

UPC

TECHNICAL UNIVERSITY OF CATALONIA. SCHOOL OF CIVIL ENGINEERING.

DEPARTMENT OF GEOTECHNICAL ENGINEERING AND GEOSCIENCES.

BARCELONA, SPAIN

-

UNIVERSITAT POLITÈCNICA DE CATALUNYA

ESCOLA TÈCNICA SUPERIOR D'ENGINYERS, CANALS I PORTS.

DEPARTAMENT D'ENGINYERIA DEL TERRENY I CARTOGRÀFICA

BARCELONA, ESPAÑA

**REGULARIZED PILOT POINTS METHOD FOR THE
CHARACTERIZATION OF HETEROGENEITY**

PHD THESIS SUBMITTED BY:

ANDRÉS ALCOLEA RODRÍGUEZ

ADVISORS:

JESÚS CARRERA RAMÍREZ

AGUSTÍN MEDINA SIERRA

BARCELONA, JUNE 2006



**HYDROGEOLOGY GROUP
TECHNICAL UNIVERSITY OF
CATALONIA**

UPC

TECHNICAL UNIVERSITY OF CATALONIA. SCHOOL OF CIVIL ENGINEERING.

DEPARTMENT OF GEOTECHNICAL ENGINEERING AND GEOSCIENCES.

BARCELONA, SPAIN

-

UNIVERSITAT POLITÈCNICA DE CATALUNYA

ESCOLA TÈCNICA SUPERIOR D'ENGINYERS, CANALS I PORTS.

DEPARTAMENT D'ENGINYERIA DEL TERRENY I CARTOGRÀFICA

BARCELONA, ESPAÑA

**REGULARIZED PILOT POINTS METHOD FOR THE
CHARACTERIZATION OF HETEROGENEITY**

PHD THESIS SUBMITTED BY:

ANDRÉS ALCOLEA RODRÍGUEZ

ADVISORS:

JESÚS CARRERA RAMÍREZ

AGUSTÍN MEDINA SIERRA

BARCELONA, JUNE 2006



**HYDROGEOLOGY GROUP
TECHNICAL UNIVERSITY OF
CATALONIA**

"All our science, measured
against reality, is primitive and
childlike and yet it is the most
precious thing we have."

Albert Einstein

To my family

ABSTRACT

Heterogeneity exerts a major control on groundwater flow and contaminant transport. Geostatistical inversion represents a powerful tool to characterize heterogeneity. The pilot points method (PPM) is arguably the most flexible and widely used among the inverse approaches. However, the PPM also suffers shortcomings. These arise from instability of the inverse problem. A traditional tactic to combat instabilities consists of adding a regularization term to the objective function. Surprisingly, this option had not been applied to the PPM in a consistent manner. This dissertation aims at filling this gap. A modification of the PPM (termed ‘regularized pilot points method’, RPPM) is presented. The main novelty consists of the addition of a plausibility term, which quantifies the departure of model parameters from their prior estimates. This term improves the identification of heterogeneity and the stability of the problem. This thesis consists of four self-contained papers.

The methodology is presented in the first paper and its performance is explored on a synthetic example. The method aims at obtaining the conditional estimation of logT from direct measurements of both logT and dependent variables (drawdowns in this case). Emphasis is placed on assessing the weighting of the plausibility term, which quantifies the importance of the prior information of parameters in the calibration. Results show that neglecting plausibility, which is the standard option in the context of PPM, leads to the best fit of dependent variables, but to an unstable identification of model parameters. On the contrary, giving too much importance to plausibility (i.e., disregarding dependent variable measurements) biases the solution towards the prior information. Thus, a proper weighting of the plausibility term is needed. This is done in the statistical framework of maximum likelihood estimation. This results in not only

statistical consistency and increased stability, but also enhanced resolution. The added stability allows the use of as many pilot points as computationally feasible (what contradicts the traditional use of the PPM).

These results are extended to the case of conditional simulation in the second paper. That is, the possibility of using the plausibility term in the case of seeking stochastic simulations of the property conditioned to direct measurements and dependent variables is explored. Results show that optimal weighting of the plausibility term is also necessary. However, to search this optimum for each conditional simulation can be tedious. A key finding of this work is that, for most simulations, the optimum value of the weighting factor of the plausibility term is the same as the one obtained by conditional estimation. This frees the modeller of the burden of having to seek the optimum weight for each simulation (usually a large number), but to obtain it just once using the RPPM in its variant of conditional estimation.

In the third paper, the RPPM is framed in the context of the universal scaling theory. The objective of this paper is to test the ability of the RPPM for reproducing the small scale variability of hydraulic conductivity. Accepting that this variability cannot be identified, the effect of the presence of small scale variability in the identification of large scale patterns of connectivity is analysed. In parallel, whether including small scale variability allows us to reproduce tailing in simulated breakthrough curves (BTCs). Results show that adding a component of small scale variability leads to increased tailing in breakthrough curves. Furthermore, the main features of BTCs (arrival time, peak concentration and slope of the tail) are reproduced. At the same time, the main patterns of connectivity are represented. This suggests that, even though the small scale variability cannot be identified accurately, it must be accounted for in meaningful transport simulations.

The motivation of the fourth paper was the hydraulic characterization of a contaminated site as a first step to the design of a remediation system. This design demands a reliable characterization of hydraulic connectivity patterns, which are best measured by hydraulic diffusivity. It can be derived using the tidal response method (TRM), which is a closed-form solution. Unfortunately, the conventional TRM assumes homogeneity. The objective of this work is to overcome this limitation and use tidal

response to identify preferential flow paths. Spatial variability is characterized using the RPPM. The procedure requires joint inversion with pumping test data to resolve diffusivity into transmissivity and storage coefficient. Actual application is complicated by the need to filter tidal effects from the response to pumping and by the need to deal with different types of data, which is addressed using maximum likelihood methods. Application to a contaminated artificial coastal fill leads to flow paths that are consistent with the materials used during construction and to solute transport predictions that compare well with observations. It is concluded that tidal responses can be used to identify connectivity patterns.

RESUMEN

La heterogeneidad controla el flujo de agua y el transporte de contaminantes en el subsuelo. La inversión geoestadística es una potente herramienta para caracterizar la heterogeneidad. De entre las metodologías de problema inverso, el método de los puntos piloto (PPM) es posiblemente el más flexible y uno de los más ampliamente utilizados. Pese a ello, el PPM algunos inconvenientes debidos a la inestabilidad del problema inverso. La inestabilidad se suele paliar añadiendo un término de regularización a la función objetivo. Sorprendentemente, esta opción no había sido contemplada de forma consistente en el PPM. Esta tesis pretende rellenar ese vacío. Se presenta una modificación del PPM (llamada método regularizado de puntos piloto, RPPM), cuya novedad consiste en la inclusión de un término de plausibilidad. Este término, que cuantifica la desviación de los parámetros del modelo con respecto a su información previa, mejora la identificación de la heterogeneidad y añade estabilidad al problema. Esta tesis contiene cuatro artículos autocontenidos.

