Available online at www.sciencedirect.com

ScienceDirect

Advances in Water Resources 29 (2006) 1678–1689

Advances in Water Resources

www.elsevier.com/locate/advwatres

Pilot points method incorporating prior information for solving the groundwater flow inverse problem

Andrés Alcolea *, Jesús Carrera, Agustín Medina

School of Civil Engineering, Technical University of Catalonia (UPC), Campus Nord D2, 1-3, Jordi Girona St., 08034 Barcelona, Spain

Received 9 November 2004; received in revised form 13 December 2005; accepted 19 December 2005
Available online 8 February 2006

Abstract

The pilot points method is often used in nonlinear geostatistical calibration. The method consists of estimating the values of the hydraulic properties at a set of arbitrary (pilot) points so as to best fit the aquifer response as measured by available indirect observations (i.e., heads or drawdowns). Though this method remains general and appealing, no prior information of the hydraulic properties is usually included in the optimization process, which constrains the number of pilot points to ensure stability. In this paper, we present a modification of the pilot points method, including prior information in the optimization process by adding a plausibility term to the objective function to be minimized. This results from formulating the inverse problem in a maximum likelihood framework. The performance of the method is tested on a synthetic example. Results show that including the plausibility term improves the identification of heterogeneity. Furthermore, this term makes the inverse problem more stable and allows the use of larger number of pilot points, thus improving the identification of the heterogeneity as well. Therefore, the use of the plausibility term is recommended.

Keywords: Geostatistics; Groundwater modeling; Inverse problem; Regularization; Synthetic problem

1. Introduction

Heterogeneity plays an important role for groundwater flow and contaminant transport in geological formations and needs to be accounted for in meaningful models. Inverse modeling represents a powerful tool to quantify the influence of heterogeneity [6,7,15,31,44]. In order to identify heterogeneity, the groundwater inverse problem is usually formulated in a geostatistical framework. Early methods [25,39,20] aimed at estimating at every point the departure from the mean log transmissivity implied by head data. These formulations are linear and their computational cost moderate. They often work fine [45], but as complexity increases, iterating is needed [9,8,45]. However, geostatistical formulations estimate log transmissivity at every cell (or element), so that the nonlinear solutions become too expensive unless special numerical methods, such as the adjoint state method, are used [33]. This allows successful practical application [34,38] to complex problems, but it is difficult to program. Therefore, one needs to reduce the number of unknowns by means of some parameterization scheme. MacLaughlin and Townley [31] discuss a number of such schemes. However, the one that is most flexible and consistent with the geostatistical assumptions is the pilot points method. Hence, it is not surprising that it has gained steam in recent years.

The pilot points method consists of (Fig. 1): (1) generating an initial spatially correlated field given a geostatistical model, (2) defining an interpolation method to obtain the value of the hydraulic properties over the model domain on the basis of their measurements and their values at the pilot point locations (model parameters) and (3) optimizing the value of the model parameters in such a way that the interpolated field (step 2) minimizes an objective function measuring the misfit between calculated and measured data (often, only heads are considered). Thus, finding the opti-

* Corresponding author. Fax: +34 93 4017251.
E-mail addresses: andres.alcolea@upc.edu (A. Alcolea), jesus.carrera@upc.edu (J. Carrera), agustin.medina@upc.edu (A. Medina).

0309-1708/$ - see front matter
doi:10.1016/j.advwatres.2005.12.009

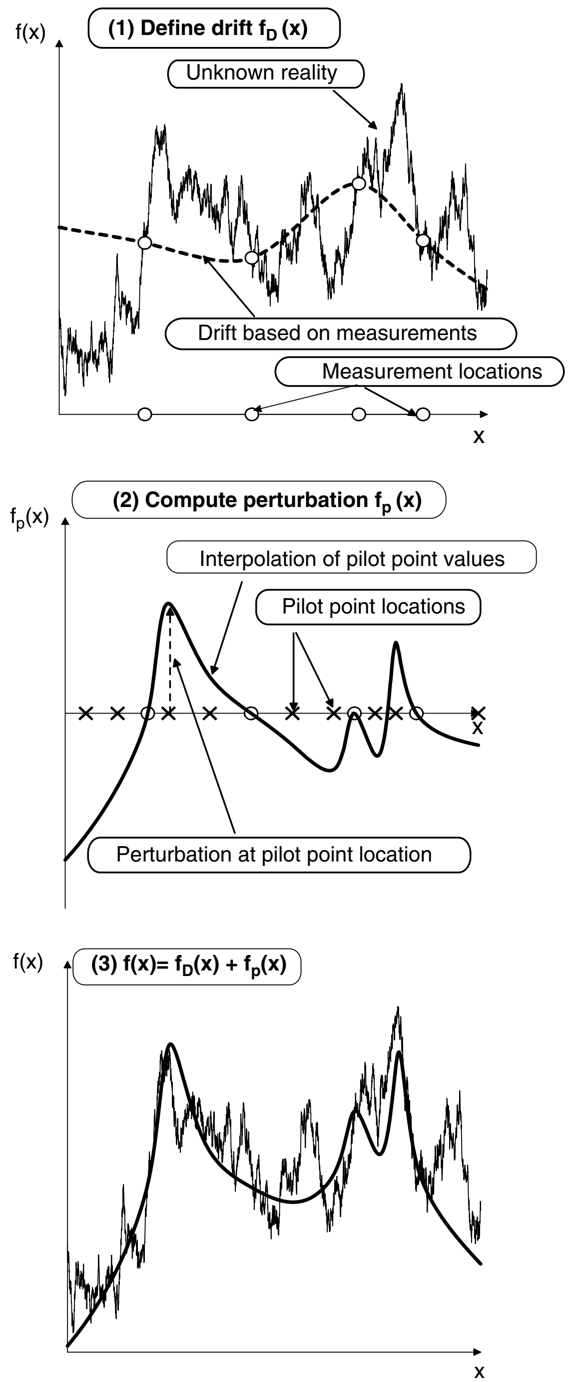

Fig. 1. Schematic description of the pilot points method for defining a spatial random function $f(x)$, as the sum of a drift, $f_D(x)$, and a perturbation $f_p(x)$. The drift is defined by conditioning on available measurements. The perturbation is obtained from interpolation of the unknown pilot point values (model parameters), which are optimized so as to obtain a good fit with available indirect observations (i.e., heads).

mum value of model parameters becomes an optimization problem. Notice that steps 2 and 3 imply the perturbation of the field generated in step 1.

This method was originally devised by de Marsily [14], but has undergone several modifications. RamaRao et al. [37] and Gómez-Hernández et al. [19] included conditional simulations in the generation of the initial field. The location of the pilot points has been studied by Lavenue and Pickens [27] and Hendricks-Franssen [23], among others. The pilot points method has become widely used and has been applied to different problems [24,42].

However, Cooley and Hill [12] and Cooley [13] identified some drawbacks. These arise from neglecting sources of model inaccuracy (i.e., errors in the conceptual model) and overparameterizing. The latter leads to instability of the optimization problem [21,22]. Instability implies (a) large values of some model parameters due to unbounded fluctuations [2], which also causes (b) large "jumps" in the value of the hydraulic properties over small distances and (c) large second derivatives of the hydraulic property field. Tactics to combat instability are based on addressing these effects.

A possibility to fight unbounded fluctuations consists of imposing upper and lower bounds on the model parameters. In the context of pilot points, RamaRao et al. [37] and Gómez-Hernández et al. [19] use this tactic. However, in general, this approach simply causes the solution to fluctuate between those arbitrary bounds, but its reliability is not improved [36].

