GEOSTATISTICAL INVERSE PROBLEM: A MODIFIED TECHNIQUE FOR CHARACTERIZING HETEROGENEOUS FIELDS

A. Alcolea[1], A. Medina[2], J. Carrera[1] and J. Jódar[1]
[1]School of Civil Engineering. UPC, Dpt. of soil engineering and geosciences. Campus Nord. Bld. D2. c/ Jordi Girona, 1-3. 08034 Barcelona, Spain; [2]School of Civil Engineering. UPC, Dpt. of Applied Mathematics III. Campus Nord. Bld. C2. c./ Jordi Girona,1-3. 08034 Barcelona, Spain

Abstract: This paper describes a modification of the self-calibrating method for generating equally likely realizations (conditional simulations) of the transmissivity field, that honour measurements of transmissivity and dependent variables (heads, concentrations, etc.). Soft data (e. g. geophysics) can also be included in the conditioning procedure as a external drift. Moreover, spatial variability patterns of the "real" field (as observed through field or lab experiments) are respected. The results of the algorithm are compared with those obtained by the most commonly used methods in groundwater, such as zonation and pilot points (conditional estimation methods). The performance of these geostatistical inverse approaches was compared on a synthetic data set, where the outcome is based on qualitative (resemblance between the obtained transmissivity fields and the 'real' one) and quantitative criteria (goodness of fit between computed and measured heads). Results show that the inclusion of head data in the conditioning procedure provides a better solution than the one obtained including only transmissivity data. Final comparison (simulations/estimations conditioned to both type of data) shows similar results. The choice of the best method depends on whether the modeller seeks small-scale variability (conditional simulation methods) or large-scale trends (conditional estimation methods).

1. INTRODUCTION

For many environmental applications, such as the selection of a waste disposal site, aquifer management or aquifer remediation, a good

characterization of the aquifer properties is absolutely necessary. The heterogeneity of some of these properties is known to control the aquifer response. For instance, it is well known that the heterogeneity of the transmissivity field has a large impact on solute or gas transport through the geosphere. The representation of aquifer behaviour is, in a wide sense, referred to as numerical modelling.

The main objective of numerical modelling is to obtain a representation of the aquifer that 1) honour all available data, such as point transmissivity, heads and concentration measurements, geological/geophysical information, etc. and 2) respect spatial variability patterns as observed through field or lab experiments. For this purpose, geostatistical inversion approaches are ideally suited, and they can be classified in two groups: conditional estimation and conditional simulation methods. While the latter provides the 'best' estimate of the unknown field, the outcome of the former is a set of equally likely realizations that honour all available data.

Several approaches can be found in each one of the groups. Zonation (*Carrera and Neuman*, 1986), kriging and pilot points method (*Certes and de Marsily,* 1991) are the most frequent among those of conditional estimation. Among others, self-calibrating method (*Gómez-Hernández et al.,* 1997), Linearized Cokriging (*Kitanidis and Vomvoris,* 1983), Linearized Semianalytical (*Rubin and Dagan,* 1987), are included in the group of conditional simulation methods.

A good review of geostatistical inverse approaches is *McLaughlin and Townley* (1996). In that paper, a common theoretical framework and a theoretical comparison are presented. However, the major attempt to compare them numerically was given by *Zimmerman et al.* (1998).

In this work we present a modification of the self-calibrating method, with especial emphasis in the algorithm, as well as a numerical comparison on a synthetic example with methods of zonation, pilot points and kriging.

2. PARAMETERIZATION METHODS

Inverse procedures need to describe the spatial and temporal variability of unknown parameters, which is referred to as parameterization. We present here a brief description of the parameterization methods used in this paper. For further information, we address the reader to reviews such as *Carrera* (1987), *Yeh* (1986). Linear parameterizations can be expressed as:

$$p(\mathbf{x},t) = \sum_{j=1}^{n} p_j f_j(\mathbf{x},t) \qquad (1)$$

where p_j are scalars called models parameters (unknowns) and $f_j(\mathbf{x},t)$ are interpolation functions. Parameterization procedures differ according to these functions. The most commonly used in groundwater have been zonation (discretization) and pilot points, defined below.