En el primer artículo se presentan la metodología y su aplicación a un ejemplo sintético. El RPPM se utiliza para obtener la estimación condicionada de $\log T$ a partir de datos de dicha propiedad y de otras variables dependientes de ella (descensos en este caso). Se enfatiza en la ponderación del término de plausibilidad, que cuantifica la importancia de la información previa de los parámetros en la calibración. Los resultados muestran que despreciando la plausibilidad (opción habitual en el contexto del PPM) se obtienen los mejores ajustes de las variables dependientes, pero las identificaciones de los parámetros son inestables. Por contra, dar demasiada importancia a la plausibilidad (despreciando las medidas de las variables dependientes) hace que la solución tienda a la información previa. Por tanto, el término de plausibilidad debe ponderarse de forma

apropiada. Esta ponderación se lleva a cabo en el contexto geoestadístico de máxima verosimilitud, lo que confiere no sólo consistencia estadística y un incremento de la estabilidad, sino también resolución adicional. La estabilidad añadida permite usar tantos puntos piloto como un esfuerzo computacional razonable tolere, lo que contradice el uso tradicional del PPM.

Estos resultados se extienden al caso de simulación condicionada en el segundo artículo. En él se explora la posibilidad de usar el término de plausibilidad en el caso de buscar simulaciones estocásticas de la propiedad condicionadas a las medidas directas y de las variables dependientes. Los resultados muestran que también se requiere una ponderación óptima del término de plausibilidad. Hallar este óptimo para cada simulación puede resultar tedioso. Sin embargo, un hallazgo clave de este trabajo es que, para la mayoría de las simulaciones, el valor óptimo del factor de ponderación del término de plausibilidad es el mismo que se obtiene en el caso de estimación condicionada. Esto libera al usuario del tedio de buscar el peso óptimo para cada simulación (generalmente un gran número) y obtenerlo una sola vez usando el RPPM en su variante de estimación condicionada.

En el tercer artículo se enmarca el RPPM en el contexto de la teoría de escalado universal. El objetivo de este artículo es probar la habilidad del RPPM para reproducir la variabilidad a escala pequeña de la conductividad hidráulica. Aceptando que dicha variabilidad no puede identificarse, se analiza su efecto en la caracterización de los patrones de conectividad a escala grande. En paralelo, se investiga si la inclusión de la variabilidad a escala pequeña permite reproducir las colas de las curvas de llegada (efecto de ‘tailing’). Los resultados muestran que añadiendo una componente de variabilidad a escala pequeña se consigue un mayor efecto de ‘tailing’ en las curvas de llegada. Además, se reproducen las características principales de éstas (tiempo de llegada, concentración de pico y pendiente de la cola). Al mismo tiempo se consigue reproducir los patrones principales de conectividad. Esto sugiere que, pese a que la variabilidad a escala pequeña no se puede identificar con precisión, debe tenerse en cuenta en las simulaciones de transporte.

La motivación del cuarto artículo es obtener la caracterización hidráulica de un acuífero contaminado, como paso previo al diseño de un sistema de remediación. Dicho

diseño requiere una caracterización fiable de los patrones de conectividad hidráulica. La mejor medida para identificarlos es la difusividad hidráulica, que puede obtenerse por aplicación del método de respuesta a la marea ('tidal response method', TRM). Desafortunadamente, el TRM convencional asume homogeneidad. El objetivo de este artículo es salvar esa limitación y usar la respuesta a las mareas para identificar caminos preferentes de flujo, usando el RPPM para caracterizar la variabilidad espacial. El procedimiento requiere la inversión conjunta de la respuesta a las mareas y de datos de ensayos de bombeo para obtener estimaciones separadas de transmisividad y coeficiente de almacenamiento. La aplicación a un acuífero real resulta complicada por la necesidad de filtrar el efecto de la marea del efecto del bombeo y por la necesidad de manejar distintos tipos de datos, para lo que se utiliza el método de máxima verosimilitud. La aplicación a un relleno artificial contaminado cercano a la costa identifica caminos preferentes de flujo consistentes con los materiales usados durante la construcción y a predicciones de transporte comparables con los datos medidos. Se concluye que la respuesta a las mareas se puede utilizar para la identificación de caminos preferentes de flujo.

RESUM

L'heterogeneïtat controla el flux de l'aigua i el transport de contaminants en el subsòl. La inversió geoestadística és una potent eina per a caracteritzar l'heterogeneïtat. Entre les metodologies de problema invers, el mètode dels punts pilot (PPM) és possiblement el més flexible i un dels més àmpliament utilitzats. Malgrat això, el PPM presenta alguns inconvenients deguts a la inestabilitat del problema invers. La inestabilitat sol pal·liar-se afegint un terme de regularització a la funció objectiu. Sorprenentment, aquesta opció no havia estat contemplada de forma consistent en el PPM. Aquesta tesi pretén emplenar aquest buit. Es presenta una modificació del PPM (que s'anomena mètode regularitzat de punts pilot, RPPM), la novetat de la qual consisteix en la inclusió d'un terme de plausibilitat. Aquest terme, que quantifica la desviació dels paràmetres del model pel que fa a la seva informació prèvia, millora la identificació de l'heterogeneïtat i afegeix estabilitat al problema. Aquesta tesi conté quatre articles autocontinguts.

En el primer article es presenten la metodologia i la seva aplicació a un exemple sintètic. El RPPM s'utilitza per a obtenir l'estimació condicionada de $\log T$ a partir de dades d'aquesta propietat i d'altres variables dependents d'ella (descensos en aquest cas). S'emfatitza en la ponderació del terme de plausibilitat, que quantifica la importància de la informació prèvia dels paràmetres en el calibratge. Els resultats mostren que menyspreant la plausibilitat (opció habitual en el context del PPM) s'obtenen els millors ajustaments de les variables dependents, però les identificacions dels paràmetres són inestables. Per contra, donar massa importància a la plausibilitat (menyspreant les mesures de les variables dependents) fa que la solució tendeixi a la informació prèvia. Per tant, el terme de plausibilitat ha de ponderar-se de forma apropiada. Aquesta ponderació es du a terme en el context geoestadístic de màxima versemblança, el que

confereix no només consistència estadística i un increment de l'estabilitat, sinó també resolució addicional. L'estabilitat afegida permet utilitzar tants punts pilot com un esforç computacional raonable toleri, el que contradiu l'ús tradicional del PPM.

Aquests resultats extenen al cas de simulació condicionada en el segon article. En ell s'explora la possibilitat d'utilitzar el terme de plausibilitat en el cas de buscar simulacions estocàstiques de la propietat condicionades a les mesures directes i de les variables dependents. Els resultats mostren que també es requereix una ponderació òptima del terme de plausibilitat. Trobar aquest òptim per a cada simulació pot resultar tediós. No obstant això, una troballa clau d'aquest treball és que, per a la majoria de les simulacions, el valor òptim del factor de ponderació del terme de plausibilitat és el mateix que s'obté per a l'estimació condicionada. Això allibera a l'usuari del tedi de buscar el pes òptim per a cada simulació (generalment un gran nombre) i obtenir-lo una sola vegada usant el RPPM en la seva variant d'estimació condicionada.