Instability is attributed to overparameterization. Thus, the second tactic to circumvent instabilities consists of reducing the number of model parameters. In the context of pilot points, a common approach consists of starting with a single pilot point and adding new candidates at each iteration of the optimization process [37]. New pilot point locations are set according to their ability for reducing the objective function, measured by the sensitivity coefficients [27]. Other researchers predefine the number of pilot points, whose location can be fixed (e.g. regular grids of 2–3 pilot points per correlation range in each direction [3,23]) or vary randomly during the optimization process [23]. Though the use of a small number of pilot points may overcome instabilities, it leads to three side effects: first, the identification of the heterogeneity loses resolution; second, the role of a good geostatistical characterization becomes critical [16] and third, the problem is very sensitive to the location of the pilot points [27].

A tactic to avoid large jumps in estimated parameters consists of penalizing them by adding regularization terms to the objective function: Tikhonov [40,41] imposes penalties to large values of the model parameters and unwarranted oscillations are penalized by Emsellem and de Marsily [17]. However, we argue that valuable information about model parameters is not included in the optimization process. This can be done by adding a plausibility term to the objective function, which helps in solving the above problems, while allowing a formal posing of the inversion [4,5]. The plausibility term is essentially a regularization criterion that penalizes the departure of the model parameters from their prior estimates (derived from the prior information of the hydraulic properties).

In the context of pilot points, the inclusion of a regularization term has not been excluded from debate. Two trends can be found in the literature. On the one hand, Certes and de Marsily [10] reject the use of this term, questioning its performance, as it depends to a large extent on the reliability of prior estimates. RamaRao et al. [37] argue that plausibility is achieved inherently, given that the initial field to be perturbed already honors (1) the available measurements of the hydraulic property and (2) the variogram describing the spatial variability patterns as observed in the field. Similar arguments are used by other researchers for rejecting the plausibility term [3,19,28,29,43]. Indeed, once the pilot point values have been estimated, the interpolation in step 2 of Fig. 1 will be consistent with measurements. However, nothing assures that the estimation is plausible at the pilot points themselves. Addressing such inconsistency is one of the motivations of this work.

On the other hand, regularization has been used by Doherty [16], who penalizes non-homogeneity of the interpolated field rather than including the prior estimates of the model parameters. Kowalsky et al. [26] include a plausibility term for the first time in the context of pilot points. These authors seek for an identification of the permeability in an unsaturated flow synthetic example, conditioned to hydrogeological data (i.e., saturation profiles at boreholes and permeability measurements) and geophysical measurements (ground penetrating radar, GPR) in a maximum a posteriori (MAP) geostatistical context. Although they include the plausibility term in the objective function, its role is not explored and its weighting is unclear. In addition, they do not introduce correlation of model parameters in the estimation process (i.e., diagonal covariance matrix). In short, a methodology for proper accounting the plausibility of pilot point values is still lacking.

The objective of this work is to present such a methodology and to show that the use of a plausibility term improves (1) the identification of heterogeneity and (2) the stability of the problem. The latter allows the modeler to use an increased number of pilot points, thus sharpening the resolution of heterogeneity. For these purposes, the method of pilot points was implemented in the code TRANSIN [32], that originally used the zonation approach within a maximum likelihood statistical framework.

This paper is organized as follows. First, the methodology is outlined. Second, the synthetic example and the results are explored. The paper ends with a discussion of the results and some conclusions about the use of the plausibility term.

2. Methodology

The proposed method is a modification of the pilot points method. Modifications include the use of a plausibility term and the way the vector of model parameters (value of the hydraulic properties at the pilot point locations) is updated through the optimization process. It can be summarized as follows (Fig. 1):

- *Step 1:* Analysis of measurements (hydrogeological, geophysical, etc.) and definition of the geostatistical model. In the example discussed here, the geostatistical model is defined by the variogram and the measurement error covariances, but more sophisticated models may be used. Some of the statistical parameters (e.g. variances of measurement errors) may remain uncertain.
- *Step 2:* Parameterization. A hydraulic property f (e.g. log-transmissivity) is expressed as the superposition of two fields: a drift $f_{\mathrm{D}}(\mathbf{x},t)$ and an uncertain residual $f_{\mathrm{p}}(\mathbf{x})$, which is a linear combination of the model parameters p_j:

$$f(\mathbf{x},t) = f_{\mathrm{D}}(\mathbf{x},t) + f_{\mathrm{p}}(\mathbf{x}) \tag{1}$$

- *Step 2.1:* Calculation of $f_{\mathrm{D}}(\mathbf{x})$. The drift can be obtained through conditional estimation (kriging/cokriging) or conditional simulation, depending on whether the modeler is seeking the characterization of large scale patterns or small scale variability, respectively. Therefore, it honors hard data (e.g. measurements of the hydraulic property $\mathbf{f}^*$) and possibly soft data (i.e., geophysical data $\mathbf{g}^*$ can be considered as external drifts). In the case of conditional simulation, $f_{\mathrm{D}}(x)$ reproduces spatial variability patterns as observed in the field (e.g. it honors the variogram as well). For the case of linear estimation, it can be expressed as

$$f_{\mathrm{D}}(\mathbf{x},t) = \sum_{i=1}^{\dim Z} \lambda_i^Z(\mathbf{x}) Z(\mathbf{x}_i,t) \tag{2}$$

where $\mathbf{x}$ is the location where f_{D} is calculated, t and $\mathbf{x}_i$ are the measurement times and locations, respectively, and λ_i^z are the (co-)kriging weights for the measurements, organized in the vector $\mathbf{Z} = (\mathbf{f}^*, \mathbf{g}^*)$. Our implementation of the methodology allows a large set of conditional estimation methods: simple kriging, residual kriging, kriging with locally varying mean, kriging with external drift, simple cokriging, ordinary cokriging and ordinary cokriging standardized to the mean value of the primary variable. In addition to these methods for conditional estimation, a sequential simulation algorithm for conditional simulations was implemented.

- *Step 2.2:* Parameterization of the uncertain residual $f_{\mathrm{p}}(\mathbf{x})$. It can be viewed as the perturbation of $f_{\mathrm{D}}(x)$ required to honor measurements of dependent variables (heads, concentrations, etc.). It is expressed as a linear combination of model parameters (value of the hydraulic property at the pilot point locations):

$$f_{\mathrm{p}}(\mathbf{x}) = \sum_{j=1}^{N_{\mathrm{p}}} \lambda_j^{pp}(\mathbf{x}) p_j \tag{3}$$

where N_{p} is the number of pilot points used to parameterize f_{p} (this number may be different for other hydraulic property) and $\lambda_j^{pp}(\mathbf{x})$ are the (co-)kriging weights for the model parameters p_j. These weights are calculated in the same way as λ_i^z for measurements. In fact, λ_i^z and λ_j^{pp} need to be calculated jointly. In our implementation, the location of the pilot points can be fixed or vary randomly as the optimization process proceeds.