- **Zonation**: A partition is made on the system. In every partition's zone, the function $f_j(\mathbf{x},t)$ has a predefined variation or a constant value (*Carrera and Neuman*, 1986).
- **Pilot points method**: The interpolation functions $f_j(\mathbf{x},t)$ are defined as kriging coefficients and p_j are the hypothetical parameters on a finite number of points, which are referred to as pilot points (*de Marsily*, 1978).

3. SUGGESTED APPROACH

The approach proposed here is a modification of the one by *Gómez-Hernández et al.* (1997). Unknown parameter (log-transmissivity in this case) is defined as the superposition of two fields: a deterministic drift and an uncertain component. The deterministic part (Y_{drift}) can be obtained through conditional simulation or kriging, depending on whether one seeks small-scale variability or large-scale trends, and therefore reproduces hard data (i.e. transmissivity measurements) and soft data (i.e. geophysical data can be included as a external drift). The uncertain part can be seen as a perturbation, such that the final field also reproduces data related to dependent variables (heads, concentrations, etc.). To overcome stability problems, this perturbation field is expressed in terms of a finite number of unknown perturbations (ΔY) at n points (similar to master locations at *Gómez-Hernández* work, but pilot points at *de Marsily*'s). Final expression of the parameterization for $\log_{10}$T field can be expressed as:

$$\mathrm{Y}(\mathbf{x}) = \mathrm{Y}_{\mathrm{drift}}(\mathbf{x}) + \sum_{\mathrm{i=1}}^{\mathrm{n}} \lambda_{\mathrm{i}}(\mathbf{x})\Delta \mathrm{Y}_{\mathrm{i}} \qquad (2)$$

where λ_i are interpolation weights, which, in this case, are obtained through kriging (seven variants were implemented: simple kriging, ordinary kriging, kriging with locally varying mean, kriging with external drift, simple cokriging, ordinary cokriging and standardized ordinary cokriging). Given that the deterministic drift honors parameter data, we seek to determine a perturbation field such that the final field also honors data related to dependent variables (heads, concentrations, etc.). Next section describes the methodology to obtain the optimal values of the unknown perturbations at the master locations.

4. INVERSION PROCEDURE

The goal is to obtain optimal values of the perturbations such that the final field also honors dependent variable measurements. A common way to achieve it is to formulate the problem in terms of a 'performance criterion', expressing the difference between actual solution and what we know about the real system (measurements). This criterion is referred to as objective function and can be expressed as (only using head measurements):

$$J = \lambda_h \left(\mathbf{h} - \mathbf{h}^*\right)^t \underline{\mathbf{C}}_h^{-1} \left(\mathbf{h} - \mathbf{h}^*\right) + \lambda_{\Delta Y} \left(\Delta\mathbf{Y} - \Delta\mathbf{Y}^*\right)^t \underline{\mathbf{C}}_{\Delta Y}^{-1} \left(\Delta\mathbf{Y} - \Delta\mathbf{Y}^*\right) \quad (3)$$

where $\mathbf{h}^*$ is the vector of all head measurements, $\mathbf{h}$ are the corresponding computed heads, $\Delta\mathbf{Y}$ is the vector of $\log_{10}$T perturbations at master points, $\Delta\mathbf{Y}^*$ their prior estimates. $\underline{\mathbf{C}}_h$, $\underline{\mathbf{C}}_{\Delta Y}$ are the corresponding covariance matrices and λ_h, $\lambda_{\Delta Y}$ are weighting coefficients.

The set of unknown perturbations that minimizes (3) makes the final field to honor all available data. It should be noticed that conditioning is enforced strictly, given that transmissivity measurements are honored by the deterministic part Y_{drift} and perturbation is zero at those points. Posed in this way, inversion becomes an optimization problem, performed by Levenberg-Marquardt's method.

One of the novelties is the inclusion of the plausibility term, accounting for the difference between prior and posterior estimations of transmissivity at the master points. Other works (e.g. *Capilla et al.*, 1997) calibrate the model only bearing in mind head or concentration measurements, obtaining solutions providing a good fit between calculated and measured values, but do not assure plausibility of estimates. This is a very important issue. As demonstrated by *Carrera and Neuman* (1986b) the inclusion of this term (regularization term in that paper) improves the conditioning of the inverse problem.

In our work, prior estimation of the perturbations at the master points are obtained through kriging, on the basis of transmissivity measurements. This formulation also improves the statistical consistency of the method. This issue will be discussed elsewhere.