En el tercer article s'emmarca el RPPM en el context de la teoria d'escalat universal. L'objectiu d'aquest article és provar l'habilitat del RPPM per a reproduir la variabilitat a escala petita de la conductivitat hidràulica. Acceptant que aquesta variabilitat no pot identificar-se, s'analitza el seu efecte en la caracterització dels patrons de connectivitat a escala gran. En paral·lel, s'investiga si la inclusió de la variabilitat a escala petita permet reproduir les cues de les corbes d'arribada (efecte de 'tailing'). Els resultats mostren que afegint una component de variabilitat a escala petita s'aconsegueix un major efecte de 'tailing' en les corbes d'arribada. A més, es reproduïen les característiques principals d'aquestes (temps d'arribada, concentració de pic i pendent de la cua). Al mateix temps s'aconsegueix reproduir els patrons principals de connectivitat. Això suggereix que malgrat que la variabilitat a escala petita no es pot identificar amb precisió, ha de tenir-se en compte en les simulacions de transport.

La motivació del quart article és obtenir la caracterització hidràulica d'un aquífer contaminat, com pas previ al disseny del sistema de regeneració. Aquest disseny requereix una caracterització fiable dels patrons de connectivitat hidràulica. La millor mesura per a identificar-los és la difusivitat hidràulica, que pot obtenir-se per aplicació del mètode de resposta a la marea ('tidal response method', TRM). Desafortunadament, el TRM convencional assumeix homogeneïtat. L'objectiu d'aquest article és salvar aquesta limitació i usar la resposta a les mareas per a identificar camins preferents de

flux, usant el RPPM per a caracteritzar la variabilitat espacial. El procediment requereix la inversió conjunta de la resposta a les mareas i de dades d'assajos de bombament per a obtenir estimacions separades de transmisivitat i coeficient d'emmagatzematge. L'aplicació a un aquífer real resulta complicada per la necessitat de filtrar l'efecte de la marea de l'efecte del bombament i per la necessitat de manipular diferents tipus de dades, per el que s'utilitza el mètode de màxima versemblança. L'aplicació a un terreny format per un rebliment artificial contaminat proper a la costa identifica camins preferents de flux consistents amb els materials emprats durant la construcció i a prediccions de transport comparables amb les dades mesurades. Es conclou que la resposta a les mareas es pot utilitzar per a la identificació de camins preferents de flux.

ACKNOWLEDGMENTS

This work would not have been possible without the help and the encouraging of many people. First, I want to thank Jesús Carrera for having guided this Thesis. His extraordinary knowledge of hydrogeology and his enthusiastic way of working were the decisive reason for starting the PhD program. In the personal side, he has been not only a great boss, but also the good friend that supplies the absolutely necessary moral support in bad moments. Second, I want to thank my co-advisor Agustín Medina for his time, his efforts and, specially, his friendship. This time with him has taught me the mathematician way of life: systematic, strict and tenacious, but funny and constructive at the same time.

I extend my thanks to the many hardworking colleagues at the Hydrogeology Group of the Technical University of Catalonia for their friendship, moral support and the good times we have spent together. Eduardo, Manuela, Sergio, thanks for the good moments and your hard work. Special thanks to Xavi, Enric and Marco, who suffered the stressing moments and always offered good ideas in the critical moments. Additional thanks to Tere, Silvia and Jordi for their valuable help during these years. Thanks also to my students. To teach them has been an enriching experience. Economic support of the Spanish Nuclear Waste Management Agency (ENRESA) and the Spanish Ministry for Education and Science (MEC) was greatly appreciated.

Friends of the Tuna group (“La Gloriosa”) deserve a special mention in this list (Carmen and Versalles’ friends, indeed, you are included). They shared the bad moments and offered a shoulder to lean on. M^a Ángeles, thank you for your love and support. Last, but not least, I thank my family for their love. It is hardly possible to express in a few words how much honor you deserve. Father, Mother, Sister, this work would not have been possible without your hopes, your love and your encouraging. This work is dedicated to them. Daddy, wherever you are, this work represents the dream you had and inculcated me. Now this dream is realized.

TABLE OF CONTENTS

Abstract	i
Resumen	v
Resum	vii
Acknowledgements	xiii
Table of Contents	xv
List of Figures	xix
List of Tables	xxvii
Introduction	I-1

PAPER I. PILOT POINTS METHOD INCORPORATING PRIOR INFORMATION FOR SOLVING THE GROUNDWATER FLOW INVERSE PROBLEM

1. Abstract	PI-1
2. Introduction	PI-1
3. Methodology	PI-6
3.1. Analysis of measurements	PI-6
3.2. Parameterization	PI-6
3.3. Calculation of prior estimates of the pilot point values p^* and corresponding a priori error covariance matrix V_p	PI-7
3.4. Objective function	PI-8
3.5. Minimization	PI-8
3.6. Updating the vector of model parameters	PI-9
3.7. A posteriori statistical analysis	PI-9
4. Application	PI-10
5. Results	PI-13
6. Conclusions	PI-20
Acknowledgements	PI-22

References	PI-22
------------------	-------

**INVERSION OF HETEROGENEOUS PARABOLIC-TYPE EQUATIONS
USING THE PILOT POINTS METHOD**

1. Abstract	PII-1
2. Introduction	PII-1
3. Methodology	PII-4
3.1. Parameterization	PII-5
3.2. Prior estimation	PII-6
3.3. Objective function	PII-6
3.4. Minimization	PII-7
3.5. Updating the vector of model parameters	PII-8
3.6. A posteriori statistical analysis	PII-8
4. Application	PII-10
5. Results	PII-14
6. Conclusions	PII-18
Acknowledgements	PII-22
References	PII-22

**REGULARIZED PILOT POINTS METHOD FOR REPRODUCING THE
EFFECT OF SMALL SCALE VARIABILITY. APPLICATION TO
SIMULATIONS OF CONTAMINANT TRANSPORT**

1. Abstract	PIII-1
2. Introduction	PIII-1
3. Inversion methodology	PIII-4
3.1. Parameterization	PIII-4

3.2. Optimization of model parameters.	PIII-4
3.3. Finding the optimum weighting scalars β_i and μ_j (a posteriori statistical analysis).	PIII-5
4. Procedure for representing small scale variability	PIII-5
4.1. Problem domain	PIII-6
4.2. Generation of the “true” conductivity fields	PIII-7
4.3. Generate drawdown data	PIII-8
4.4. Calibration of the Y fields using the regularized pilot points method.	PIII-8
4.5. Application of the calibrated Y fields to a transport prediction	PIII-10
5. Results	PIII-11
6. Conclusions	PIII-19
Acknowledgements	PIII-21
References	PIII-22
Appendix	PIII-25

REGULARIZED PILOT POINTS METHOD FOR REPRODUCING THE EFFECT OF SMALL SCALE VARIABILITY. APPLICATION TO SIMULATIONS OF CONTAMINANT TRANSPORT

1. Abstract	PIV-1
2. Introduction	PIV-1
3. Site description	PIII-3
4. Methodology	PIV-5
4.1. Tidal fluctuation response	PIV-5
4.2. Hydraulic tests	PIV-7
4.3. Model calibration	PIV-8
4.4. Transport prediction	PIV-11
5. Results	PIV-12
6. Conclusions	PIV-19

Acknowledgments	PIV-20
References	PIV-20
Appendix. Kriging with external drift for reconstructing heads at a borehole	PIV-23
Concluding comments	C-1

