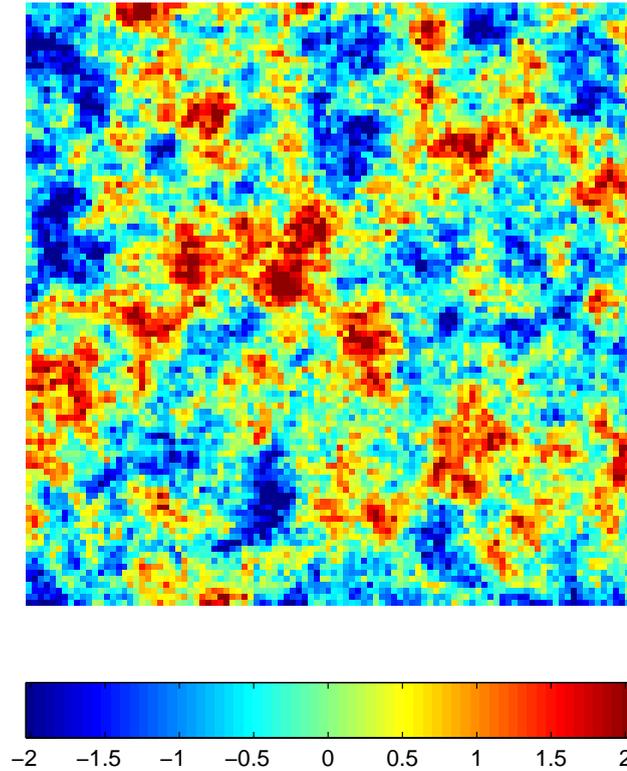


Figure 3

Map of a particular realization of a permeability field (in log-scale) with $\sigma_{ln}^2 = 1$, $l_c = 6$ gridblocks, $k_w = k_g = 100$ mD.

TABLE 1

Properties of the grid, of the well and of the fluid and rock.

Grid properties	
Number of gridblocks	$N_x \times N_y = 799 \times 799$
Gridblock dimensions	$\Delta_x \times \Delta_y = 10m \times 10m$
Well location	$(N_{xw}, N_{yw}) = (300, 300)$
Well properties and boundary conditions	
Well radius	$r_w = 8 \cdot 10^{-2}m$
Well rate	$Q = 1.157 \cdot 10^{-3}m^3/s$
BC on the reservoir border	no flux
Fluid and rock properties	
Porosity	$\phi = 0.1$
Fluid viscosity	$\mu = 1cp$
Total compressibility	$c_t = 10^{-4}bar$

The grid and the flow properties are found in the Table 1. The permeability field has been generated by a sequential Gaussian simulation provided by the GSLIB library [9]. It has a lognormal distribution with a variance, $\sigma_{ln}^2 = 1$ and a correlation length, $l_c = 6$ gridblocks. The well permeability and its geometric mean are $k_w = k_g = 100$ mD. A well test has been performed for 100 days. Figure 4 represents the apparent permeability with respect to time. We can see that the well test can be divided in four main periods:

1. The well test is only sensitive to the wellbore area and the apparent permeability is equal to the well permeability.
2. The well test is in a transient period and the fluctuations are more important. Actually, the well test performs a permeability average over volumes increasing with time.

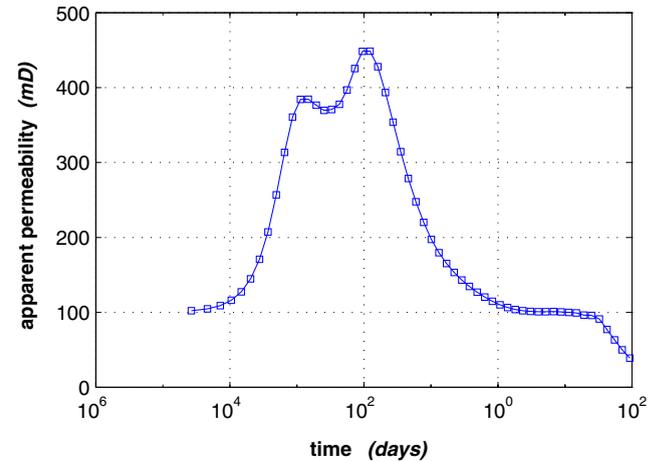


Figure 4

Apparent permeability $k_{app}(t)$ vs. time in days (log-scale for the time axis).

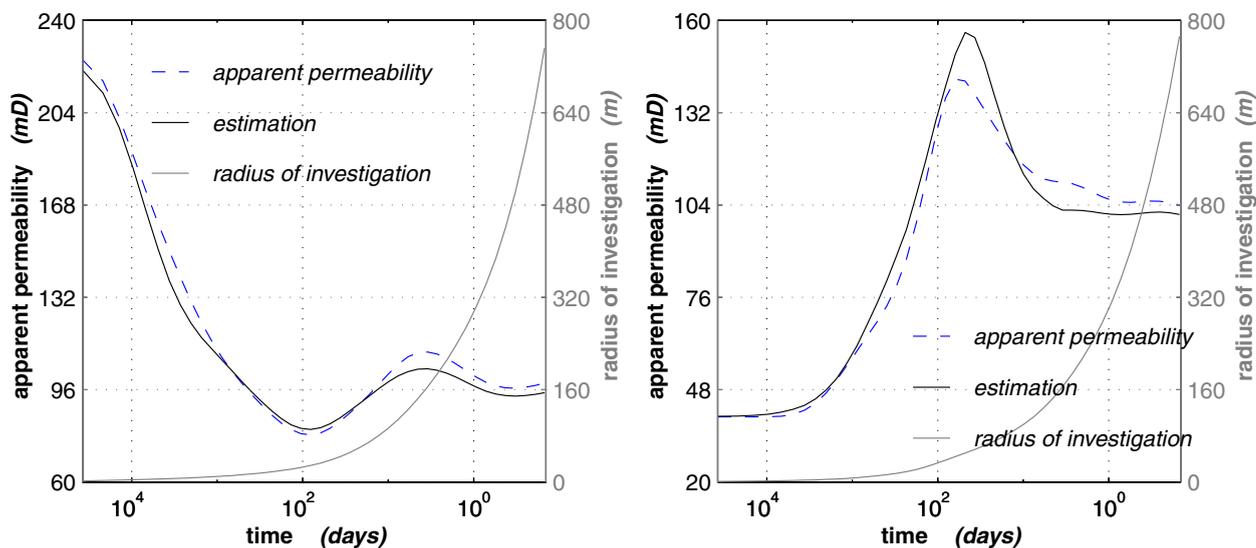


Figure 5

$k_{app}(t)$ and $\hat{k}(t)$ vs. time for two realizations of a permeability field with $l_c = 30\text{ m}$ and $\sigma_m^2 = 1$. The radius of investigation $R_{app}(t)$ is also plotted.

This averaging property pointed out by Oliver [4] has been discussed in the second Section (*cf. Eq. (7)*).