- *Step 3:* Calculation of prior estimates of the pilot point values $\mathbf{p}^*$ and corresponding a priori error covariance matrix $\mathbf{V_p}$, by conditional estimation to measurements in vector $\mathbf{Z}$. Notice that correlation is included during the estimation process. As a result, the variance of pilot points located close to measurement points will be small. Moreover, pilot point values should be close (i.e., highly correlated) when pilot point locations are close. Therefore, $\mathbf{V_p}$ is a full matrix, as opposed to diagonal.
- *Step 4:* Objective function. Following Medina and Carrera [33], the optimum set of model parameters minimizes the objective function:

$$F = \sum_{i=1}^{\text{nstat}} \beta_i \left(\mathbf{u}_i - \mathbf{u}_i^*\right)^t \mathbf{V}_{\mathbf{u}_i}^{-1} \left(\mathbf{u}_i - \mathbf{u}_i^*\right) + \sum_{j=1}^{\text{ntypar}} \mu_j \left(\mathbf{p}_j - \mathbf{p}_j^*\right)^t \mathbf{V}_{\mathbf{p}_j}^{-1} \left(\mathbf{p}_j - \mathbf{p}_j^*\right) \tag{4}$$

where "nstat" denotes number of state variables $\mathbf{u}_i$ with available measurements $\mathbf{u}_i^*$ and covariance matrix $\mathbf{V}_{\mathbf{u}_i}$ ($i = 1$ for heads/drawdowns, $i = 2$ for concentrations, $i = 3$ for fluxes, etc.); "ntypar" is the number of types of model parameters being optimized, with prior information $\mathbf{p}_j^*$ and covariance matrix $\mathbf{V}_{\mathbf{p}_j}$ ($j = 1$ for pilot points linked to transmissivities, $j = 2$ for storativities, etc.) β_i and μ_j are weighting scalars correcting errors in the specification of $\mathbf{V}_{\mathbf{u}_i}$ and $\mathbf{V}_{\mathbf{p}_j}$. In this work, we used only drawdown data as state variable for identifying the heterogeneity of transmissivity (however, the methodology is general and can be applied to the estimation of different hydraulic properties). Thus, we will term hereinafter F_d the term of drawdowns ($\mathbf{s}$ hereinafter) and F_p the one of model parameters linked to transmissivities, being the simplified objective function [33]:

$$F = F_\mathrm{d} + \mu F_\mathrm{p} = (\mathbf{s} - \mathbf{s}^*)^t \mathbf{V}_\mathbf{s}^{-1} (\mathbf{s} - \mathbf{s}^*) + \mu (\mathbf{p} - \mathbf{p}^*)^t \mathbf{V}_\mathbf{p}^{-1} (\mathbf{p} - \mathbf{p}^*) \tag{5}$$

The objective function stated in Eq. (4) (or its particularization in Eq. (5)) can be based on favoring the best match (F_d) and ensuring plausibility and stability (F_p). However, it can also be derived in a statistical framework. Gavalas et al. [18] derived it by maximizing the posterior pdf of the model parameters, MAP, while Carrera and Neuman [4] arrived to it by maximizing the likelihood of the parameters given the data (maximum likelihood estimation, MLE). Here, we use the formulation of Medina and Carrera [33], who prefer working with the expected value of the likelihood function, as this allows the most stable estimation of statistical parameters, i.e., β_i and μ_j.
- *Step 5:* Minimization. The minimization of Eq. (4) is performed by means of Levenberg–Marquardt's method. This method belongs to the Gauss–Newton family and it consists of linearizing the dependence of state variables on model parameters, while imposing that the parameter change $\Delta\mathbf{p}^k$ at the k-th iteration is constrained. This leads to a linear system of equations [11,30,35]:

$$(\mathbf{H}^k + \delta^k \mathbf{I}) \Delta\mathbf{p}^k = -\mathbf{g}^k \tag{6}$$

where $\mathbf{H}^k$ is an approximation of the Hessian matrix of F (Eq. (4)) and $\mathbf{g}^k$ its gradient at $\mathbf{p}^k$ (vector of model parameters at iteration k), $\mathbf{I}$ is the identity matrix and δ^k is a positive scalar (Marquardt's parameter).
- *Step 6:* Updating the vector of model parameters. After each iteration, the vector of model parameters is updated as

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \Delta\mathbf{p}^k \tag{7}$$

Prior to updating, the components of vector $\Delta\mathbf{p}^k$ are examined. If any of them is larger than a given threshold, all of them are reduced accordingly. Thus, an upper bound (per iteration) limits the maximum step size.

Steps 5 and 6 are repeated until one the of the following conditions is met [32]: (a) the maximum increment of parameters (per iteration) is very small, (b) the change in the objective function between two consecutive iterations is negligible, (c) the gradient norm is very small or (d) the ratio between the gradient norm and its value at the first iteration is small enough. The algorithm also stops if the number of iterations or failed iterations (those increasing the objective function) reach threshold values. In our experience, (d) is possibly the best check of convergence and, in this work, a reduction factor of 10^{-6} of the norm of the gradient was adopted as indicator of convergence (this condition was achieved in most of the cases presented in the next section).

To verify uniqueness, it is advisable to repeat the estimation starting from different initial values for model parameters. Starting from the drift (zero values to model parameters) is a good strategy. Starting from large values for pilot point perturbations usually leads to convergence. On the contrary, starting from too low values often leads to poor convergence.
- *Step 7:* A posteriori statistical analysis. The optimization process is repeated using different values of the weighting scalars β_i and μ_j, whose optimum values are the ones leading to the maximum of the expected likelihood, equivalent to the minimum of the support function [33]:

$$S_2 = N + \ln|\mathbf{H}| + N \ln\left(\frac{F}{N}\right) - \sum_{i=2}^{\text{nstat}} n_i \ln \beta_i - \sum_{j=1}^{\text{ntypar}} k_j \ln \mu_j \tag{8}$$

Here, N is the total number of data, n_i and k_j are the number of state variable i and parameter type j data, respectively and $\mathbf{H}$ is the first order approximation of the Hessian matrix of the objective function at the end of the optimization process.

3. Application

The objective of this example is to test, first, the role of the plausibility term in the improvement of the identification of heterogeneity, and second, its sensitivity to the number of pilot points. Results are explored on the basis of a synthetic example consisting of the simultaneous interpretation of three pumping tests in a square domain of $400 \times 400\ \mathrm{m}^2$. In essence, the procedure follows the steps of Meier et al. [34].

The flow domain is enlarged to avoid spurious boundary effects to a squared global domain of $3800 \times 3800\ \mathrm{m}^2$ (Fig. 2). Two different finite element discretizations apply, being more refined the central part (zone of interest).

The "true" log transmissivity field ($\log_{10} T$ hereinafter; Fig. 2a) was generated with code TRANSIN [32] by sequential simulation conditional to a set of measurements defining two channels of high transmissivity. The "true" variogram is spherical, with a range of 200 m and a variance of 2, without nugget effect. Values of the "true" $\log_{10} T$ field range from -9.1 to 0.5, with a mean value of $-4\,[\log_{10}(\mathrm{m}^2/\mathrm{s})]$. In this work, only heterogeneity of the $\log_{10} T$ field was explored. Storativity was assumed to be constant and known over the whole domain, with a value of 10^{-4}.

Thirteen measurements of $\log_{10} T$ were selected from the "true" field as conditioning data. These measurements were purposefully located in such a way that the initial drift of Eq. (2) (calculated by ordinary kriging; Fig. 2b) was radically different from the "true" field (Fig. 2a). Notice that, indeed, the high $\log_{10} T$ channels crossing the zone of interest are missed by the drift. Thus, the performance of the model is heavily dependent on the calibration of the perturbation field f_{p}. We chose this setup to ensure that the plausibility term, which biases the estimation towards the drift, would hinder finding a good solution.