LIST OF FIGURES

PILOT POINTS METHOD INCORPORATING PRIOR INFORMATION FOR SOLVING THE GROUNDWATER FLOW INVERSE PROBLEM

<p>Figure 1. Schematic description of the pilot points method for defining a spatial random function $f(\mathbf{x})$, as the sum of a drift, $f_D(\mathbf{x})$, and a perturbation $f_p(\mathbf{x})$. 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).</p>	PI-3
<p>Figure 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 thirteen $\log_{10}T$ measurements (circles denote observation wells, while crosses mark pumping wells).</p>	PI-11
<p>Figure 3. “True” drawdowns after pumping ($t = 7200$ seconds) at wells B-1 (a), B-2 (b) and B-3 (c). The zone of interest (central square of $400 \times 400 \text{ m}^2$) has been enlarged two hundred meters each side. . .</p>	PI-13
<p>Figure 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 $\text{RMSE}_{\log_{10}T}$ (dashed horizontal line displays theoretical threshold value of $\sqrt{2}$).</p>	PI-17

Figure 5. Qualitative comparison of the results: Row 1. “True” field, log10T conditioning measurements, common scale bar and situation of pilot points. Row 2: drifts to be perturbed (column 1 obtained by ordinary kriging of the log10T measurements; columns 2 and 3 by sequential simulation). Rows 3-6 display the results after conditioning to log10T and drawdown measurements with varying weight μ . The “true” field is resembled when optimum weight is assigned, as measured by S2 (in all cases, when μ equals 0.1). Conditional estimation resembles the large scale patterns of the “true” field, despite the identifications of the spatial variability are oversmoothed. PII-19

Figure 6. Time evolution of measured (circles) and computed drawdowns in response to pumping at B-3 at selected observations points. Results of conditional estimation (black line) and two of the conditional simulations are presented. Notice that the fit of drawdown data is almost the same in the three cases. PII-21

REGULARIZED PILOT POINTS METHOD FOR REPRODUCING THE EFFECT OF SMALL SCALE VARIABILITY. APPLICATION TO SIMULATIONS OF CONTAMINANT TRANSPORT

Figure 1. Flow domain and location of conditioning measurements. The inset bounds the zone of interest (model domain for transport prediction). PIII-6

Figure 2. Zone of interest of the “true” Y fields. Notice that the large scale trends of NH (low Y in the middle, high Y in the lower two corners, etc) are reproduced in all realizations. Still, small scale variability, as reflected by the granularity of the fields, increases from NH to HH. PIII-7

Figure 3. “True” steady state drawdowns (in meters) at the zone of interest. Crosses at pictures in row 1 depict the location of pumping wells.	PIII-9
Figure 4. “True” log10 concentrations after 9900 seconds (row 1), 39000 seconds (row 2) and 1.4 10 ⁵ seconds (row 3).	PIII-10
Figure 5. Y fields obtained for the largest relevance of the small scale variability (HH test; subset of pictures at the right of the Figure) and when this variability is neglected (NH test). Two columns are displayed for each subset. Left column displays the Y field conditioned to Y measurements only. Right column displays the Y field conditioned to Y and drawdown measurements. Row 1 contains the “true” Y fields. Row 2 contains the estimated fields and rows 3 and 4 contain the results of two out of twenty simulations.	PIII-13
Figure 6. Histogram of Y fields. Results of conditional simulation are the ensemble of 20 realizations.	PIII-15
Figure 7. Total objective function for all realizations, when both Y and drawdown measurements are used for conditioning.	PIII-16
Figure 8. BTCs of the transport prediction for the Y fields conditioned to Y measurements only (left column) and Y and drawdown measurements (right column). “True” BTCs are depicted with dots, the BTC corresponding to the conditional estimation with thick line and the 20 BTCs corresponding to conditional simulation with thin grey line.	PIII-18
Figure 9. Mass balance of the transport prediction. a) Released volume of water through the upper contour of the zone of interest, b) Tracer mass in the aquifer at the end of the simulation.	PIII-19

GEOSTATISTICAL INVERSE MODELLING OF A COASTAL AQUIFER FROM TIDAL RESPONSE AND HYDRAULIC TEST DATA

Figure 1. Site description state. The study area (below) lays on top a Quaternary conglomerate (zone 1). Zone 2 accommodates two pipelines for sea water pumping. Zones 4, 5 and 6 represent land gained from the sea, covered with tetrapod marine defences. They cover the sea-shore except in its middle part (depicted by crosses), where a concrete wall panel was built for protection. Zones indicated by ‘7’ are concrete structures that cover almost the whole saturated thickness. Contamination has been detected in the three shaded zones, where the discharge pipeline (dashed line) was presumably broken. Observation boreholes are depicted by black circles. PIV-4

Figure 2. Kriged vs measured heads at borehole S24 during injection test 1 at borehole S5-1. PIV-7

Figure 3. Filtered response to injection test 1 at borehole S24. PIV-7

Figure 4. Estimated log-transmissivities using tidal response and injection tests as calibration data. a) hydraulic information-based model (i.e., zonation not accounted for explicitly). b) geology-based model. Dashed lines depict the contact between anthropogenic zones. Pilot points are depicted by dots. PIV-16

Figure 5. Calculated (line) vs measured (circle) tidal response using log10T field calibrated by the hydraulic information-based model (Figure 4a). Some measurements at boreholes S5, S7 and S5-3 are not depicted to clarify the figure. PIV-17

Figure 6. Calculated (line) and measured (circles) response to injections at boreholes S5-1 (boreholes S24 –on top right-, S29 and S12-1) and S25 (borehole S24 on bottom right) using log10T field calibrated by the hydraulic information-based model (Figure 4a). PIV-18

Figure 7. Measured (dots) and calculated concentrations obtained with the geology-based model (continuous line) and the hydraulic information-based (dashed line) models.	PIV-18
Figure A1. Measured (circles) and reconstructed (line) tidal response at borehole S5 using S9 measurements as external drift.	PIV-24

LIST OF TABLES

PILOT POINTS METHOD INCORPORATING PRIOR INFORMATION FOR SOLVING THE GROUNDWATER FLOW INVERSE PROBLEM

Table 1. Summary of results of the sensitivity analysis to the weighting factor and to the number of pilot points. Minimum values for each set are written in bold characters.	PI-15
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------

INVERSION OF HETEROGENEOUS PARABOLIC-TYPE EQUATIONS USING THE PILOT POINTS METHOD

Table 1. Summary of results of the sensitivity analysis to the weighting factor, for the conditional estimation CE (from Alcolea et al 2006) and two out of ten conditional simulations CS. Minimum values for each set are written in bold characters.	PII-16
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------

Table 2. Values of the weighting factor for which the estimation errors are minima (CE and CS denote conditional estimation and conditional simulation, respectively). The tested values of were 10^{-1} , 10^0 , 10^1 and 10^2 . The optimum identification of the spatial variability, as measured by the support function of the expected likelihood S2 (Equation 10) is always attained when equals 0.1.	PII-17
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------

REGULARIZED PILOT POINTS METHOD FOR REPRODUCING THE EFFECT OF SMALL SCALE VARIABILITY. APPLICATION TO SIMULATIONS OF CONTAMINANT TRANSPORT

Table 1. Statistical parameters of nested structures defining variograms used for the generation of the four “true” log-conductivity fields. PIII-8

Table 2. Statistical moments of the distribution of measurements, “true” and calibrated fields (in bold, values of mean and variance which are closest to the “true” ones). PIII-14

Table 3. Root mean square error of drawdowns (RMSEs) for all realizations. In bold, conditional simulations yielding worse drawdown fits than the corresponding conditional estimation. . . PIII-17