3. When the investigated volume is large enough, the apparent permeability tends to the geometric mean which is the effective permeability for a 2D lognormal permeability field [28].
4. During the last period, the apparent permeability decreases sharply since the boundary conditions start to influence the well test. The corresponding time is typically dependent on a maximum radius of investigation [31] the value of which is equal to one third of the well-boundary distance:

$$R_{\max} = \frac{\Delta_x \times N_x}{6} \quad (34)$$

This value is in agreement with Oliver's approximation [4] (Indeed we have seen in the second Section that the kernel $G(r, t)$ performs a permeability average within the range $[0.12R_{app}(t), 2.34R_{app}(t)]$)

Here, we pointed out the averaging process performed by the well test. Other effects, such as the influence of the permeability variability or of the correlation length can be found in Gautier [31].

4.2 Fast Evaluation of Well Tests

The fast evaluation of well tests is performed by first transforming the cartesian permeability grid into a homogeneous radial permeability field defined by a set of rings defined

by (12). Using (7) we are then able to compute an estimation of the apparent permeability. In Figure 5, we show two examples. The permeability field description can be found in the Table 1 and the 50 ring radii used for the estimation are in geometrical progression with a minimum and a maximum radius, equal to $\Delta_x/2$ and $N_x\Delta_x/2$ respectively. We can see that for both cases the difference between the apparent permeability and its estimation is small and fluctuates more for intermediate values of the investigated radius.

To check the global accuracy of the fast estimation, we computed the average of the estimation errors for well tests performed on 70 realizations of two permeability fields with $\sigma_m^2 = 1$ and $l_c = 50, 100\text{ m}$ (Fig. 6 and 7). The bold line represents the errors (in %) and the thin line, the averaged radius of investigation. Globally, the errors remain less than 10%, and are smaller for short and long times. Both estimation errors have a small peak at the beginning. This peak appearing for an averaged radius of investigation close to half the size of a gridblock is due to simulator numerical errors and is independent of the permeability field properties. Of more importance is the second peak which is located at a distance from the well close to the correlation length. This is in accordance with the homogenization process performed by the well test. For short times (short radius of investigation), we have seen that both the apparent permeability and the estimation are close to the wellbore area permeability. For long times, if the reservoir is large enough, the apparent permeability tends to the effective permeability. The estimation tends to the permeability of the last ring which is

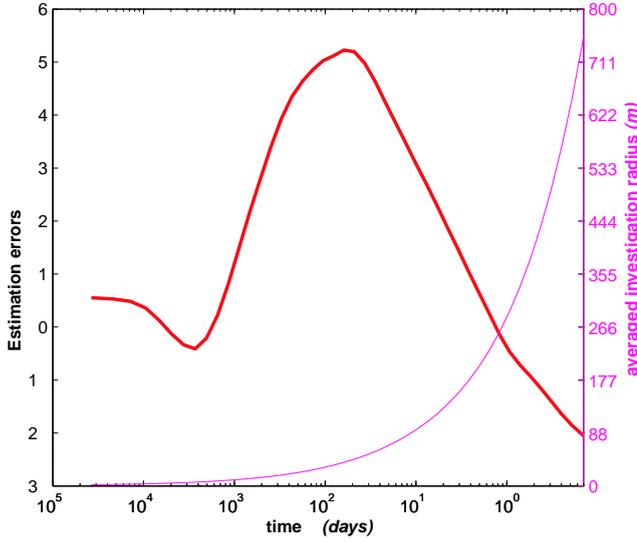


Figure 6

Estimation errors averaging (in %) vs. time for a sample of 70 realizations of a permeability field ($l_c = 50\text{ m}$, $\sigma_{ln}^2 = 1$). The “average” radius of investigation is represented by the thin line.

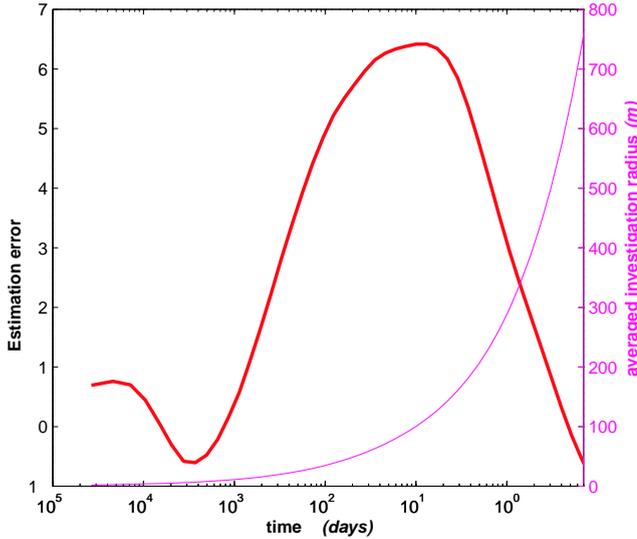


Figure 7

Estimation errors averaging (in %) vs. time for a sample of 70 realizations of a permeability field ($l_c = 100\text{ m}$, $\sigma_{ln}^2 = 1$). The “average” radius of investigation is represented by the thin line.

basically equal to the geometric mean if the last ring contains a sufficient amount of gridblocks. Therefore, the peak at the correlation length means that the apparent permeability is no longer sensitive to the well permeability and that it is not yet sensitive to the homogenization process. The fluctuations are bigger during this period and the approximation is slightly worse.

4.3 Geostatistical Parameter Estimation

We will now describe the GPs estimation using well test data. First we have to make some assumptions. As pointed out in the former section, the fluid flow simulator is not precise for short times (typically for $R_{app}(t) < \Delta_x/2$). For long times, the well test is sensitive to the boundary conditions which are independent of the permeability field structure. The last time step used in the optimization process will have a corresponding radius of investigation close to R_{max} defined in (34). Therefore the set of time steps $(t_i)_{i=1\dots N_t}$ will be selected such that:

$$R_{min} \leftarrow \frac{\Delta_x}{2} < R_{app}(t_i) < \frac{\Delta_x \times N_x}{6} \rightarrow R_{max} \quad (35)$$

We will consider that the observation error matrix is the identity matrix multiplied by a proportionality factor σ_w^2 . We will consider hereinafter that $\sigma_w^2 = (10\%)^2$. 10% represent the tolerance of the discrepancies between the observation data and the estimations. Therefore \mathbf{R} in (27) is defined by:

$$\mathbf{R} = \sigma_w^2 \mathbf{I}_{N_t} \quad (36)$$

Finally, we also include a parameter σ_c^2 which adds an information (obtained by other measurements): our knowledge of the well permeability⁽³⁾. We will also consider that this parameter is fixed and that $\sigma_c^2 = (10\%)^2$.

Let us consider an example of a well test performed on a 799×799 permeability field (the other data are in Table 1). The variance and the correlation length of the variogram are $\sigma_{ln}^2 = 1$ and $l_c = 50\text{ m}$ respectively. Figure 8 represents the apparent permeability with respect to the radius of investigation.