Drawdown data comes from three independent pumping tests (but analyzed simultaneously) in the most productive wells of the central domain (pumping rates of $10^{-2}\ \mathrm{m}^3/\mathrm{s}$ at wells B1, B2 and B3 in Fig. 2). Transient drawdowns were simulated at grid nodes (Fig. 3), assuming a zero drawdown as initial condition and prescribed at the boundaries. Drawdown measurements were calculated at the 13 points where $\log_{10} T$ measurements are available (a total of 936 drawdown data). A Gaussian white noise was added to those measurements, simulating acquisition errors, with

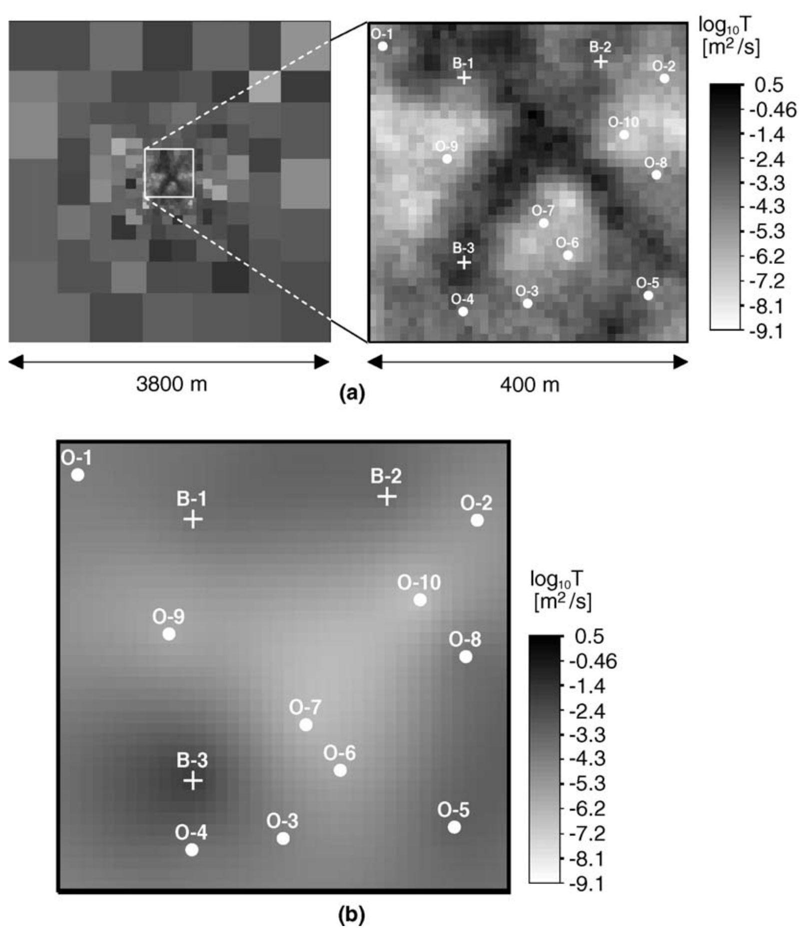

Fig. 2. Test problem description. (a) Flow domain, "true" $\log_{10} T$ field and location of conditioning measurements. All boundaries have a prescribed drawdown condition (zero). White square limits the zone of interest. Pumping tests are performed independently at points B-1, B-2 and B-3. (b) Kriging of the 13 $\log_{10} T$ measurements (circles denote observation wells, while crosses mark pumping wells).

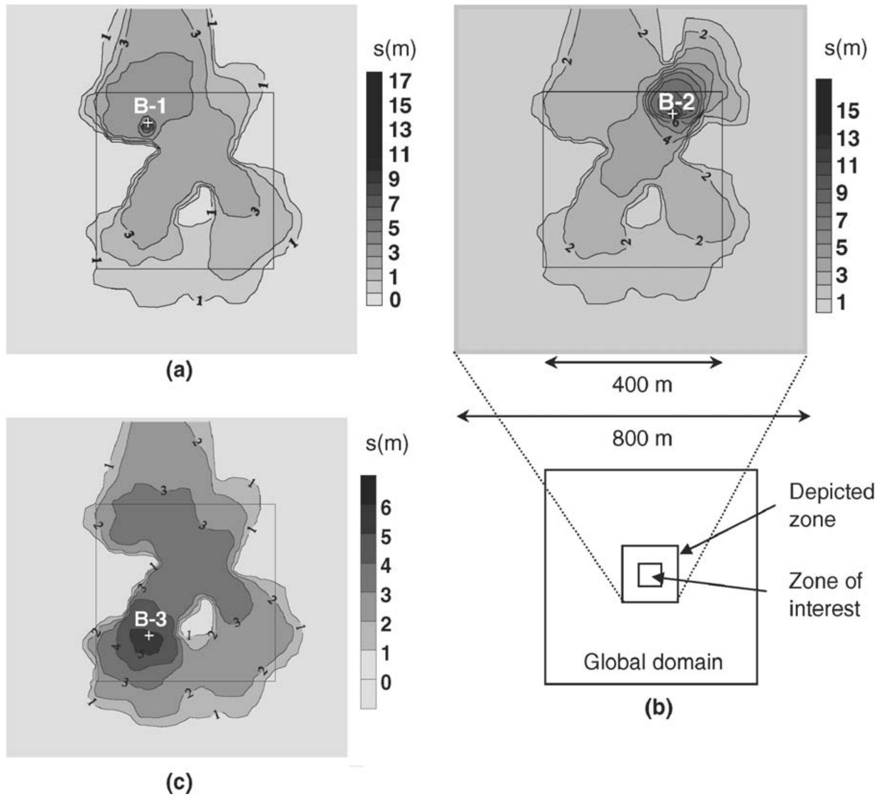

Fig. 3. "True" drawdowns after pumping ($t = 7200$ s) at wells B-1 (a), B-2 (b) and B-3 (c). The zone of interest (central square of 400×400 m^2) has been enlarged 200 m each side.

a standard deviation of 0.3 m for pumping at wells B1 and B2 and 0.15 m at well B3 (1% of the maximum drawdown at each one of the tests).

A total of 28 cases were solved, varying the weighting factor μ (Eq. (5)) and the number of pilot points employed. To explore the role of the plausibility term, seven values of μ were tested, ranging from 10^{-3} to 10^2 (lower values were not considered due to convergence problems). This range of values was selected by taking into account that the optimum value of μ should be one if the variogram is error-free ($\log_{10} T$ variogram used in the calibrations was the "true" one). High values of μ give too much weight to the plausibility term. This should result in a poor identification of heterogeneity, as the estimation would be biased towards the kriged field (Fig. 2). On the contrary, small values of μ tend to disregard the plausibility term, thus risking instability.

Regarding the number and location of pilot points, four regular networks were tested, containing 41, 65, 97 and 241 pilot points (Fig. 5, column 1). Sixteen of them are located in the outer part of the domain (coarse discretization in Fig. 2). The remaining ones (i.e., 25, 49, 81 and 225, respectively) fall within the zone of interest, corresponding to 2.5, 3.5, 4.5 and 7.5 pilot points per correlation range in each direction. Notice that only the coarsest network, containing 41 pilot points, acknowledges the "rule of thumb" of using 2–3 pilot points per correlation range [3,23]. Observe that the number of $\log_{10} T$ measurement locations does not constrain the number of pilot points (13 $\log_{10} T$ measurement locations versus a minimum of 41 pilot points), due to the inclusion of the plausibility term.

Additionally, we explored the sensitivity of the convergence rate to the threshold value limiting the maximum variation of model parameters after each iteration of the Levenberg–Marquardt's method. Values of 0.1, 1 and 2 orders of magnitude of variation were tested.