5. SYNTHETIC EXAMPLE

In this section we present the comparison between 6 geostatistical inverse approaches, including the one proposed here. All of them were applied to a set of synthetic data, where the 'real' system was perfectly known a priori.

Flow domain is a square of 4000x4000 m^2 area, where inflows are prescribed to be 0.1 m^3/d at the left boundary and the head level was set to 0 m at the right boundary. Upper and lower boundaries are supposed to be impervious. There are also two internal sinks of 3 m^3/d in the middle part of the flow domain. (Figure 1)

Log-transmissivity is considered as a random field with a zero mean and a spherical isotropic covariance function, with a variance of 4.0 and a range of 1000 m (1/4 of the domain length). For the purpose of the transmissivity estimation/simulation, the domain is divided into 1600 squared blocks of 100x100 m^2 area.

Flow regime is transient with steady-state initial conditions; under steady conditions, no pump is assumed in the middle of the domain. Wells pump only during half part of the test. The storage coefficient was taken as constant and perfectly known, with a value of 10^{-5}.

This problem setup (see Figure 1) was considered as the model for the 'real' system and was used to derive the conditioning measurements (head and $\log_{10}T$ data) at 25 observation wells.

For the application of the pilot points method and the proposed approach, a uniform grid was generated, using three master points per correlation range, as suggested by *Gómez-Hernández et al.* (1997). This leads to a total number of 144 master points, a number large enough to reproduce spatial variability patterns, but small enough considering computational effort.

Six methods (summarized at Table 1) were applied to the set of synthetic data and evaluated both qualitatively (resemblance between the obtained transmissivity and head fields and the 'real' ones) and quantitatively, in terms of the errors in computed log-transmissivities.

Table 1. Summary of methods applied to the set of synthetic data

Group	Conditioning data	Method	Acronym
Conditional simulation	$\log_{10}T$	Suggested approach	CS-T
	$\log_{10}T$, h	Suggested approach	CS-Th
Conditional estimation	$\log_{10}T$, h	Pilot points	CETh-PP
	$\log_{10}T$	Ordinary Kriging	CET-K
	$\log_{10}T$, h	Kriging as drift + perturbation	CETh-MP
	$\log_{10}T$, h	Zonation	CETh-Z

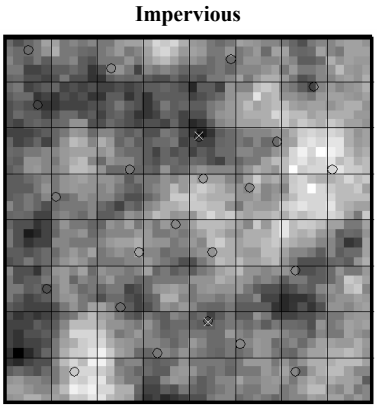


Figure 1. Synthetic example setup. "Real" transmissivity field, boundary conditions and position of the measurement points (circles).

An error vector $\mathbf{e}_j$ was defined for each simulation 'j', (unique in the case of conditional estimation):

$$\mathbf{e}_j^i = \mathbf{Y}_{calc,j}^i - \mathbf{Y}_{true,j}^i \quad i=1,N_b \quad j=1,N_S \tag{4}$$

where $\mathbf{Y}_{calc}$ and $\mathbf{Y}_{true}$ are the vector of calibrated and 'real' transmissivities of all blocks at simulation j.

Comparison is evaluated in terms of:

1. Mean error (ME): $ME = \frac{1}{N_s}\sum_{j=1}^{N_s}\frac{1}{N_b}\sum_{i=1}^{N_b}\mathbf{e}_{ji}$
2. Mean deviation error (MDE): $MDE = \frac{1}{N_s}\sum_{j=1}^{Ns}\left(\frac{1}{N_b}\mathbf{e}_j^t\mathbf{e}_j\right)^{1/2}$
3. Heads objective function: $\mathbf{J}_\mathbf{h}^\mathbf{j} = \left(\mathbf{h}^\mathbf{j} - \mathbf{h}^*\right)^t\left(\mathbf{h}^\mathbf{j} - \mathbf{h}^*\right)$

where N_s is the number of conditional simulations and N_b is the number of transmissivity blocks (50 and 1600, respectively).