For this particular example, we have made a complete exploration of the GPs space and we have computed for each value of the GPs the negative log-likelihood function $L(\sigma_{ln}^2, l_c)$. The L contour plot is represented in Figure 9. We can see that the minimum is for values of the GPs close to $\sigma_{ln}^2 = 1$, $l_c = 40$. The minimum of L is more precisely determined in Figure 10 where we used the optimization algorithm with different first guesses. We see that, in all the cases, the algorithm converges to a same point ($\sigma_{ln}^2 = 1$, $l_c = 42$) in 5-10 iterations.

The next two figures (11(a) and 11(b)) represent the GPs estimation for well tests performed over two other realizations of the same permeability field and show us that the estimation depends on the realization. A statistical analysis of the estimation is therefore necessary.

Figure 12 represents the histograms of the correlation length and of the variance estimation based on well tests performed on 70 realizations of the permeability field that we have already considered. Statistical data are presented

(3) The developments of the optimization formalism considering this new information are straightforward but not presented here. The interested reader can refer to Chapter 5 of Gautier’s Thesis [31]

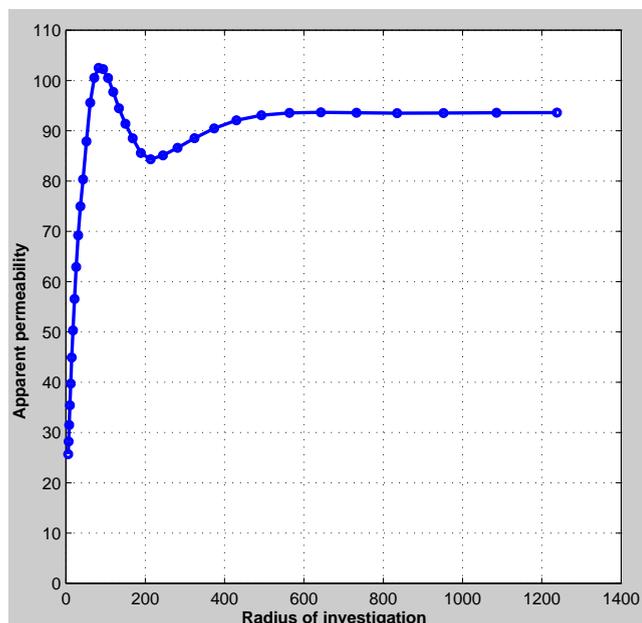


Figure 8

Well test performed on a realization of a 799×799 permeability field ($l_c = 50\text{ m}$, $\sigma_{ln}^2 = 1$).

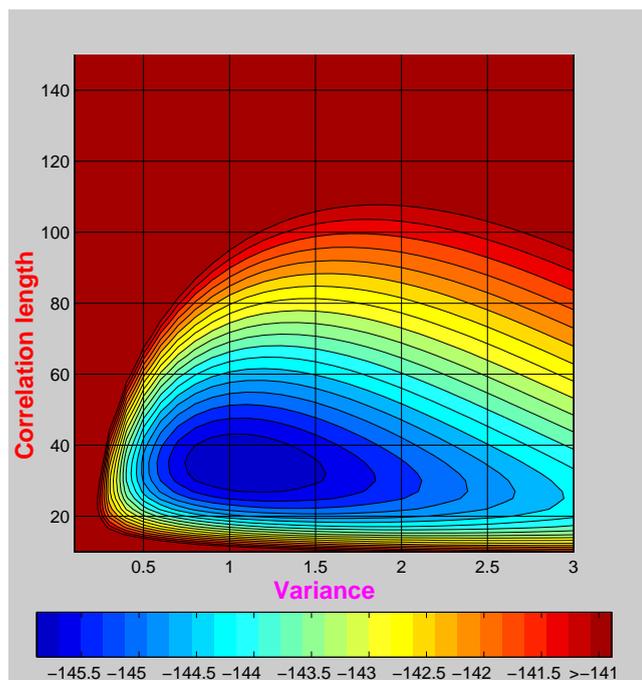


Figure 9

Negative log-likelihood function $L(\sigma_{ln}^2, l_c)$ for the well test plotted on Figure 8.

in the fourth row of Tables 2 and 3. For this particular example, we found that the averages of the variance and of the correlation length estimations (resp 0.85 and 73 m) gives

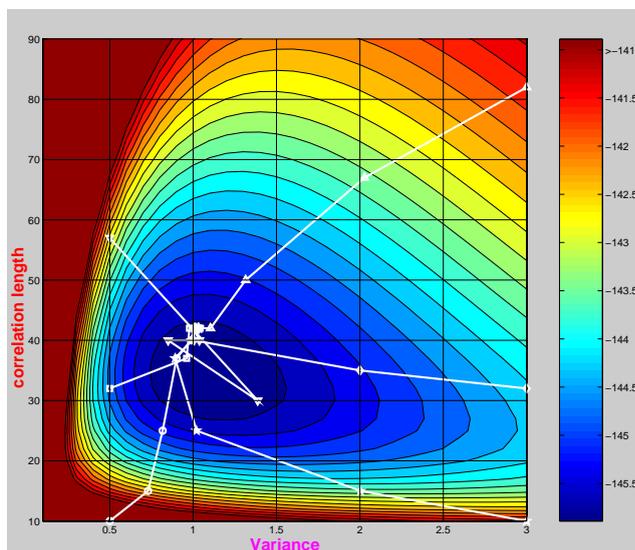


Figure 10

Minimization of the negative log-likelihood function $L(\sigma_{ln}^2, l_c)$ for the well test plotted on Figure 8 and for different first guesses.

a good order of magnitude of the theoretical values (resp. 1 and 50 m).

We performed the same study for other permeability fields, with $\sigma_{ln}^2 = 1, 2, 4$ and $l_c = 30, 50, 100\text{ m}$. In order to make comparisons between the results, we have plotted in Figures 13 and 14 the histograms of \tilde{l}_c/l_c and $\tilde{\sigma}_{ln}^2/\sigma_{ln}^2$ (\tilde{l}_c and $\tilde{\sigma}_{ln}^2$ represent the estimations). The statistics are presented in Tables 2 and 3. The main characteristics of these statistics are as follows:

- The mean of the correlation length estimation overestimates the theoretical value. The difference lies between 20 – 30 m. The standard deviation is quite high (more than half of the mean value).
- The mean of the variance estimation is close to its theoretical value. Nevertheless, the uncertainty remains high and is close to the mean value.

It is not clear if this overestimation is due to the estimation method or if it is simply due to an artifact which depends on the numerical scheme used to solve the well test Equation. In fact, the harmonic mean has been used to compute the transmissivity between two adjacent grid-blocks. The well test performs a moving average of the transmissivities which are a function of the permeabilities. Therefore, to a certain extent, the apparent permeability reflects the statistical properties of the transmissivities [33]. A theoretical study would be necessary to relate the statistical properties of the transmissivities and of the permeabilities.