4. Results

The performance of the method was evaluated both qualitatively ($\log_{10} T$ maps and drawdown fits) and quantitatively. For the latter, an error vector **e** is defined as the difference between calculated and "true" values of $\log_{10} T$ at the zone of interest (1600 blocks of 10×10 m^2). We analyzed the following statistics:

(1) Total objective function and its drawdowns and parameters components (F, F_d and F_p in Eq. (5), respectively). These are not good comparison criteria as they grow (F_d and F) or decrease (F_p) monotonically with μ.
(2) Support function of the expected likelihood (Eq. (8)), whose minimization should lead to the optimum value of μ.
(3) Mean absolute error: measures the match between calculated and "true" values of $\log_{10} T$.

Table 1
Summary of results of the sensitivity analysis to the weighting factor μ and to the number of pilot points

Test problem		Objective function (Eq. (5))				Estimation errors	
N_p, number of pilot points	Weighting factor μ	Total objective function (F)	Drawdown objective function (F_d)	Parameter objective function (F_p)	S_2 (Eq. (8))	$\bar{e}_{\log_{10}T}$ (Eq. (9))	$\mathrm{RMSE}_{\log_{10}T}$ (Eq. (10))
–	$\mu \to \infty$	1.156×10^6	1.156×10^6	–	–	1.390	1.831
41	10^{-3}	**1161**	**1158**	3254	1399	3.534	5.146
	10^{-2}	1192	1166	2663	**1358**	3.416	4.988
	10^{-1}	1390	1196	1945	1455	2.913	4.241
	3×10^{-1}	1708	1287	1402	1516	2.492	3.887
	10^0	3638	2377	1261	2345	2.629	3.878
	10^1	6571	3496	308	2885	1.853	2.632
	10^2	34,050	16,160	**179**	4462	**1.848**	**2.456**
65	10^{-3}	**756**	**754**	1753	1109	1.704	2.705
	10^{-2}	768	758	1049	1027	1.343	1.983
	10^{-1}	826	778	480	**1014**	1.157	1.685
	3×10^{-1}	905	805	453	1068	0.999	1.389
	10^0	1105	844	261	1237	**0.992**	**1.363**
	10^1	3953	1682	227	2450	1.302	1.842
	10^2	19,141	6091	**131**	3995	1.505	2.062
97	10^{-3}	**741**	**737**	3690	1214	2.016	2.938
	10^{-2}	759	744	1501	1075	1.431	2.080
	10^{-1}	787	753	348	**1007**	0.961	**1.331**
	3×10^{-1}	875	784	302	1074	**0.950**	1.386
	10^0	1033	829	203	1205	1.025	1.456
	10^1	3070	1318	175	2267	1.408	2.001
	10^2	17,426	5566	**119**	4018	1.525	2.081
241	10^{-3}	**726**	**723**	2681	1321	1.749	2.503
	10^{-2}	736	727	851	1188	1.151	1.577
	10^{-1}	771	744	273	**1116**	**0.771**	**1.034**
	3×10^{-1}	816	760	188	1131	0.852	1.142
	10^0	922	787	135	1227	0.809	1.090
	10^1	2835	1185	108	2479	1.194	1.758
	10^2	15,358	4598	**105**	4423	1.525	2.138

Minimum values for each set are written in bold characters.

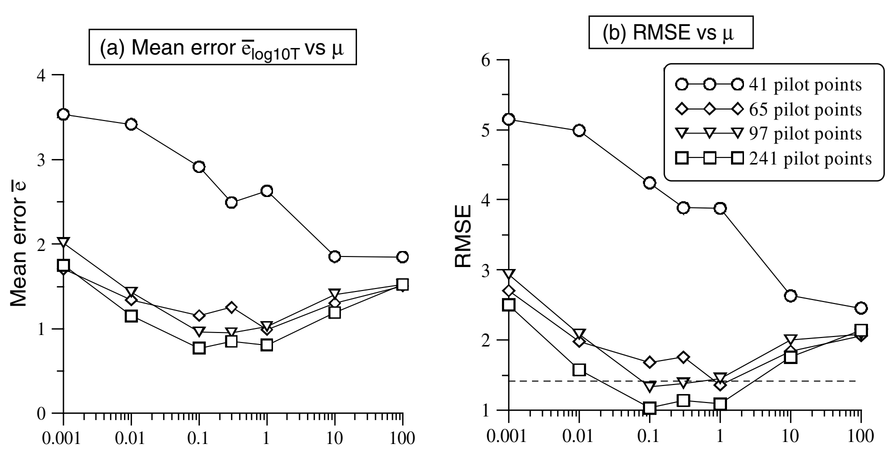

Fig. 4. Log-transmissivity estimation errors versus μ, weighting factor of the plausibility term: (a) Mean error $\bar{e}_{\log_{10}T}$, (b) Root mean square error $\mathrm{RMSE}_{\log_{10}T}$ (dashed horizontal line displays theoretical threshold value of $\sqrt{2}$).

$$\bar{e}_{\log_{10}T} = \frac{1}{1600}\sum_{i=1}^{1600}|e_i| = \frac{1}{1600}\sum_{i=1}^{1600}|\log_{10}T^{\mathrm{calc}} - \log_{10}T^{\mathrm{true}}| \quad (9)$$

We used this criterion rather than the raw one measuring the estimation bias (identical but without absolute value), given that the latter, also evaluated, was close to zero in most cases, as expected. Therefore, it did not shed new light to this research.

(4) Root mean square error of $\log_{10} T$: this is the basic raw criterion to evaluate the goodness of the identification. Theoretically, it should be smaller than the a priori deviation (square root of the variogram sill, $\sqrt{2}$ in this case), if conditioning is good. The analogous magnitude for drawdowns, $\mathrm{RMSE_d}$, was also calculated.

$$\mathrm{RMSE}_{\log_{10}T} = \left(\frac{1}{1600}\mathbf{e}^t\mathbf{e}\right)^{1/2} \tag{10}$$

Table 1 summarizes the results concerning the identification of heterogeneity. Figs. 4a and b display the quantitative comparisons in terms of the estimation errors, $\bar{e}_{\log_{10}T}$ and $\mathrm{RMSE}_{\log_{10}T}$. Qualitative comparisons of $\log_{10} T$ estimates