Table 4. Peak time, peak concentration and late-time slope of the BTCs for the different realizations (mean value and variance for conditional simulations). PIII-20

GEOSTATISTICAL INVERSE MODELLING OF A COASTAL AQUIFER FROM TIDAL RESPONSE AND HYDRAULIC TEST DATA

Table 1. Description of injection tests in the study area. PIV-6

Table 2. Available transmissivity measurements in the study area. This information arises from the standard analysis (e.g., homogeneous medium) of the injection tests at boreholes S5-1 and S25. PIV-8

Table 3. Summary of boundary conditions of the flow characterization model. PIV-10

Table 4. Summary of boundary conditions and parameters of the transport model. $f(t)$ denotes the time function simulating the mass of solute shifted from the non-saturated zone towards the saturated zone. This arises from a reactive transport model.	PIV-12
Table 5. Summary of average residuals (mean difference between calculated and measured state variables at a given borehole) of the calibration. P1, P2 and P3 denote the flow problems of tidal response and injections at boreholes S5-1 and S25, respectively. For instance, P3-S24 denotes measurements at borehole S24 corresponding to injection at borehole S25.	PIV-14
Table 6. Mean estimated transmissivities (m^2/d) in each anthropogenic zone using the hydraulic information-based and the geology-based models.	PIV-15

INTRODUCTION

INTRODUCTION

Heterogeneity has a large impact on groundwater flow and contaminant transport (Dagan, 1989; Carrera, 1993). Hydraulic conductivity K is the hydraulic parameter which exhibits most variability (Fogg et al, 2000). K displays many scales of heterogeneity at any given sample size (Neuman, 1990). For simplicity, only two scales are considered for modelling purposes: large scale heterogeneity, defined by the spatial patterns of connectivity, and small scale variability, defined by high frequency fluctuations of the property.

Inverse modelling represents a potentially powerful tool to characterize heterogeneity. Inversion is formulated in a geostatistical framework. Geostatistical inverse approaches estimate $\log K$ at every cell (or element). The optimum set of model parameters minimizes an objective function that quantifies the misfit between calculated and measured data. Often, only state variable data (e.g., heads) are considered and prior information of model parameters is disregarded. The formulation of these approaches is often non linear, so that solution of complex problems becomes too expensive. Therefore, one needs to reduce the number of unknowns by means of a parameterization scheme (McLaughlin and Townley, 1996).

The pilot points method is arguably the most flexible and widely used among parameterization schemes. This method was originally devised by de Marsily (de Marsily et al, 1984) and it has been applied to a number of problems (Vesselinov et al, 2001; Hernandez et al, 2003, among others). Yet, it suffers a number of limitations which have not been excluded from debate. First, the original method of de Marsily obtained the spatially correlated field through conditional estimation. This yields a single 'best' solution that minimizes the estimation error variance and honours available

measurements of the property. However, the resulting field is oversmoothed and does not allow a realistic representation of heterogeneity. To overcome this problem, some authors (RamaRao et al, 1995; Gomez-Hernandez et al, 1997; Capilla et al, 1997; Hendricks-Franssen, 2001) included conditional simulation, yielding a set of equally likely realizations of the property conditioned to available measurements. Therefore, each of these simulations reproduces the expected variability and, furthermore, they can be used to evaluate the uncertainty of predictions.

A second limitation arises from overparameterization and instability of the inverse problem (Cooley and Hill, 1995; Cooley, 2000). Instability implies unbounded fluctuations in the values of some model parameters. Instability has coerced the traditional use of the pilot points method. A possibility to fight unbounded fluctuations consists of imposing upper and lower bounds on model parameters (RamaRao et al, 1995; Gomez-Hernandez et al, 1997). In general, this causes the solution to fluctuate between these arbitrary bounds, but its reliability is not improved. A tactic to circumvent overparameterization consists of reducing the number of model parameters. Though the use of a small number of pilot points may overcome instabilities, it leads to a loss of resolution in the identification of heterogeneity.

Another tactic to combat instabilities consists of adding a regularization term to the objective function (whose minimization leads to optimal model parameters). Regularization has been used by Doherty (2003), who penalized non homogeneity of the unknown field. Kowalsky et al (2004) included the concept of parameter plausibility for the first time in the context of pilot points. These authors penalized the departure of model parameters from their prior information. Unfortunately, the role of the plausibility term was not explored. This dissertation is a step in this direction. It is aimed at showing that the use of a plausibility term improves (1) the identification of heterogeneity and (2) the stability of the problem. The latter allows the modeller to use an increased number of pilot points (in fact, as many as computationally feasible), thus sharpening the resolution of heterogeneity. This document consists of four papers, which are self-contained and can be read independently, and concluding comments.

In the first paper, the methodology, termed ‘regularized pilot points method’ (RPPM), is presented and its performance is explored on a groundwater flow synthetic

example. Conditional estimation of $\log T$ is performed on the basis of drawdown data and prior information of parameters. Emphasis is placed on assessing the optimal weighting of the plausibility term, which quantifies the importance of prior information of parameters in the calibration. Specifically, the role of the plausibility term in the improvement of the identification of heterogeneity and the sensitivity of the methodology to the number of pilot points are explored.

The second paper presents a comparison between the variants of conditional estimation and conditional simulation implemented in the RPPM. That is, the possibility of using a plausibility term in the case of seeking stochastic simulations of the unknown field ($\log T$) conditioned on direct measurements of the property and of dependent variables is explored.

In the third paper, the RPPM is framed in the context of the universal scaling theory (Neuman, 1990). The objective of this paper is to test the ability of the RPPM for reproducing the effect of small scale variability. Heterogeneity of $\log K$ is simulated by two nested variograms of short and long range, representing the small and large scales of variability, respectively. We aim at evaluating whether the presence of high frequency fluctuations impedes the characterization of high connectivity patterns. In the negative case, whether including small scale variability allows reproducing tailing in breakthrough curves.

Application of the RPPM to a real case is summarized in the fourth paper. The motivation of that work was the hydraulic characterization of a contaminated site as a first step to the design of a remediation system. This design demands an accurate characterization of hydraulic connectivity patterns, which are best measured by hydraulic diffusivity D (Knudby and Carrera, 2005). Tidal response data and prior information of model parameters should suffice for the characterization of D . However, calculations needed to design the remediation system demand the resolution of D into transmissivity and storage coefficient. To this end, response to two injection tests was added to the calibration data set. Two model structures are applied for the characterization of flow properties. They differ on whether or not the geological mapping is explicitly used for zonation.

References

- Capilla JE, Gomez-Hernandez 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: 175–188.
- Carrera J. An overview of uncertainties in modelling groundwater solute transport. *J Contam Hydrol* 1993, 13: 23-48
- 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): 1159–1163.
- Cooley RL, Hill MC. Comment on RamaRao et al (1995) and Lavenue et al (1995). *Water Resour Res* 2000; 36(9):2795–2797.
- Dagan G. *Flow and transport in porous formations*. 465 pp., Springer-Verlag, New York, 1989.
- 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: 831–849.
- Doherty J. Groundwater model calibration using pilot points and regularization. *Ground Water* 2003; 41(2):170–177.
- Fogg GE, Carle SF, Green C. Connected-network paradigm for the alluvial aquifer system. In *Theory, Modeling and Field Investigation in Hydrogeology: A Special Volume in Honor of Shlomo P. Neuman's 60th Birthday*, edited by D. Zhang and C.L. Winter. *Geol Soc Am Special Paper* 348, 2000: 25-42.
- Gomez-Hernandez 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):162–174.
- Hendricks-Franssen HJ. Inverse stochastic modeling of groundwater flow and mass transport. Ph.D. thesis, Technical University of Valencia, Spain, 2001.
- 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): 329–338.
- Knudby C, Carrera J. On the relationship between indicators of geostatistical, flow and transport connectivity. *Adv Wat Resour* 2005; 28 (4): 405-421.