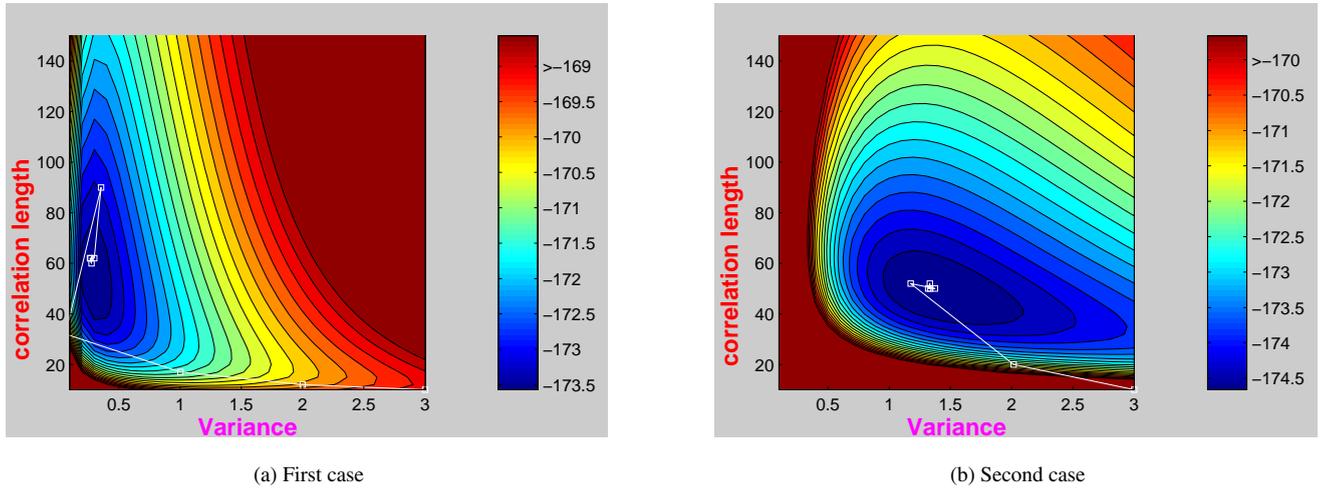


Figure 11

Minimization of the negative log-likelihood function $L(\sigma_{ln}^2, l_c)$ based on a well test performed on a permeability field with $(\sigma_{ln}^2 = 1, l_c = 50 m)$.

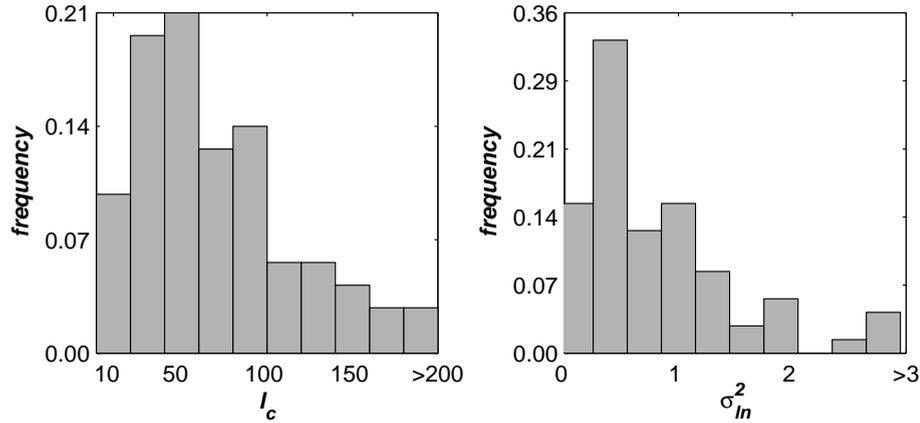


Figure 12

Histograms of l_c estimation (left) and of σ_{ln}^2 estimation (right) from well tests performed on 70 realizations of a permeability field with $(\sigma_{ln}^2 = 1, l_c = 50)$.

We have also studied the influence of the control parameters $(\sigma_w^2, \sigma_c^2, R_{min}, R_{max})$, and the evolution of the estimation statistics with respect to the number of tests [31]. For this latter effect, it has been shown, that, typically, 5-10 estimations of the GPs are needed to obtain a mean value close to the stable one.

CONCLUSIONS

Using a general Bayesian formalism, we built a method to estimate the dominant geostatistical parameters, l_c and σ_{ln}^2 , using well test data. The proposed method is rigorous, and

general. To get practical results, we had to couple up-scaling techniques and inversion procedures.

Due to the rather low dependency of the well test on these GPs, the obtained estimations have important fluctuations. Accurate well test data are needed, and the results are improved when a lot of independent well test data are available.

The proposed technique was implemented in the lognormal case, and it would need to be generalized to discontinuous permeability distributions that occur often in practice. It could also be adapted to account for other fluid-flow data, such as interference tests, and multiphase flow which are likely to contain more information.