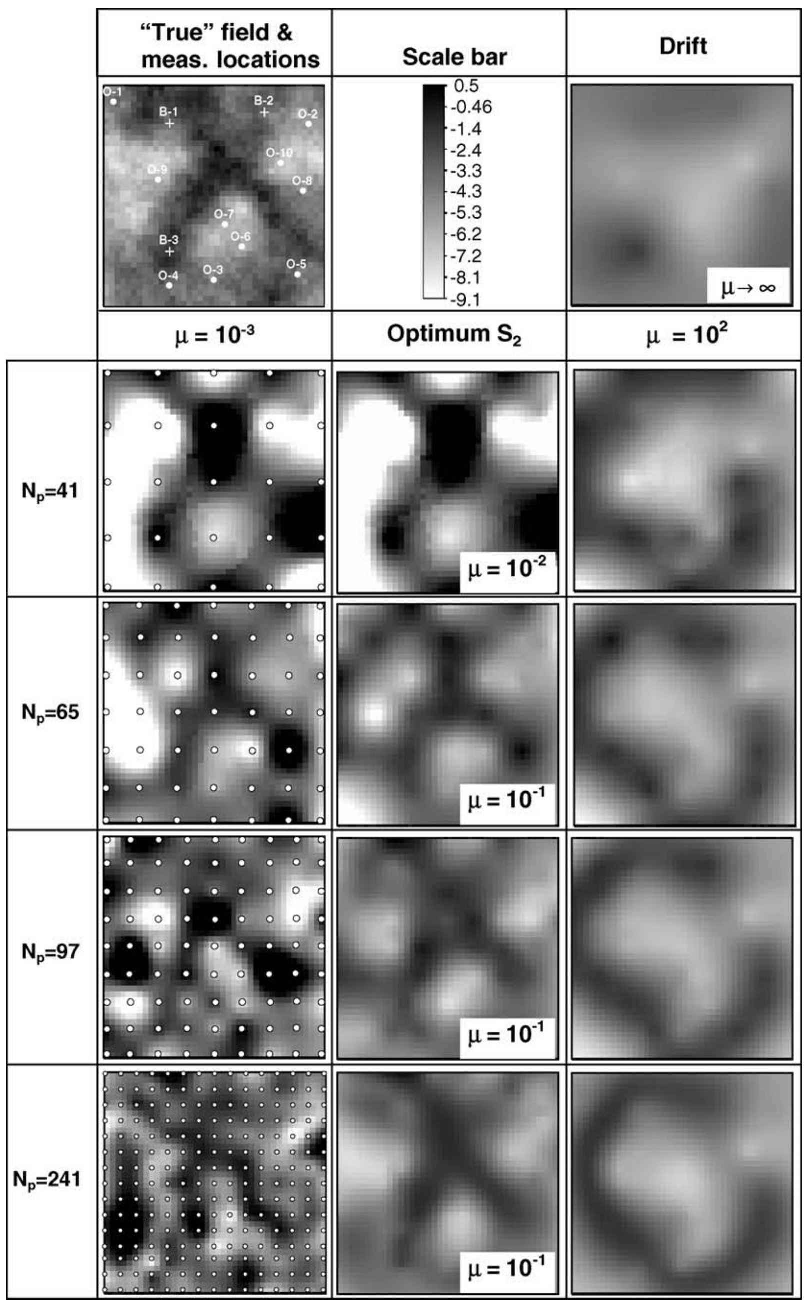

Fig. 5. Qualitative comparison of results: Row 1. "True" $\log_{10} T$ field and measurements, common scale bar and the drift to be perturbed (obtained by ordinary kriging of the $\log_{10} T$ measurements). Rows 2–5 display log-transmissivities obtained after conditioning to $\log_{10} T$ and drawdown measurements with varying number (N_p) of pilot points (circles in column 1) and weighting factor μ. Results look unstable when little weight (10^{-3}) is assigned to prior information (consistently worst estimation errors; column 1). They look too smooth when too much weight (10^2) is assigned (column 3). They resemble the "true" field when both optimum weight and a large number of pilot points are used (optimum identification as measured by criterion S_2 is displayed in column 2; in the insets, the corresponding values of μ).

are presented in Fig. 5. Fig. 6 displays the best matching of drawdown data (μ equals 10^{-3}).

The first observation that becomes apparent from Table 1 is the strong effect of the plausibility term. The relative importance given to this term is measured by the value of the weighting factor μ. Using small values of μ (small importance of the plausibility term, disregarding the prior estimates in the optimization process) consistently leads to the minimum value of F_{d} (best fit of drawdowns, Fig. 6) and to the worst identification of $\log_{10}T$ in all cases (Fig. 5, column 1). That is, for any given number of pilot points, largest estimation errors, as measured by $\bar{e}_{\log_{10}T}$ and $\mathrm{RMSE}_{\log_{10}T}$, are obtained for values of $\mu = 10^{-3}$. In this case, the lack of constraint in the plausibility term makes the problem somewhat unstable. Thus, estimated values at pilot point locations fluctuate wildly, leading to a "lumpy" appearance of the solution (Fig. 5, column 1). Similar appearance of estimated fields can be found in Zimmerman et al. [45] and Alcolea et al. [1]. The use of variable locations of the pilot points helps alleviating this problem [23].

Similarly, large values of the weighting factor also yield poor results. The final solution tends to be too smooth (Fig. 5, column 3), because it is biased towards the drift, which contains little information about the actual variability of the "true" field. In fact, the second largest values of the estimation errors ($\bar{e}_{\log_{10}T}$ and $\mathrm{RMSE}_{\log_{10}T}$) were obtained with a value of 10^2 for μ in three (65, 97 and 241 pilot points) out of the four sets of pilot points.

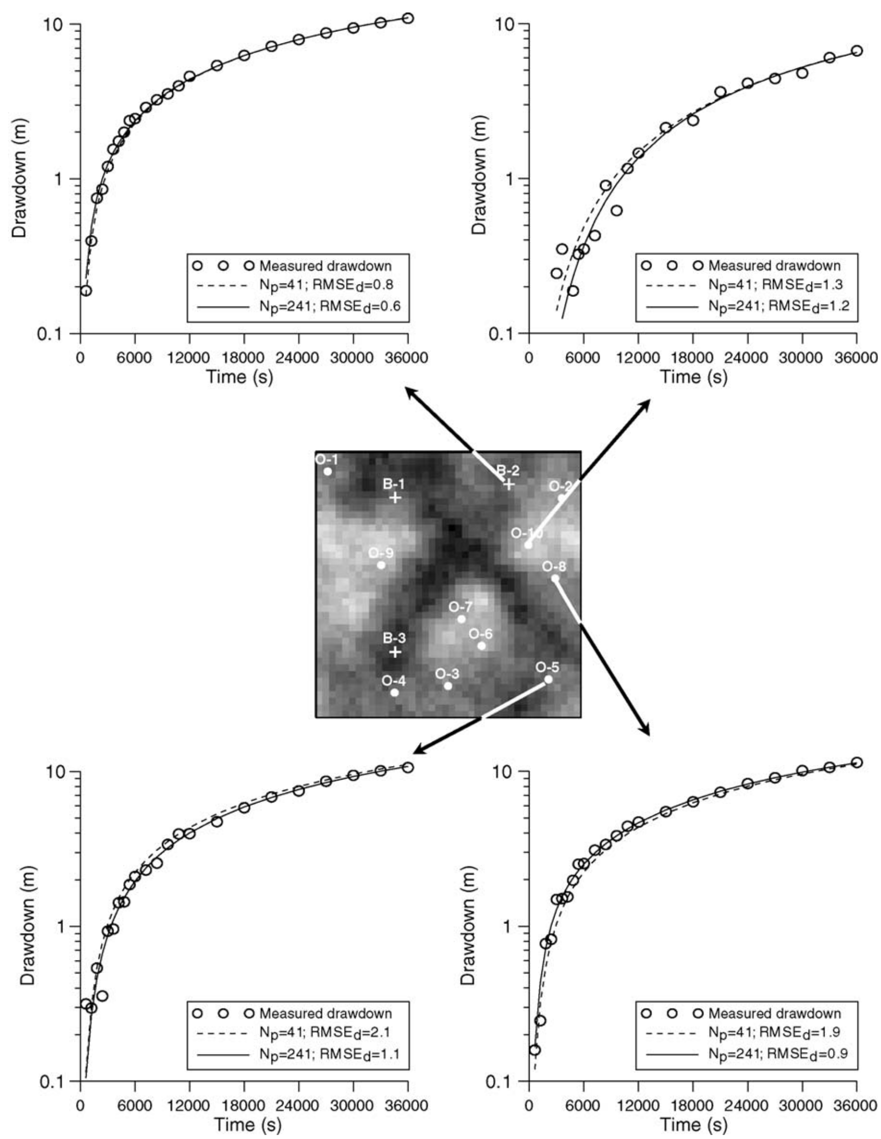

Fig. 6. Time evolution of measured (circles) and computed (lines) drawdowns in response to pumping in B-3 at selected observation points. Weighting scalar of the plausibility term is 10^{-3} (best fit of measured drawdowns). The number of pilot points and the root mean square error of drawdowns, $\mathrm{RMSE}_{\mathrm{d}}$, are presented in the insets. Notice that the fits for 41 and 241 pilot points are very similar, despite the large differences between the corresponding $\log_{10}T$ fields (Fig. 5, column 1) and the calculated $\mathrm{RMSE}_{\mathrm{d}}$. In fact, they are visually identical for all runs with drawdown objective function below 1000 (F_{d} in Table 1). The fit for 41 pilot points is not as good, but would also be considered acceptable. This implies that fitting drawdowns cannot be used as the sole criterion for the identification of transmissivities.