The first criterion measures the estimation biases and should be close to zero. The second one measures the difference between the true field and the obtained one, and (*Carrera and Glorioso*, 1991) should be smaller than the field variance (4 in this case). The third one measures the quality of the fit between calculated and measured heads at the observation points.

5.1 Visual comparison

Consider Figure 2, displaying the results of one of the realizations obtained by the proposed method. Comparing maps at column 1, one can observe that the simulation conditioned to $\log_{10}$T data (b1) reproduces the large-scale patterns of the real field. However, there is still a large difference between the real field and the proposed one. This uncertainty can also be observed comparing maps (a2, 'real' heads) and (b2, predicted heads). Because head measurements were not included as conditioning data, measured heads do not have necessarily to be reproduced by the model, as shown at picture (b3).

This difference is reduced by adding the perturbation field (the one being calibrated on the basis of head measurements). Final solution is presented on row (c). The reduction of the uncertainty of the initial drift (conditioned only to $\log_{10}$T data) can be observed in maps and pictures at row 'c'. Final field also reproduces large-scale patterns and is more alike than the initial drift. Also, head measurements are reproduced.

Figure 3 displays the results obtained by conditional estimation methods. The most important remark is that $\log_{10}$T fields are inherently smooth. However, large-scale spatial patterns of the 'real' field are also honored, even in the cases where only $\log_{10}$ T data were used.

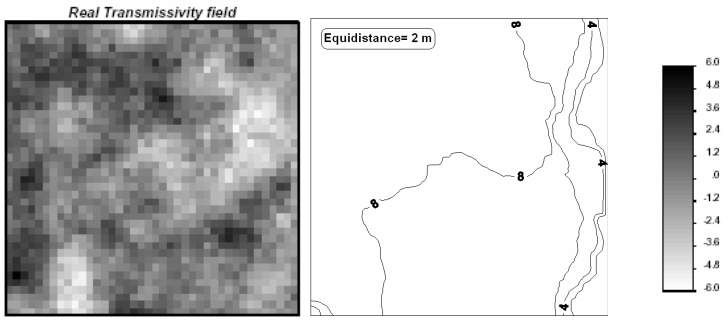
Considering rows (d) and (e) the similarity between true and calibrated fields is striking, if one seeks large-scale trends. Consider now map (c1), using pilot points method. One can see some singularities in the calibrated field, as measurements are fully respected. Row (e) displays the results obtained by the proposed method, using kriging as initial drift, jointly with the calibration of the perturbation field using the master points. This one does not present singularities on the final transmissivity map (e1), even though $\log_{10}$T measurements are also respected.

Figure 4 displays a comparison between the average field of the 50 conditional simulations and the one obtained through zonation. As one can see, they are very similar. However, the average field is still sharp, probably because only 50 realizations were considered.

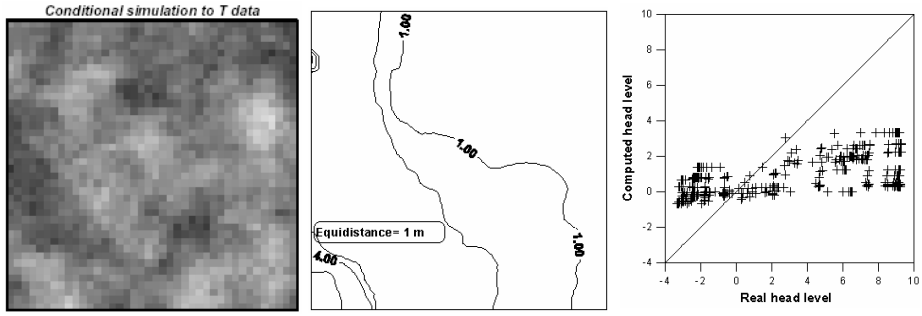
5.2 Numerical comparison

Table 2 displays the numerical aspects of the comparison. Considering mean error, all methods yielded similar results. Mean error was, in all cases, too high, but close to zero. So that, final solutions have a little bias.