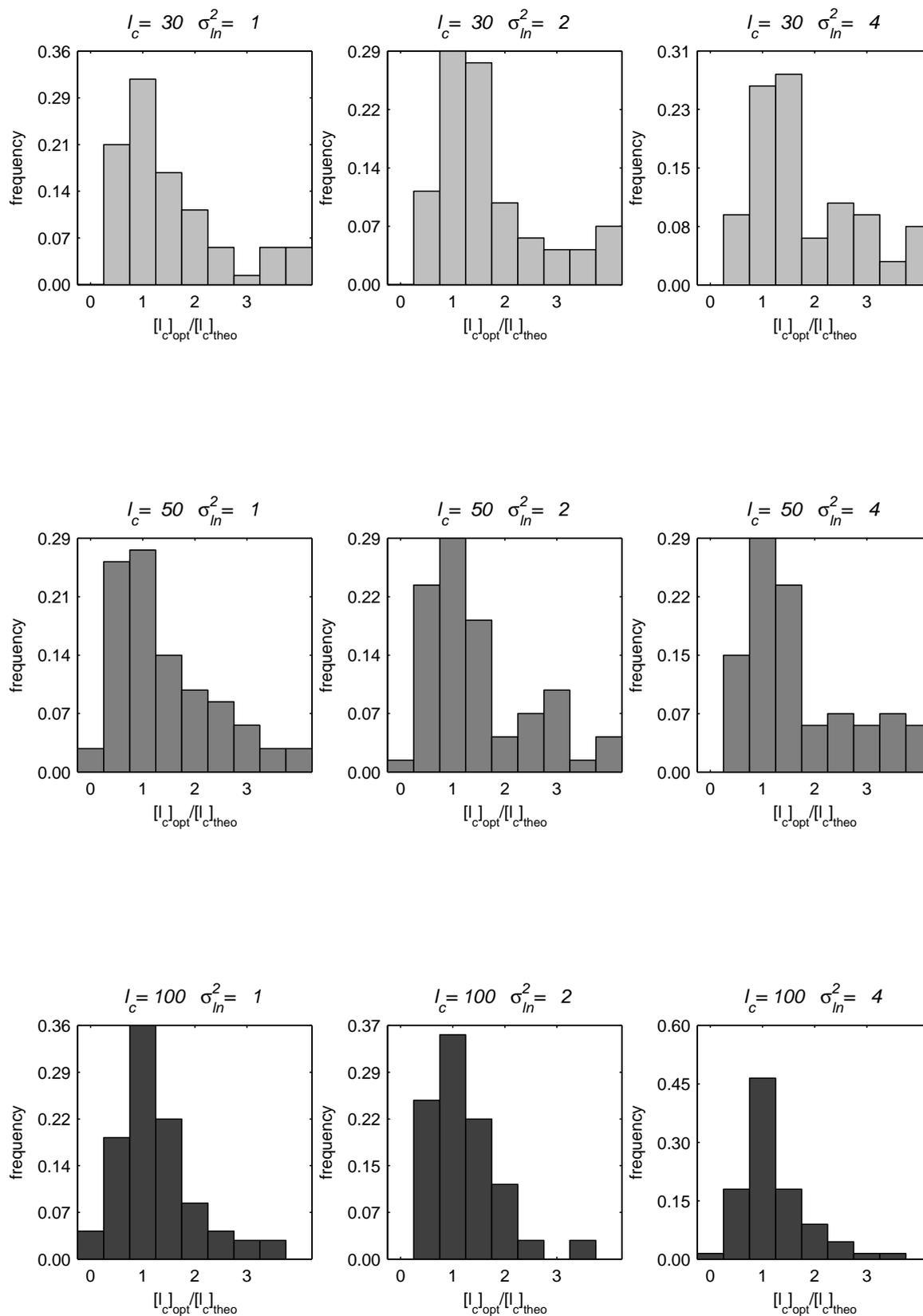


Figure 13

Histograms of \tilde{l}_c/l_c from well tests performed on 70 realizations of 9 permeability fields with $\sigma_{ln}^2 = 1, 2, 4$ and $l_c = 30, 50, 100$.

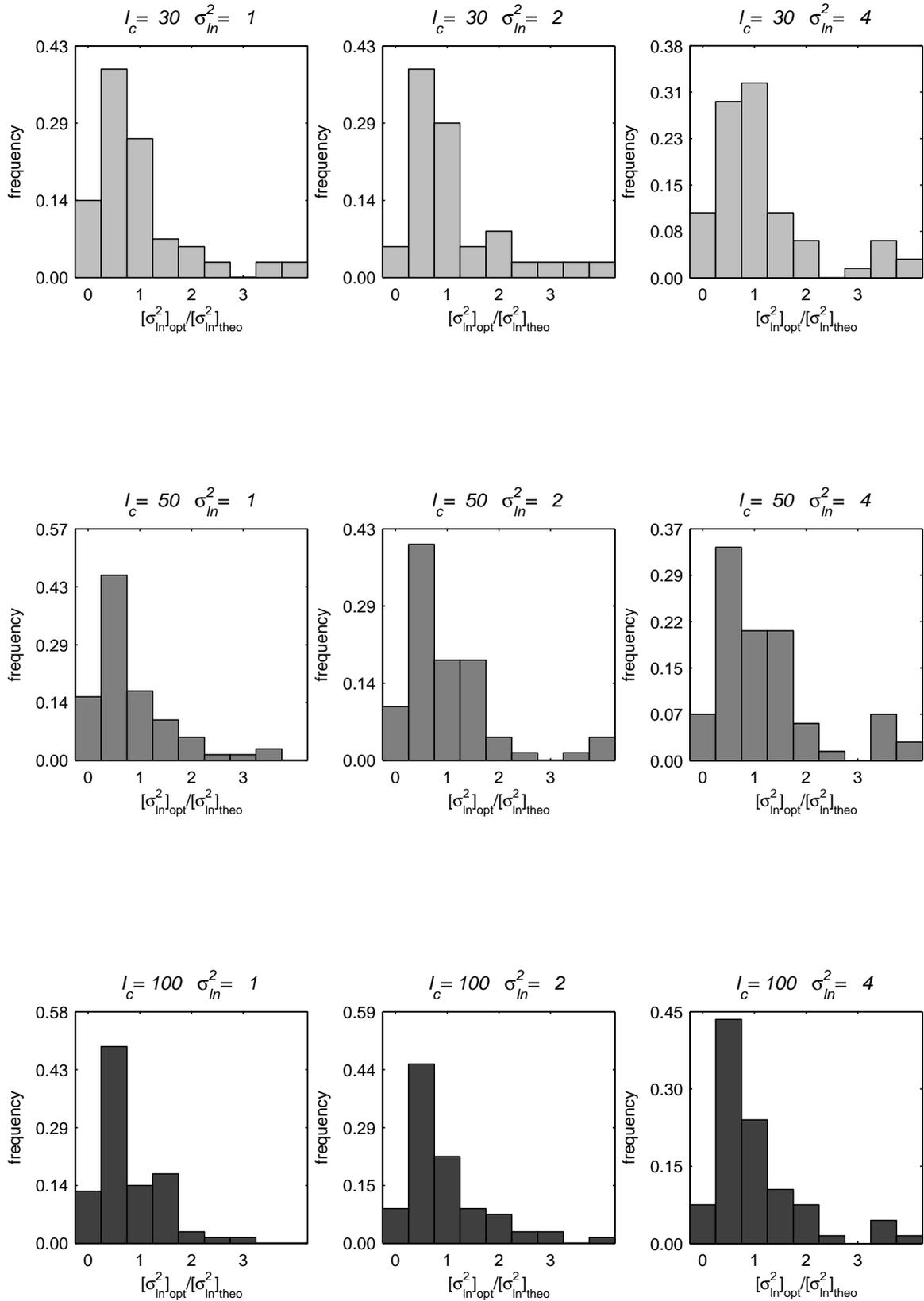


Figure 14

Histograms of $\tilde{\sigma}_{In}^2 / \sigma_{In}^2$ from well tests performed on 70 realizations of 9 permeability fields with $\sigma_{In}^2 = 1, 2, 4$ and $l_c = 30, 50, 100$.

TABLE 2

Statistics for the estimation of the correlation length for 70 realizations of nine permeability fields. The theoretical correlation length values are $l_c = 30, 50, 100\text{ m}$ and the log variance values are $\sigma_{ln}^2 = 1, 2, 4$.

l_c	σ_{ln}^2	l_c	l_c	quartiles		
Theo	Theo	Mean	SD	25%	50%	100%
$l_c = 30\text{ m}$	$\sigma_{ln}^2 = 1$	50	38	25	35	65
	$\sigma_{ln}^2 = 2$	54	39	27	45	65
	$\sigma_{ln}^2 = 4$	60	44	30	48	72
$l_c = 50\text{ m}$	$\sigma_{ln}^2 = 1$	73	56	35	60	95
	$\sigma_{ln}^2 = 2$	77	57	39	60	96
	$\sigma_{ln}^2 = 4$	89	61	48	70	115
$l_c = 100\text{ m}$	$\sigma_{ln}^2 = 1$	124	76	75	110	110
	$\sigma_{ln}^2 = 2$	126	80	75	120	120
	$\sigma_{ln}^2 = 4$	119	61	85	110	110

TABLE 3

Statistics for the estimation of σ_{ln}^2 for 70 realizations of nine permeability fields. The theoretical correlation length values are $\sigma_{ln}^2 = 30, 50, 100\text{ m}$ and the log variance values are $\sigma_{ln}^2 = 1, 2, 4$.