Optimum identifications (Fig. 5, column 2), as measured by criterion S_2 (Eq. (8)), are obtained when μ equals 10^{-1}, except when the number of pilot points is small (41 pilot points), while F_d increases minimally (i.e., the fit of drawdowns does not deviate too much from its optimum, obtained when $\mu = 10^{-3}$). We attribute this value (theoretically it should have been 1) to the procedure for designing the "true" field. This field is not typically multigaussian as assumed in this application. The fact that the method reacts by lowering μ (with respect to its theoretical optimum) suggests that the procedure is indeed robust with respect to the basic assumptions of the geostatistical model.

A disturbing finding is that, if the plausibility term is not weighted properly, the identification of the heterogeneity is even worse than the drift, calculated by conditional estimation to values of hydraulic property only (i.e., not conditioning to drawdown data). However, the use of a maximum likelihood framework allows the estimation of the weighting factor μ (step 7 in the methodology). Therefore, the use of the plausibility term is advisable.

Regarding the number of pilot points, estimation errors decrease substantially when using more than 41 pilot points. Mean error (Fig. 4a) shows that, above this number, the improvement is small. However, the estimation variance (Fig. 4b) is reduced considerably only when a large number of pilot points is used (97 and 241). In fact, the estimation variance is smaller than the a priori deviation only in these cases and near the optimum weight of the plausibility term. In addition, the identification of heterogeneity gains precision (Fig. 5, column 2) as the number of pilot points increases. A similar conclusion can be obtained (regardless of the importance of the plausibility term) concerning drawdown data matching. Even though the fits using 41 and 241 pilot points are very similar (Fig. 6), the drawdown component of the objective function (F_d, Table 1) decreases as the number of pilot points increase. Thus, the larger is the number of pilot points, the better is the match to drawdown data and the identification of heterogeneity. On the other hand, CPU time required for the calibration increases proportionally to the number of pilot points (Fig. 7).

Concerning convergence rate (i.e., number of iterations in the different cases), each of the 28 basic cases in Table 1 was repeated three times, using different threshold values to limit the size of the updating vector (values 0.1, 1 and 2 were explored, allowing modifications of the model parameters of 0.1, 1 and 2 orders of magnitude, respectively, at each iteration of the optimization process). For any basic case, both qualitative and quantitative results of the three runs were almost identical, varying only the number of iterations of the optimization process (i.e., the smaller is the threshold variation prescribed, the larger is the number of iterations needed for yielding the same solution). The number of iterations needed with a threshold value of 0.1 is about twice the one needed for a value of 2 (number of iterations was similar using values of 1 and 2). Therefore, setting too restrictive bounds in the variations of model parameters does not make sense, because the computational effort increases while the solution remains unaltered.

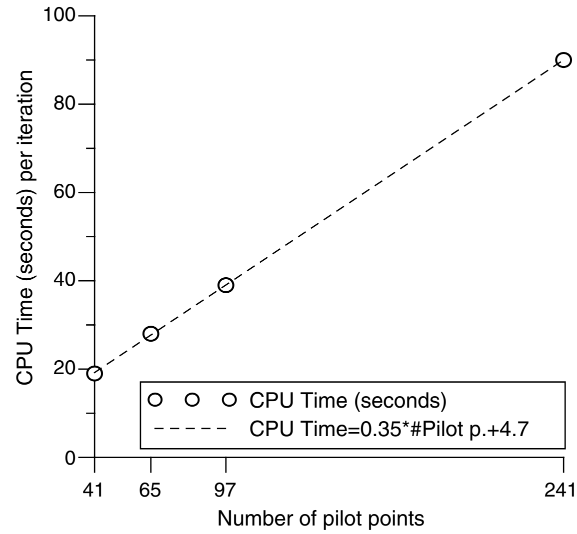

Fig. 7. CPU time (s) required for one iteration of the algorithm (steps 5 and 6). Results obtained in a Pentium IV platform.

5. Conclusions

A modification of the pilot points method has been presented, including a plausibility term in the optimization process. The suggested approach was tested on a synthetic example, exploring three items concerning the identification of heterogeneity: (1) the role of the plausibility term, (2) the sensitivity to the number of pilot points and (3) the effect of reducing the variation of the model parameters during the inversion process.

Regarding the role of the plausibility term, we have found that neglecting it, which is the standard approach in the context of pilot points, leads to the best fit of drawdown data, but to an unstable identification of the model parameters. This instability is translated in large variations of the model parameters and manifested qualitatively in a "lumpy" appearance of the estimated field. On the contrary, to give too much importance to the plausibility term biases the solution towards the drift. If the geostatistical model contains little information of the actual variability patterns (as in this case), the estimated field yields also a poor identification of the heterogeneity.

In fact, a disturbing finding is that, in most cases, conditioning to drawdown data worsens the results if the plausibility term is not weighted properly. However, the use of a statistical framework (maximum likelihood in this case) allows the estimation of the optimum weight of the plausibility term. In the synthetic example, values ranging from 0.1 to 1 (the latter was the theoretical optimum) offered the best identifications of the heterogeneity, as measured by the estimation errors, while drawdown fits were close to the optimum ones, obtained when μ is minimum. It should be noticed that good fits to measured drawdowns were obtained when neglecting (assigning very low weights to) prior information. Still, nearly as good fits were obtained with stable estimations when moderate weights were assigned to prior information.

Concerning the number of pilot points, the comparison of the estimation errors has shown that the use of a refined network with a large number of pilot points offers a precise identification of the heterogeneity and a good fit of drawdown data, while reducing the importance of a good geostatistical characterization. In short, the use of the plausibility term permits the use of a large number of pilot points, thus overcoming the risk of instabilities. In fact, one should use as large a number of pilot points as computationally feasible.

It should be stressed that the nature of the example did not favor the use of a plausibility term. First, a large number of drawdown data, coming from three different tests, was available. Therefore, one would tend to think that the problem is well posed and that little is gained by adding plausibility. Second, prior information was not very good. Only 13 measurement points were available and they missed the channels of the true field (recall Fig. 2). Therefore, one might fear that the plausibility term would bias the estimation to a wrong solution, as indeed occurred when too much weight was given to this term. The fact that the solution was still good suggests that the approach is robust. We attribute the relatively low weight assigned to prior information to the fact that reality was not really multigaussian, as assumed.

The inclusion of a reduction factor in the variation of the model parameters does not offer any improvement to the identification of the heterogeneity. Results using three values of this reduction factor yield virtually identical results in all runs. Thus, the reduction in the variation of the model parameters only adds computational effort, while the solution remains unchanged.

Finally, we stress that the prior information is a valuable data for quantifying heterogeneity, even when it is poorly informative. Thus, the use of a plausibility term including this information (usually disregarded in the context of pilot points) needs to be considered.