(a1) Real field (a2)



(b1) CS-T case (b2) (b3)



(c1) CS-Th case (c2) (c3)

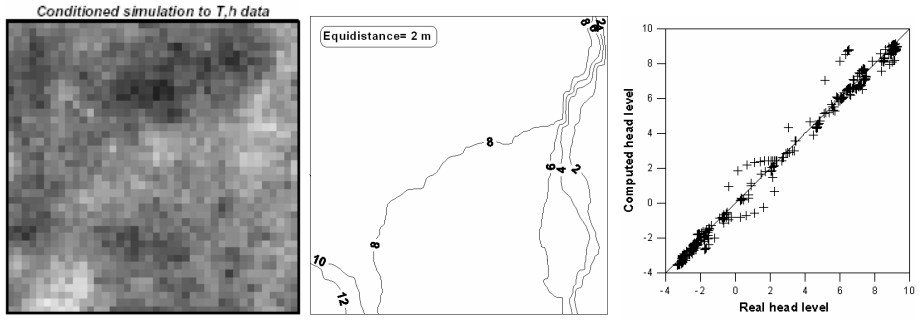
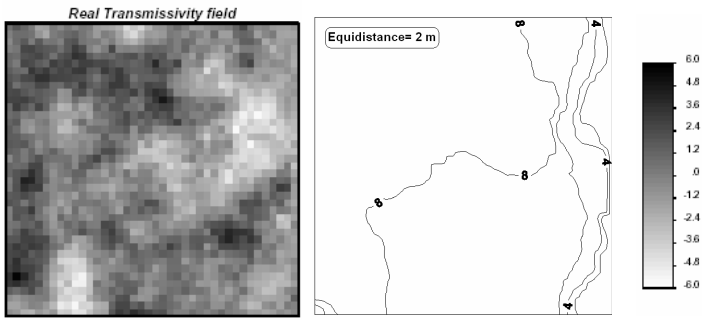
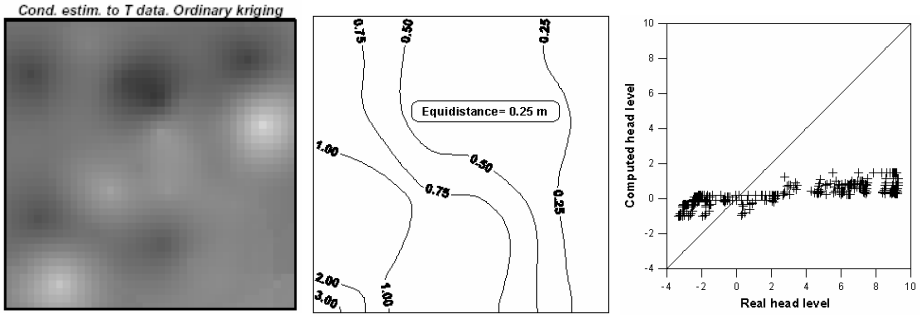


Figure 2. Results concerning conditional simulations. Row (a): 'real' $\log_{10}$T field and 'real' head level field (steady-state). Row (b): conditional simulation to $\log_{10}$T data. Row (c): Final field obtained with the proposed method, with field (b1) as initial drift. Column 1: $\log_{10}$T maps. Column 2: head level map (steady state). Column 3: Plot of computed vs. measured head level.

(a1) Real field (a2)



(b1) CET-K case (b2) (b3)



(c1) CETh-PP case (c2) (c3)

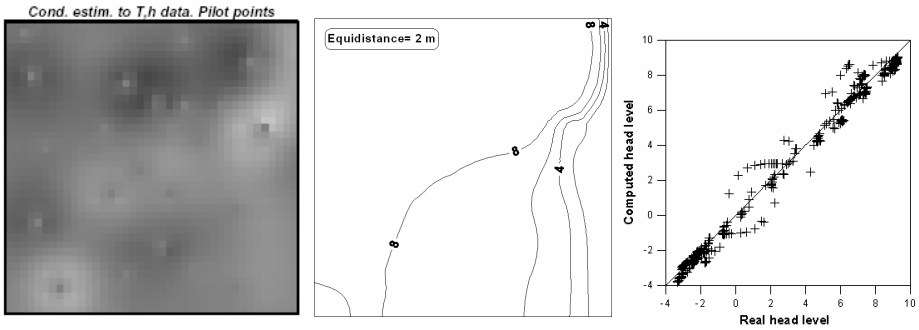


Figure 3a. Results concerning conditional estimation methods. Row (a): 'real' $\log_{10}$T field and 'real' head level field (steady-state). Row (b): conditional estimation to $\log_{10}$T data using ordinary kriging. Row (c): conditional estimation to $\log_{10}$T and head data using pilot points method. Column 1: $\log_{10}$T maps. Column 2: head level map (steady state). Column 3: Plot of computed vs. measured head level.