Theoretical l_c	Theoretical σ_{ln}^2	Mean of σ_{ln}^2 estimation	Standard deviation of σ_{ln}^2 estimation
$l_c = 30\text{ m}$	$\sigma_{ln}^2 = 1$	0.77	0.92
	$\sigma_{ln}^2 = 2$	2.27	2.09
	$\sigma_{ln}^2 = 4$	4.63	3.97
$l_c = 50\text{ m}$	$\sigma_{ln}^2 = 1$	0.85	0.78
	$\sigma_{ln}^2 = 2$	2.09	2.23
	$\sigma_{ln}^2 = 4$	4.84	4.17
$l_c = 100\text{ m}$	$\sigma_{ln}^2 = 1$	0.78	0.60
	$\sigma_{ln}^2 = 2$	1.87	1.69
	$\sigma_{ln}^2 = 4$	4.05	3.43

REFERENCES

- Jacquard, P. and Jain, C. (1965) Permeability distribution from field pressure data. *Soc. Pet. Eng. J.*, **5**(4):281–294.
- Chavent, G., Dupuy, M. and Lemonnier, P. (1975) History matching by use of optimal control. *Soc. Pet. Eng. J.* **74–86**, **15**(1):74–86.

- Gavalas, G.R. and Shah, P.C. (1976) Reservoir history matching by bayesian estimation. *Soc. Pet. Eng. J.* **337–350**, **16**(6):337–350.
- Oliver, D.S. (1989) The averaging process in permeability estimation from well test data. Number SPE 19845, pages 813–833, San Antonio.
- Alabert, F.G. (1989) Constraining description of randomly heterogeneous reservoirs to pressure test data: A monte carlo study. Number SPE 19600, pages 207–221, San Antonio.
- Sagar, R.K. (1993) *Reservoir description by integration of well test data and spatial statistics*. PhD thesis, University of Tulsa.
- Sun, N.Z. (1993) *Inverse Problems in Groundwater Modeling*. Elsevier.
- Yeh, W.W.G., Yoon, Y.S. and Lee, K.S. (1981) Aquifer parameter identification with kriging and optimum parametrization. *Water Resources Research*, **17**(3):664–672.
- Deutsch, C.V. and Journel, A.G. (1992) *Geostatistical Software Library and User's Guide*. Oxford University Press.
- Eschard, R., Doligez, B., Rahon, D., Ravenne, C. and Leloch, G. A new approach for reservoirs description and simulation using geostatistical methods. Edinburg, 21/22 February 1991.
- Yadavalli, S.K., Roadifer, R.D., Jones, J.R., Krase, L.D. and Yeh, N.S. Use of pressure transient data to obtain geostatistical parameters for reservoir characterization. Number SPE 28432, pages 719–732, New Orleans, LA, September 1994.
- Yadavalli, S.K., Jones, J.R., Krase, L.D., Yeh, N.S. and Roadifer, R.D. Use of pressure transient data to estimate permeability variograms. Number SPE 30587, pages 599–608, Dallas, October 1995.
- Feitosa, G. (1993) *Well test analysis for heterogeneous reservoirs*. PhD thesis, University of Tulsa.
- Yortsos, T.C. Estimation of permeability variogram from stochastic analysis of pressure transients of multiple wells in heterogeneous reservoirs. Number SPE 36510, pages 599–608, Denver, Colorado, October 1996.
- Dagan, G. (1989) *Flow and Transport in Porous Formations*. Springer-Verlag, Berlin.
- Carrera, J. and Neuman, S.P. (1986) Estimation of aquifer parameters under transient and steady state conditions: 1. maximum likelihood method incorporating prior information. *Water Resources Research*, **22**(2):225–233.
- Kitanidis, P.K. (1996) On the geostatistical approach to the inverse problem. *Advances in Water Resources*, **31**(10):2411–2419.
- Kitanidis, P.K. (1995) Quasi-linear geostatistical theory for inversing. *Water Resources Research*, **31**(10):2411–2419.
- Kitanidis, P.K. and Vomvoris, E. (1983) A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulation. *Water Resources Research*, **19**(3):677–690.
- Hoeksema, R.J. and Kitanidis, P.K. (1984) An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. *Water Resources Research*, **20**(7):1003–1020.
- Daviau, F. (1986) *Interprétation des essais de puits*. Éditions Technip.

- 22 Horne, R.N. (1990) *Modern well test analysis*. Petroway.
- 23 Kruehl Romeu, R. (1994) *Écoulements en milieux hétérogènes : prise de moyenne de perméabilité en régime permanent et transitoire*. PhD thesis, Université de Paris VI.
- 24 Bourdarot, G. (1996) *Essais de puits : méthodes d'interprétation*. Éditions Technip.
- 25 Gelhar, L.W. (1993) *Stochastic Subsurface Hydrology*. Prentice Hall.
- 26 Indelman, P. (1996) Averaging of unsteady flow in heterogeneous media of stationary conductivity. *J. Fluid. Mech.*, **310**:39–60.
- 27 Nøtinger, B. and Gautier, Y. (1998) Use of the fourier-laplace transform and of diagrammatical methods to interpret pumping tests in heterogeneous media. *Advances in Water Resources*, **21**:581–590.
- 28 Matheron, G. (1967) *Eléments pour une théorie des milieux poreux*. Masson, Paris.
- 29 Hoeksema, R.J. and Kitanidis, P.K. (1985) Analysis of the spatial structure of properties of selected aquifers. *Water Resources Research*, **21**(4):563–572.
- 30 Carrera, J. and Neuman, S.P. (1986) Estimation of aquifer parameters under transient and steady state conditions: 1. Uniqueness, stability, and solution algorithms. *Water Resources Research*, **22**(2):211–227.
- 31 Gautier, Y. (1997) *Estimation de paramètres géostatistiques à partir de tests de puits*. PhD thesis, Université de Paris VI.
- 32 Blanc, G., Guérillot, D., Rahon, D. and Roggero, F. Building geostatistical models constrained by dynamic data- a posteriori constraint. Number SPE 35478, pages 19–33, Stavanger, Norway, April 1996.
- 33 Nøtinger, B. and Kruehl Romeu, R. (1994) About the influence of the transmissibility evaluation rule about the results of fluid flow simulations in heterogeneous reservoirs. *Water Resources Research*.
- 34 Tarantola, A. (1987) *Inverse Problem Theory*. Elsevier.
- 35 Priestley, M.B. (1987) *Spectral Analysis and Time Series*. Academic Press.