- 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.
- McLaughlin D, Townley LLR. A reassessment of the groundwater inverse problem. *Water Resour Res* 1996; 32(5):1131–61.
- Neuman, SP. Universal scaling of hydraulic conductivities and dispersivities in geologic media. *Water Resour Res* 1990; 26(8): 1749–1758.
- 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): 475–493.
- 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):3019–3041.

PAPER I

**PILOT POINTS METHOD INCORPORATING PRIOR INFORMATION FOR
SOLVING THE GROUNDWATER FLOW INVERSE PROBLEM**

Andrés Alcolea, Jesús Carrera and Agustín Medina

Advances in Water Resources. *In press*. DOI:10.1016/j.advwatres.2005.12.009.

1. Abstract

The pilot points method is often used in non linear 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.

2. 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 (Carrera, 1987; Carrera et al, 2005, de Marsily et al, 1999, McLaughlin and Townley, 1996; Yeh, 1986). In order to identify heterogeneity, the groundwater inverse problem is usually formulated in a geostatistical framework. Early methods (Kitanidis and Vomvoris, 1983; Rubin and Dagan, 1987; Gutjahr and Wilson, 1989) 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 (Zimmerman et al, 1998), but as complexity increases, iterating is needed (Carrera and Glorioso, 1991; Carrera et al, 1993; Zimmerman et al, 1998). However, geostatistical formulations estimate log transmissivity at every cell (or element), so that the non linear solutions become too expensive unless special numerical methods, such as the adjoint state method, are used (Medina and Carrera, 2003). This allows successful practical

application (Meier et al, 2001; Rötting et al, 2006) to complex problems, but it is difficult to program. Therefore, one needs to reduce the number of unknowns by means of some parameterization scheme. McLaughlin and Townley (1996) 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 (Figure 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 optimum 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 (de Marsily et al, 1984), but has undergone several modifications. RamaRao et al (1995) and Gómez-Hernández et al (1997) included conditional simulations in the generation of the initial field. The location of the pilot points has been studied by Lavenue and Pickens (1992) and Hendricks-Franssen (2001), among others. The pilot points method has become widely used and has been applied to different problems (Hernandez et al, 2003; Vesselinov et al, 2001).

However, Cooley and Hill (2000) and Cooley (2000) 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 (Hadamard, 1902 and 1932). Instability implies (a) large values of some model parameters due to unbounded fluctuations (Bastin and Duque, 1981), 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.

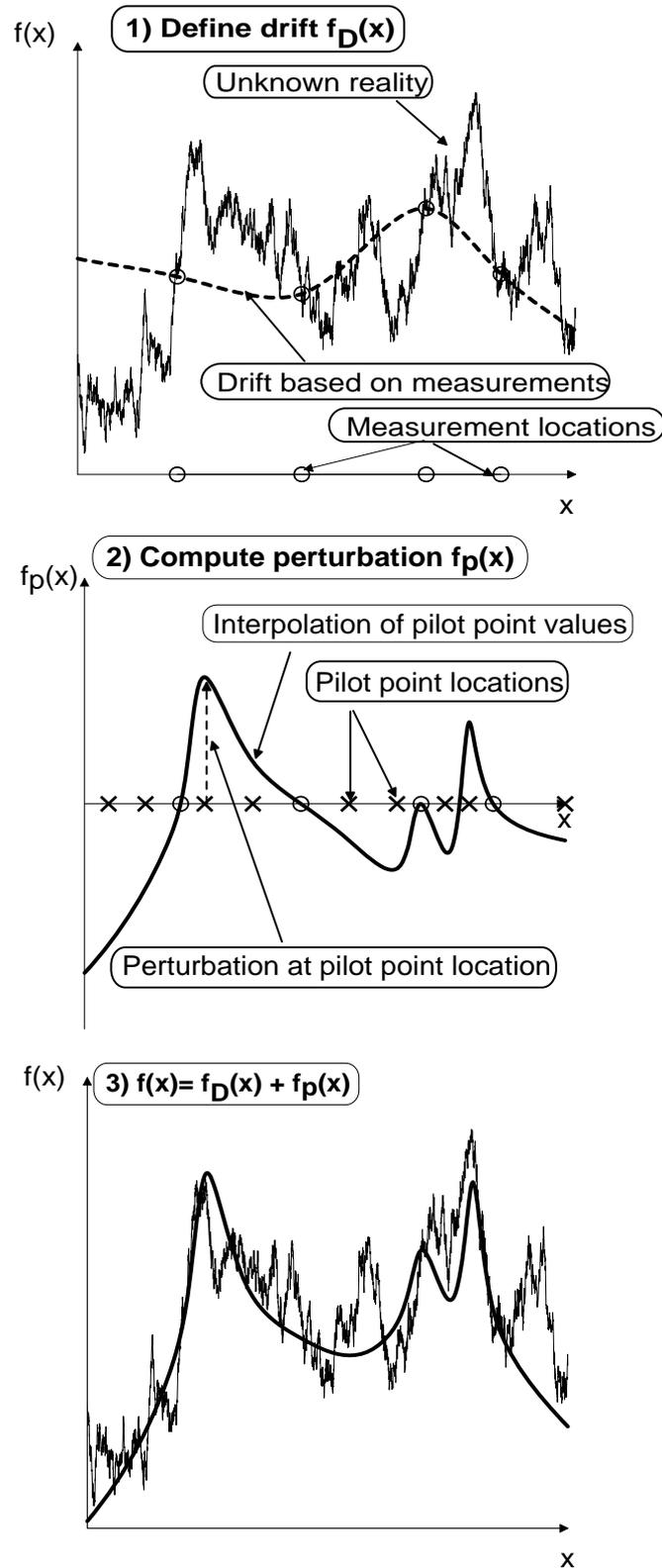


Figure 1. Schematic description of the pilot points method for defining a spatial random function $f(x)$, as the sum of a drift, $f_D(x)$, and a perturbation $f_p(x)$. The drift is defined by conditioning on available measurements. The perturbation is obtained from interpolation of the unknown pilot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

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 (1995) and Gómez-Hernández et al (1997) use this tactic. However, in general, this approach simply causes the solution to fluctuate between those arbitrary bounds, but its reliability is not improved (Neuman, 1973).

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 (RamaRao et al, 1995). New pilot point locations are set according to their ability for reducing the objective function, measured by the sensitivity coefficients (Lavenue and Pickens, 1992). 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; Capilla et al, 1997; Hendricks-Franssen, 2001) or vary randomly during the optimization process (Hendricks-Franssen, 2001). 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 (Doherty, 2003) and third, the problem is very sensitive to the location of the pilot points (Lavenue and Pickens, 1992).