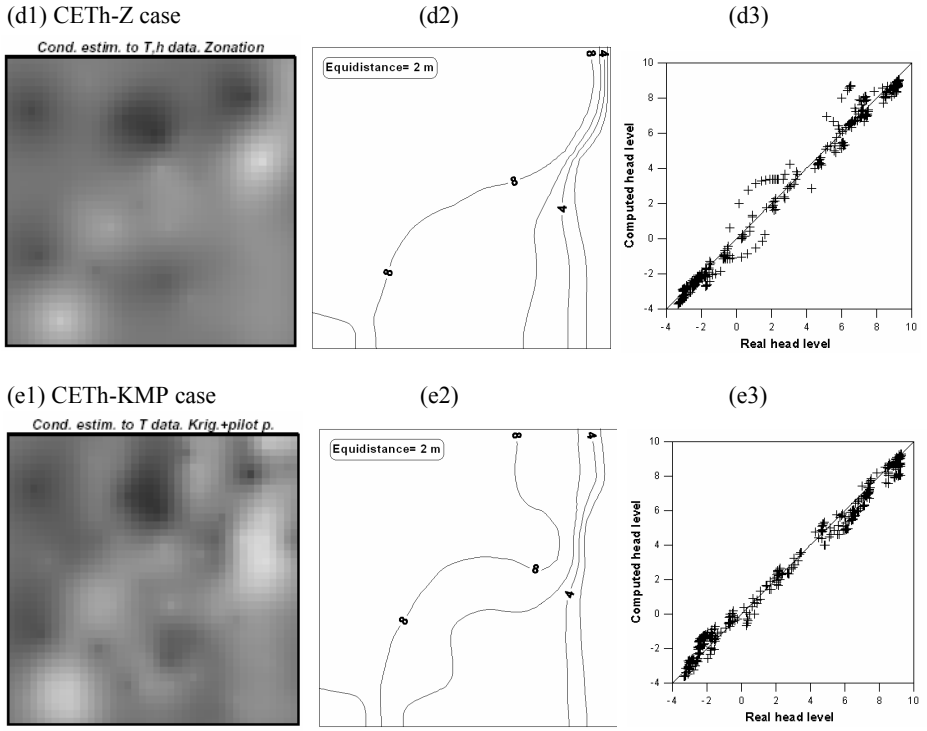


*Figure 3*b. Results concerning conditional estimation methods. Row (d): conditional estimation to $\log_{10}$T and head data using the zonation approach. Row (e): conditional estimation to $\log_{10}$T and head data using the proposed method, using a kriged field as initial drift. Column 1: $\log_{10}$T maps. Column 2: head level map (steady state). Column 3: Plot of computed vs. measured head level.

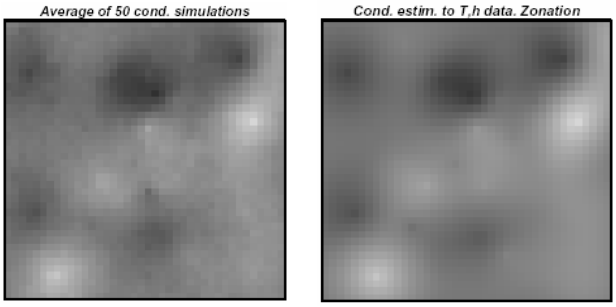


Figure 4. Comparison between the average field obtained with 50 conditional simulations to transmissivity and head level data and the $\log_{10}$T map obtained through zonation (conditional estimation to $\log_{10}$T and head data).

Considering mean deviation error, the suggested approach using kriging as initial drift displays a better behavior than the rest. All of the approaches

yielded a mean deviation error under the standard deviation of the real field, showing, in general, a good performance.

The power of the suggested approach is shown considering head level fits (as calculated by J_h). In thirteen of the fifty conditional simulations, the suggested approach performed better than the zonation method, subjectively considered as the second best. Poor results of ordinary kriging are due to the fact of considering only $\log_{10}T$ measurements as conditioning data. In general, to consider both types of measurements as conditioning data improves the quality of the final estimation.