Final manuscript received in February 2004

APPENDIX A

Detailed Derivations of the Optimization Algorithm

Here, we will describe the different steps of the optimization algorithm. An important part of these results is based on the research of Kitanidis [17]. The first step is equivalent to minimizing $-2 I(\mathbf{y})$ defined in (28) and appears in many papers. One can refer to Tarantola [34] for a thorough review. We use an iterative Gauss-Newton method:

$$\mathbf{y}_{n+1} = \mathbf{y}_n - \mathbf{H}_{\mathbf{y}_n}^{-1} \left(\frac{\partial I(\mathbf{y})}{\partial \mathbf{y}} \right)_{\mathbf{y}_n} \quad (37)$$

The $(N_r \times N_r)$ matrix, $\mathbf{H}_{\mathbf{y}_n}$, is the Hessian of $I(\mathbf{y})$ computed at the point \mathbf{y}_n :

$$\mathbf{H}_{\mathbf{y}_n} = \left(\frac{\partial^2 I(\mathbf{y})}{\partial \mathbf{y}^2} \right)_{\mathbf{y}_n} \quad (38)$$

Exact computation of the Hessian requires the second order derivatives of $I(\mathbf{y})$. But in the Gauss-Newton method, it is not necessary to have a very accurate evaluation of the Hessian, which is computed by differentiating twice Equation (28), ignoring second-order terms:

$$\mathbf{H}_{\mathbf{y}_n} = \mathbf{J}_{\mathbf{y}_n}^t \mathbf{R}^{-1} \mathbf{J}_{\mathbf{y}_n} + \mathbf{C}^{-1} \quad (39)$$

where $\mathbf{J}_{\mathbf{y}_n}$ is the $(N_r \times N_r)$ Jacobian matrix of $\mathbf{G}(\mathbf{y})$ evaluated at \mathbf{y}_n (see its expression in Appendix B). By deriving Equation (28):

$$\left[\frac{\partial I(\mathbf{y})}{\partial \mathbf{y}} \right]_{\mathbf{y}_n} = -\mathbf{J}_{\mathbf{y}_n} \mathbf{R}^{-1} (\mathbf{d} - \mathbf{G}(\mathbf{y}_n)) + \mathbf{C}^{-1} (\mathbf{y}_n - \bar{\mathbf{y}}) \quad (40)$$

Therefore, including Equations (39) and (40) in Equation (37), we obtain:

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \left(\mathbf{J}_{\mathbf{y}_n}^t \mathbf{R}^{-1} \mathbf{J}_{\mathbf{y}_n} + \mathbf{C}^{-1} \right)^{-1} \times \left(\mathbf{J}_{\mathbf{y}_n}^t \mathbf{R}^{-1} (\mathbf{d} - \mathbf{G}(\mathbf{y}_n)) + \mathbf{C}^{-1} (\mathbf{y}_n - \bar{\mathbf{y}}) \right) \quad (41)$$

Or using two classical relations⁽⁴⁾ and a some matrix algebra:

$$\mathbf{y}_{n+1} = \bar{\mathbf{y}} + \mathbf{C} \mathbf{J}_{\mathbf{y}_n}^t (\mathbf{R} + \mathbf{J}_{\mathbf{y}_n} \mathbf{C} \mathbf{J}_{\mathbf{y}_n}^t)^{-1} \times ((\mathbf{d} - \mathbf{G}(\mathbf{y}_n)) + \mathbf{J}_{\mathbf{y}_n} (\mathbf{y}_n - \bar{\mathbf{y}})) \quad (42)$$

We will use hereinafter this formula which involves only one computation of the inverse of a matrix. After convergence, when the value $\tilde{\mathbf{y}}$ has been determined, a first order development of $\mathbf{G}(\mathbf{y})$ can be handled in the vicinity of $\tilde{\mathbf{y}}$:

$$\mathbf{G}(\mathbf{y}) = \mathbf{G}(\tilde{\mathbf{y}}) + \mathbf{J}_{\tilde{\mathbf{y}}} (\mathbf{y} - \tilde{\mathbf{y}}) \quad (43)$$

$$(4) \left(\mathbf{J}_{\mathbf{y}_n}^t \mathbf{R}^{-1} \mathbf{J}_{\mathbf{y}_n} + \mathbf{C}^{-1} \right)^{-1} = \mathbf{C} - \mathbf{C} \mathbf{J}_{\mathbf{y}_n}^t (\mathbf{R} + \mathbf{J}_{\mathbf{y}_n} \mathbf{C} \mathbf{J}_{\mathbf{y}_n}^t)^{-1} \mathbf{J}_{\mathbf{y}_n} \mathbf{C} \\ \left(\mathbf{J}_{\mathbf{y}_n}^t \mathbf{R}^{-1} \mathbf{J}_{\mathbf{y}_n} + \mathbf{C}^{-1} \right) \mathbf{J}_{\mathbf{y}_n}^t \mathbf{R}^{-1} = \mathbf{C} \mathbf{J}_{\mathbf{y}_n}^t (\mathbf{R} + \mathbf{J}_{\mathbf{y}_n} \mathbf{C} \mathbf{J}_{\mathbf{y}_n}^t)^{-1}$$

Replacing $\mathbf{G}(\mathbf{y})$ by its value near $\tilde{\mathbf{y}}$ in Equation (27), one obtains after some computing, using classical Gaussian function integration [34]:

$$p(\mathbf{d}|\Theta) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{d}-\bar{\mathbf{d}})^t \Sigma^{-1} (\mathbf{d}-\bar{\mathbf{d}})} \quad (44)$$

with:

$$\bar{\mathbf{d}} = \langle \mathbf{d} \rangle = \mathbf{G}(\tilde{\mathbf{y}}) + \mathbf{J}_{\tilde{\mathbf{y}}} (\bar{\mathbf{y}} - \tilde{\mathbf{y}}) \quad (45)$$

$$\Sigma = \mathbf{R} + \mathbf{J}_{\tilde{\mathbf{y}}} \mathbf{C} \mathbf{J}_{\tilde{\mathbf{y}}}^t \quad (46)$$

Now, we have to maximize $p(\mathbf{d}|\Theta)$ with respect to Θ , and therefore to minimize the opposite of the negative log-likelihood function $L(\Theta)$ defined by:

$$L(\Theta) = -2 \log(p(\mathbf{d}|\Theta)) \\ = |\Sigma| + (\mathbf{d} - \bar{\mathbf{d}})^t \Sigma^{-1} (\mathbf{d} - \bar{\mathbf{d}}) \quad (47)$$