A tactic to avoid large jumps in estimated parameters consists of penalizing them by adding regularization terms to the objective function: Tikhonov (1963a and b) imposes penalties to large values of the model parameters and unwarranted oscillations are penalized by Emsellem and de Marsily (1971). 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 (Carrera and Neuman, 1986a and b). 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 (1991) 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 (1995) 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 (Capilla et al, 1997; Gomez-Hernandez et al, 1997; Lavenue et al, 1995; Lavenue and de Marsily, 2001; Wen et al, 2002). Indeed, once the pilot point values have been estimated, the interpolation in step 2 of Figure 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 (2003), who penalizes non-homogeneity of the interpolated field rather than including the prior estimates of the model parameters. Kowalsky et al (2004) 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 (Medina et al, 2000), 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.

3. 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 (Figure 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 $f_D(\mathbf{x}, t)$ and an uncertain residual $f_p(\mathbf{x})$, which is a linear combination of the model parameters p_j :

$$f(\mathbf{x}, t) = f_D(\mathbf{x}, t) + f_p(\mathbf{x}) \quad (1)$$

- Step 2.1: Calculation of $f_D(\mathbf{x})$. 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, $f_D(\mathbf{x})$ 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:

$$f_D(\mathbf{x}, t) = \sum_{i=1}^{\dim Z} \lambda_i^Z(\mathbf{x}) Z(\mathbf{x}_i, t) \quad (2)$$

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 $f_p(\mathbf{x})$. It can be viewed as the perturbation of $f_D(\mathbf{x})$ 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):

$$f_p(\mathbf{x}) = \sum_{j=1}^{N_p} \lambda_j^{pp}(\mathbf{x}) p_j \quad (3)$$

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 (2003), the optimum set of model parameters minimizes the objective function:

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

where “nstat” denotes number of state variables \mathbf{u}_i with available measurements \mathbf{u}_i^* and covariance matrix \mathbf{V}_{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}_{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}_{u_i} and \mathbf{V}_{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 (Medina and Carrera, 2003):

$$F = F_d + \mu F_p = (\mathbf{s} - \mathbf{s}^*)^t \mathbf{V}_s^{-1} (\mathbf{s} - \mathbf{s}^*) + \mu (\mathbf{p} - \mathbf{p}^*)^t \mathbf{V}_p^{-1} (\mathbf{p} - \mathbf{p}^*) \quad (5)$$

The objective function stated in equation 4 (or its particularization in equation 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 (1976) derived it by maximizing the posterior pdf of the model parameters, MAP, while Carrera and Neuman (1986a) 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 (2003), 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 equation 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 (Cooley, 1985; Marquardt, 1963; Nowak and Cirpka, 2004):

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

where \mathbf{H}^k is an approximation of the Hessian matrix of F (equation 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:

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \Delta \mathbf{p}^k \quad (7)$$

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 of the following conditions is met (Medina et al, 2003): (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 (Medina and Carrera, 2003):

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

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.

4. 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 $400 \times 400 \text{ m}^2$. In essence, the procedure follows the steps of Meier et al (2001).

The flow domain is enlarged to avoid spurious boundary effects to a squared global domain of $3800 \times 3800 \text{ m}^2$ (Figure 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; Figure 2a) was generated with code TRANSIN (Medina et al, 2000) 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 [$\log_{10}(\text{m}^2/\text{s})$]. 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 equation 2 (calculated by ordinary kriging; Figure 2b) was radically different from the “true” field (Figure 2a). Notice that, indeed, the high $\log_{10}T$ channels

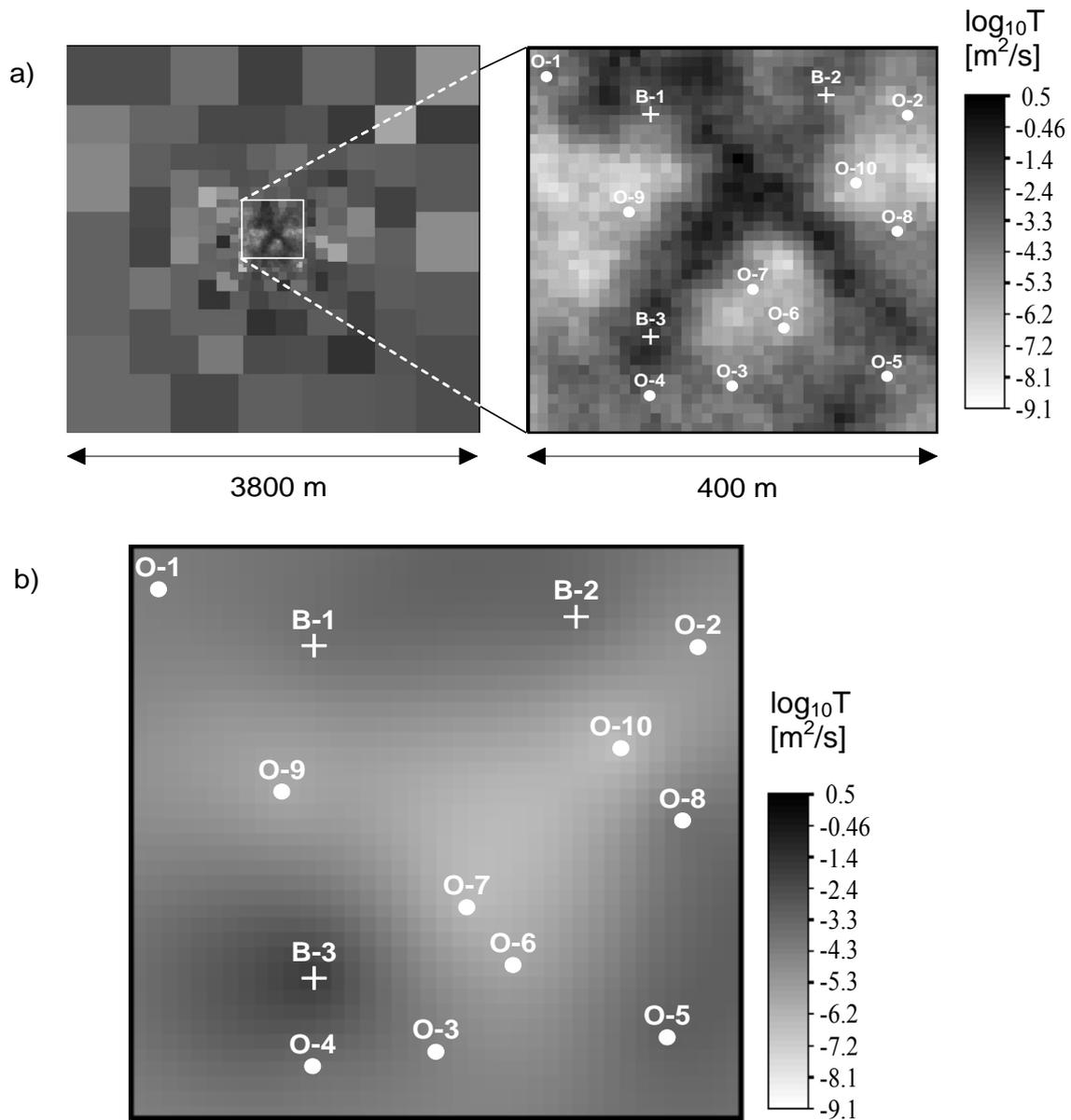


Figure 2. Test problem description. a) Flow domain, “true” $\log_{10}T$ field and location of conditioning measurements. All boundaries have a prescribed drawdown condition (zero). White square limits the zone of interest. Pumping tests are performed independently at points B-1, B-2 and B-3. b) Kriging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

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} m³/s at wells B1, B2 and B3 in Figure 2). Transient drawdowns were simulated at

grid nodes (Figure 3), assuming a zero drawdown as initial condition and prescribed at the boundaries. Drawdown measurements were calculated at the thirteen 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 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 (equation 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 (Figure 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 (Figure 5, column 1). Sixteen of them are located in the outer part of the domain (coarse discretization in Figure 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 (Capilla et al, 1997; Hendricks-Franssen, 2001). Observe that the number of $\log_{10}T$ measurement locations does not constrain the number of pilot points (13 $\log_{10}T$ measurement locations vs 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.