Table 2. Numerical comparison among the methods listed at Table 1.

Method	ME	MDE	J_h
CSTh (average 50 simul.)	0.33	1.87	282
Minimum CSTh	0.23	1.72	142
Maximum CSTh	0.44	1.97	791
CETh-KPP	0.10	1.53	206
CETh-PP	0.30	1.65	193
CET-OK	0.52	1.72	18200
CETh-Z	0.34	1.61	188

6. CONCLUSIONS

A modification of the self-calibrating method for generating equally likely realizations (conditional simulations) of the transmissivity field is presented. Final solutions honor measurements of transmissivity and dependent variables (heads, concentrations, etc.). Soft data (e.g. geophysics) can also de included in the conditioning procedure as an external drift.

Transmissivity field is defined as the superposition of a deterministic drift (obtained through kriging or conditional simulation), that honours log10T measurements

and reproduces spatial variability of the field being simulated and an uncertain perturbation field. The latter is optimized such that the final field also honours dependent variables measurements (heads in this work, although other type of measurements can be included easily).

Actual modifications consists of the addition of a penalty/regularization term in the objective function, considering plausibility of the model parameters, as well as the chance of using a kriged field as initial drift.

The algorithm is compared with the most frequently used conditional estimation methods (ordinary kriging, zonation and pilot points) on a set of synthetic data. The comparison is evaluated qualitatively and numerically. Both conditional estimation and conditional simulation approaches yielded good reproductions of the real system. The choice of the most appropriate method is somewhat subjective. It depends on whether the modeler seeks

small-scale variability (conditional simulation) or large-scale trends (conditional estimation). However, single optimal estimate provided by conditional estimation should be used with caution for non-linear predictions. It is also (once more) corroborated that the inclusion of head measurements as conditioning data improves the quality (reduces the uncertainty) of the final estimation.

ACKNOWLEDGEMENTS

This work was funded by Spanish Nuclear Waste Management Company (ENRESA).

REFERENCES

1. Capilla, J. E., J. Gómez-Hernández and A. Sahuquillo, "Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data- I. Demonstration on a synthetic aquifer", *Journal of Hydrology, 203, pp 175-188*, 1997.
2. Carrera, J. and S.P. Neuman, "Estimation of aquifer parameters under transient and steady-state conditions, 1, Maximum Likelihood Method incorporating prior information", *Water Resources Research, 22 (2), pp. 199-210,* 1986a.
3. Carrera, J. and S.P. Neuman, "Estimation of aquifer parameters under transient and steady-state conditions, 2, Uniqueness, Stability and Solution Algorithms", *Water Resources Research, 22 (2), pp. 211-227,* 1986b.
4. Carrera, J., " State of the Art of the Inverse Problem Applied to the Flow and Solute Transport Equations". Groundwater Flow and Quality Modeling (Custodio et al. Eds.), Riedel, Dordrecht, pp 549-583, 1987.
5. Carrera, J. and L. Glorioso, "On geostatistical formulations of the groundwater flow inverse problem", *Advances in Water Resources, 14, pp 273, 282 (1991)*
6. Certes, C. and de G. Marsily, "Application of the Pilot Point Method to the Identification of Aquifer Transmissivities", *Advances in Water Resources, 14, pp. 284-300 (1991)*
7. de Marsily, G., "De l'identification des systèmes en hydrogeologuiques (tome 1)", *Ph.D. Thesis, L'Univ. Pierre et Marie Curie-Paris VI, Paris, pp. 58-130 (1991)*
8. Gómez-Hernández, J., A. Sahuquillo and J.E. Capilla, "Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data- I. Theory", *Journal of Hydrology, 203, pp 162-174*, 1997.
9. Kitanidis, P.K. and E.G. Vomvoris, "A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations", *Water Resources Research 19(3), pp 677-690 (1983).*
10. McLaughlin, D. and LL. R. Townley, "A reassessment of the groundwater inverse problem"*, Water Resources Research 32(5), pp 1131-1161, (1996)*
11. Yeh, W.W.G., "Review of parameter identification procedures in groundwater hydrology: The inverse problem", *Water Resources Research, 22 (2), pp. 95-108,* 1986.