This minimization is performed using a second Gauss-Newton algorithm. Assuming that an estimation Θ_n has been computed, we have:

$$\Theta_{n+1} = \Theta_n - \lambda_n \Xi_n^{-1} \left(\frac{\partial L(\Theta)}{\partial \Theta} \right)_{\Theta_n} \quad (48)$$

where λ_n is a scalar parameter which can be chosen to improve the convergence rate of the algorithm, $(\partial L(\Theta)/\partial \Theta)$ and Ξ are the gradient and the Hessian of L with respect to Θ . The gradient is computed using matrix derivation formulae⁽⁵⁾

$$\frac{\partial L(\Theta)}{\partial \Theta} = \text{Tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \Theta} \right) \\ - (\mathbf{d} - \bar{\mathbf{d}})^t \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \Theta} \Sigma^{-1} \right) (\mathbf{d} - \bar{\mathbf{d}}) \\ + 2 (\mathbf{d} - \bar{\mathbf{d}})^t \Sigma^{-1} \frac{\partial \bar{\mathbf{d}}}{\partial \Theta} \quad (49)$$

When observations are sufficiently numerous, the Hessian can be well approximated by the information Fisher matrix [35] which is defined as:

$$\mathbf{M}_n = \left\langle \left(\frac{\partial L(\Theta)}{\partial \Theta} \right) \left(\frac{\partial L(\Theta)}{\partial \Theta} \right)^t \right\rangle_{\Theta_n} \quad (50)$$

or, equivalently:

$$\mathbf{M}_n = \left\langle \left(\frac{\partial^2 L(\Theta)}{\partial \Theta^2} \right) \right\rangle_{\Theta_n} \quad (51)$$

(5) Assuming that the inverse and the derivatives (with respect to Θ) of a matrix \mathbf{A} exist, we have the following equations:

$$\frac{\partial |\mathbf{A}|}{\partial \Theta} = \text{Tr} \left(\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \Theta} \right) \\ \frac{\partial \mathbf{A}^{-1}}{\partial \Theta} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \Theta} \mathbf{A}^{-1}$$

Therefore the Hessian matrix Ξ_n can be asymptotically estimated by its mean value given by the information Fisher matrix. After calculation:

$$\mathbf{M}_n = \text{Tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \Theta} \Sigma^{-1} \frac{\partial \Sigma}{\partial \Theta} \right) + \frac{\partial \bar{\mathbf{d}}}{\partial \Theta} \Sigma^{-1} \frac{\partial \bar{\mathbf{d}}}{\partial \Theta} \quad (52)$$

and finally to update Θ_n , we use the following equation:

$$\Theta_{n+1} = \Theta_n - \lambda_n \mathbf{M}_n^{-1} \left(\frac{\partial L(\Theta)}{\partial \Theta} \right)_{\Theta_n} \quad (53)$$

Expressions of the Jacobian matrix of $\mathbf{G}(\mathbf{y})$, and of the Fisher matrix can be found in Appendix B.

APPENDIX B

Derivation of the Jacobian and the Fisher Matrices

Expression of the Jacobian Matrix

$$\mathbf{J} = \frac{\partial \mathbf{G}(\mathbf{y})}{\partial \mathbf{y}} = \begin{bmatrix} \frac{\partial G^1(\mathbf{y})}{\partial y_1} & \frac{\partial G^1(\mathbf{y})}{\partial y_2} & \dots & \frac{\partial G^1(\mathbf{y})}{\partial y_{N_r}} \\ \frac{\partial G^2(\mathbf{y})}{\partial y_1} & \frac{\partial G^2(\mathbf{y})}{\partial y_2} & \dots & \frac{\partial G^2(\mathbf{y})}{\partial y_{N_r}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G^{N_t}(\mathbf{y})}{\partial y_1} & \frac{\partial G^{N_t}(\mathbf{y})}{\partial y_2} & \dots & \frac{\partial G^{N_t}(\mathbf{y})}{\partial y_{N_r}} \end{bmatrix} \quad (54)$$

using Equations (7) and (26), and performing the differentiations:

$$\mathbf{J} = - \begin{bmatrix} G_{11}e^{-y_1} & G_{12}e^{-y_2} & \dots & G_{1N_r}e^{-y_{N_r}} \\ G_{11}e^{-y_1} & G_{12}e^{-y_2} & \dots & G_{1N_r}e^{-y_{N_r}} \\ \vdots & \vdots & \ddots & \vdots \\ G_{N_t1}e^{-y_1} & G_{N_t2}e^{-y_2} & \dots & G_{N_tN_r}e^{-y_{N_r}} \end{bmatrix} \quad (55)$$

which is the expression of the Jacobian of $\mathbf{G}(\mathbf{y})$.

Expression of the Fisher Matrix

To compute the Fisher matrix (52) we have to find the derivatives of Σ and $\bar{\mathbf{d}}$ with respect to Θ . Using (45) and (46), we have:

$$\frac{\partial \bar{\mathbf{d}}}{\partial \Theta} = \mathbf{J}_{\bar{\mathbf{y}}} \frac{\partial \bar{\mathbf{y}}}{\partial \Theta} \quad (56)$$

and

$$\frac{\partial \Sigma}{\partial \Theta} = \mathbf{J}_{\bar{\mathbf{y}}} \frac{\partial \mathbf{C}}{\partial \Theta} \mathbf{J}_{\bar{\mathbf{y}}}^t \quad (57)$$

To compute the covariance matrix derivative with respect to the GPs, we decompose the sums in (25) into two different contributions:

$$\begin{aligned} C_{ij} = & \sum_{(k \in \Omega_i), (l \in \Omega_j), k \neq l} \omega_k \omega_l \rho(d_{kl}) \\ & + \sum_{(k \in \Omega_i), (l \in \Omega_j), k=l} \omega_k \omega_l \rho(0) \end{aligned} \quad (58)$$

Let's assume that the correlation function is exponential with a variance, σ_{ln}^2 and a correlation length l_c , ie $\rho(r) = \sigma_{ln}^2 e^{-r/l_c}$. Therefore $\Theta = [\sigma_{ln}^2 \ l_c]^t$ and:

$$\frac{\partial C_{ij}}{\partial \sigma_{ln}^2} = \frac{C_{ij}}{\sigma_{ln}^2} \quad (59)$$

and

$$\frac{\partial C_{ij}}{\partial l_c} = \frac{1}{l_c^2} D_{ij} \quad (60)$$

with

$$D_{ij} = \sum_{(k \in \Omega_i), (l \in \Omega_j), k \neq l} \omega_l \omega_k d_{kl} \rho(d_{kl}) \quad (61)$$

The derivative components of $\bar{\mathbf{y}}$ defined by (24) with respect to the GPs, are:

$$\frac{\partial \bar{y}_j}{\partial \theta_i} = \sum_{l \in \Omega_j} \omega_l \frac{\partial \langle Y_l \rangle}{\partial \theta_i} \quad (62)$$

Since we assume in all our applications that the mean of the permeability field is constant:

$$\frac{\partial \bar{\mathbf{y}}}{\partial \Theta} = \mathbf{0} \quad (63)$$