Acknowledgments

This work was funded by ENRESA (Spanish Agency for Nuclear Waste Disposal) and MEC (Spanish Ministry for Education and Science). Authors gratefully acknowledge J. Jódar (UPC) for his invaluable help and three anonymous reviewers for their comments.

References

[1] Alcolea A, Jodar J, Medina A, Carrera J. Geostatistical inverse problem: a modified technique for characterizing heterogeneous fields. In: Sanchez-Vila X, Carrera J, Gómez-Hernández JJ, editors. Proc GEOENV2002. Kluwer Academic Publishers; 2002. p. 175–87.

[2] Bastin G, Duque C. Modelling of steady state groundwater flow systems. In: Proc LASTED international symposium on modelling, identification and control, 1981. p. 10–6.

[3] Capilla JE, Gómez-Hernández JJ, Sahuquillo A. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 2. Demonstration on a synthetic aquifer. J Hydrol 1997;203:175–88.

[4] Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions. 1. Maximum likelihood method incorporating prior information. Water Resour Res 1986;22(2): 199–210.

[5] Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions. 2. Uniqueness, stability and solution algorithms. Water Resour Res 1986;22(2):211–27.

[6] Carrera J. State of the art of the inverse problem applied to the flow and solute transport problems. In: NATO ASI Series: groundwater flow and quality modeling, 1987. p. 549–85.

[7] Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ. Inverse problem in hydrogeology. Hydrogeol J 2005(13):206–22.

[8] Carrera J, Medina A, Galarza G. Groundwater inverse problem. Discusión on geostatistical formulations and validation. Hydrogeologie 1993(4):313–24.

[9] Carrera J, Glorioso L. On geostatistical formulations of the groundwater flow inverse problem. Adv Water Resour 1991;14(5):273–83.

[10] Certes C, de Marsily G. Application of the pilot point method to the identification of aquifer transmissivities. Adv Water Resour 1991;14(5):284–300.

[11] Cooley RL. Regression modeling of groundwater flow. USGS Open File Report 1985:85–180.

[12] Cooley RL, Hill MC. Comment on RamaRao et al. (1995) and Lavenue et al. (1995). Water Resour Res 2000;36(9):2795–7.

[13] Cooley RL. An analysis of the pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. Water Resour Res 2000;36(4):1159–63.

[14] de Marsily GH, Lavedan G, Boucher M, Fasanino G. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In: Verly G et al., editors. Geostatistics for natural resources characterization. Part 2. D. Reidel Pub Co; 1984. p. 831–49.

[15] de Marsily GH, Delhomme JP, Delay F, Buoro A. 40 years of inverse problems in hydrogeology. CR Acad Sci Ser II A Earth Planet Sci 1999;329(2):73–87.

[16] Doherty J. Groundwater model calibration using pilot points and regularization. Ground Water 2003;41(2):170–7.

[17] Emsellem Y, de Marsily GH. An automatic solution for the inverse problem. Water Resour Res 1971;7(5):1264–83.

[18] Gavalas GR, Shaw PC, Seinfield JH. Reservoir history matching by Bayesian estimation. J Soc Petrol Eng 1976;261:337–50.

[19] Gómez-Hernández JJ, Sahuquillo A, Capilla JE. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 1. Theory. J Hydrol 1997;204(1–4):162–74.

[20] Gutjahr AL, Wilson JR. Co-kriging for stochastic flow models. Transp Porous Media 1989;4(6):585–98.

[21] Hadamard J. Sur les problèmes aux dérivées partielles et leur signification physique [On the problems about partial derivatives and their physical significance]. Bull Univ Princeton 1902;13:49–52.

[22] Hadamard J. Le probleme de Cauchy et les equations aux derivees Partielles Lineaires Hyperboliques [The Cauchy's problem and the linear hyperbolic partial differential equations]. Paris: Hermann; 1932.

[23] Hendricks-Franssen HJ. Inverse stochastic modeling of groundwater flow and mass transport. Ph.D. thesis, Technical University of Valencia, Spain, 2001.

[24] Hernandez AF, Neuman SP, Guadagnini A, Carrera J. Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion. Stochast Environ Res Risk Assess 2003;17(5): 329–38.

[25] Kitanidis PK, Vomvoris EG. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one dimensional simulations. Water Resour Res 1983;19(3):677–90.

[26] Kowalsky MB, Finsterle S, Rubin Y. Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone. Adv Water Resour 2004;27:583–99.
[27] Lavenue AM, Pickens JF. Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model. Water Resour Res 1992;28(6):1543–69.
[28] Lavenue M, RamaRao BS, de Marsily GH, Marietta MG. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 2. Application. Water Resour Res 1995;31(3):495–516.
[29] Lavenue M, de Marsily G. Three-dimensional interference test interpretation in a fractured aquifer using the pilot-point inverse method. Water Resour Res 2001;37(11):2659–75.
[30] Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 1963;11:431–41.
[31] McLaughlin D, Townley LLR. A reassessment of the groundwater inverse problem. Water Resour Res 1996;32(5):1131–61.
[32] Medina A, Alcolea A, Carrera J, Castro LF. Modelos de flujo y transporte en la geosfera: Código Transin IV. [Flow and transport modelling in the geosphere: the code TRANSIN IV]. In: IV Jornadas de Investigación y Desarrollo Tecnológico de Gestión de Residuos Radiactivo de ENRESA. Technical publication 9/2000. p. 195–200.
[33] Medina A, Carrera J. Geostatistical inversion of coupled problems: dealing with computational burden and different types of data. J Hydrol 2003;281:251–64.
[34] Meier P, Medina A, Carrera J. Geostatistical inversion of Cross–Hole pumping tests for identifying preferential flow channels within a shear zone. Ground Water 2001;39(1):10–7.
[35] Nowak W, Cirpka A. A modified Levenverg–Marquardt algorithm for quasi-linear geostatistical inversing. Adv Water Resour 2004;27(7):737–50.
[36] Neuman SP. Calibration of distributed parameter groundwater flow models viewed as a multiple-objective decision process under uncertainty. Water Resour Res 1973;9(4):1006–21.
[37] RamaRao BS, Lavenue M, de Marsily GH, Marietta MG. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 1. Theory and computational experiments. Water Resour Res 1995;31(3):475–93.
[38] Rötting T, Carrera J, Bolzicco J, Salvany JM. Stream-stage response tests and their joint interpretation with pumping tests. Ground Water, in press.
[39] Rubin Y, Dagan G. Stochastic identification of transmissivity and effective recharge in steady groundwater flow: 1. Theory. Water Resour Res 1987;23(7):1185–92.
[40] Tikhonov AN. Solution of incorrectly formulated problems and the regularization method. Sov Math Dokl 1963;4:1035–8.
[41] Tikhonov AN. Regularization of incorrectly posed problems. Sov Math Dokl 1963;4:1624–7.
[42] Vesselinov VV, Neuman SP, Illman WA. Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff. 2. Equivalent parameters, high-resolution stochastic imaging and scale effects. Water Resour Res 2001;37(12):3019–41.
[43] Wen XH, Deutsch CV, Cullick AS. Construction of geostatistical aquifer models integrating dynamic flow and tracer data using inverse technique. J Hydrol 2002;255:151–68.
[44] Yeh WWG. Review of parameter estimation procedures in groundwater hydrology: The inverse problem. Water Resour Res 1986;22: 95–108.
[45] Zimmerman DA, de Marsily GH, Gotway CA, Marietta MG, Axness CL, Beauheim RL, et al. A comparison of seven geostatiscally based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resour Res 1998;34(6): 1373–413.