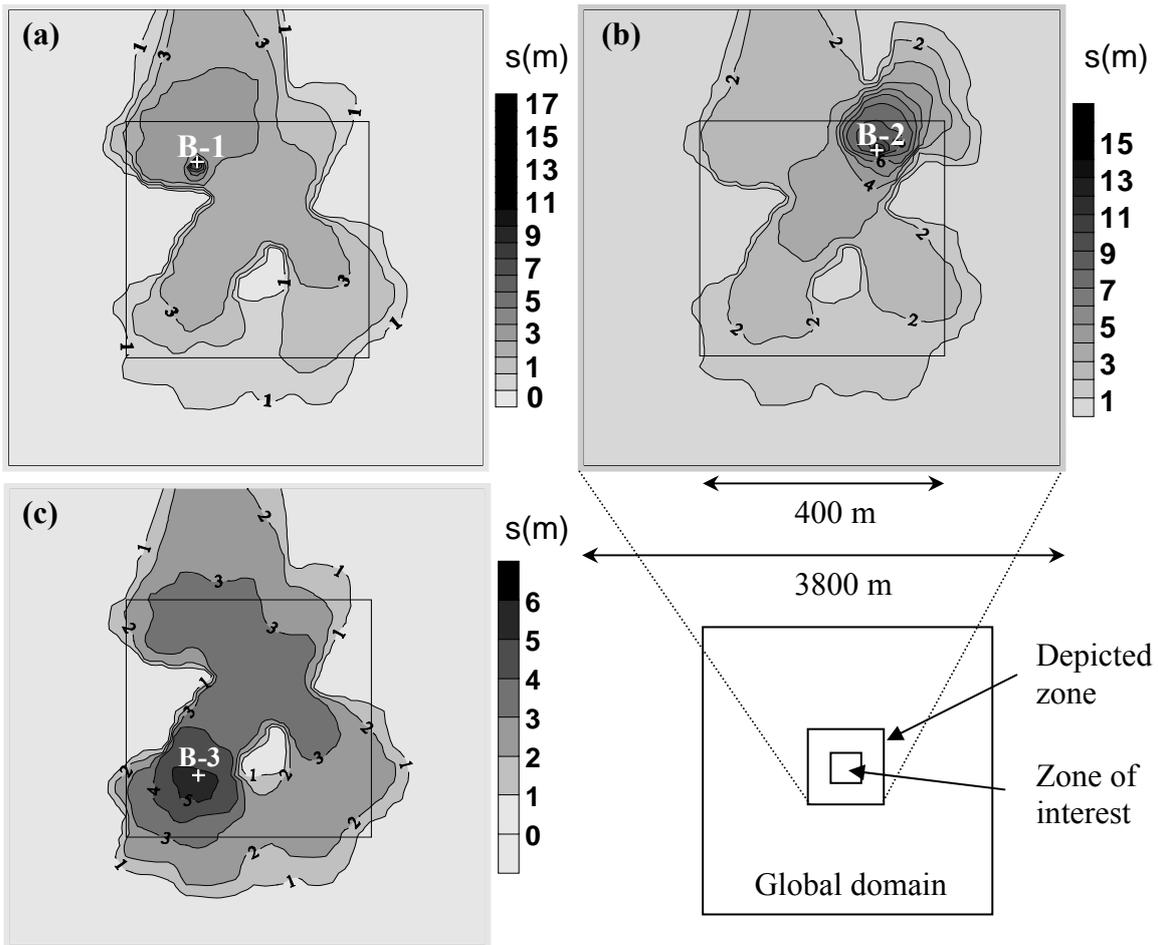


Figure 3. “True” drawdowns after pumping ($t = 7200$ seconds) at wells B-1 (a), B-2 (b) and B-3 (c). The zone of interest (central square of $400 \times 400 \text{ m}^2$) has been enlarged two hundred meters each side.

5. 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 \mathbf{e} is defined as the difference between calculated and “true” values of $\log_{10}T$ at the zone of interest (1600 blocks of $10 \times 10 \text{ m}^2$). We analyzed the following statistics:

- 1) Total objective function and its drawdowns and parameters components (F , F_d and F_p in equation 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 (equation 8), whose minimization should lead to the optimum value of μ .
- 3) Mean absolute error: measures the match between calculated and “true” values of $\log_{10}T$.

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

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, RMSE_d , was also calculated.

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

Table 1 summarizes the results concerning the identification of heterogeneity. Figures 4a and 4b display the quantitative comparisons in terms of the estimation errors, $\bar{e}_{\log_{10}T}$ and $\text{RMSE}_{\log_{10}T}$. Qualitative comparisons of $\log_{10}T$ estimates are presented in Figure 5. Figure 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, Figure 6) and to the worst identification of $\log_{10}T$ in all cases (Figure 5, column 1). That is, for any given number of pilot points, largest estimation errors, as measured by $\bar{e}_{\log_{10}T}$ and $\text{RMSE}_{\log_{10}T}$,

Table 1. Summary of results of the sensitivity analysis to the weighting factor μ and to the number of pilot points. Minimum values for each set are written in bold characters.

Test problem		Objective function (Equation 5)			Estimation errors		
N_p , number of pilot points	Weighting factor μ	Total obj. func. (F)	Drawdown obj. func. (F _d)	Param. obj. func. (F _p)	S_2 (Eq. 8)	$\bar{e}_{\log_{10}T}$ (Eq. 10)	RMSE $\log_{10}T$ (Eq.11)
---	$\mu \rightarrow \infty$	$1.156 \cdot 10^6$	$1.156 \cdot 10^6$	---	---	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
	$3 \cdot 10^{-1}$	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	34050	16160	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
	$3 \cdot 10^{-1}$	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	19141	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
	$3 \cdot 10^{-1}$	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	17426	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
	$3 \cdot 10^{-1}$	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	15358	4598	105	4423	1.525	2.138

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 (Figure 5, column 1). Similar appearance of estimated fields can be found in Zimmerman et al

(1998) and Alcolea et al (2002). The use of variable locations of the pilot points helps alleviating this problem (Hendricks-Franssen, 2001).

Similarly, large values of the weighting factor also yield poor results. The final solution tends to be too smooth (Figure 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 $RMSE_{\log_{10}T}$) 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.

Optimum identifications (Figure 5, column 2), as measured by criterion S_2 (equation 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 (Figure 4a) shows that, above this number, the improvement is small. However, the estimation variance (Figure 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 (Figure 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 (Figure 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 (Figure 7).

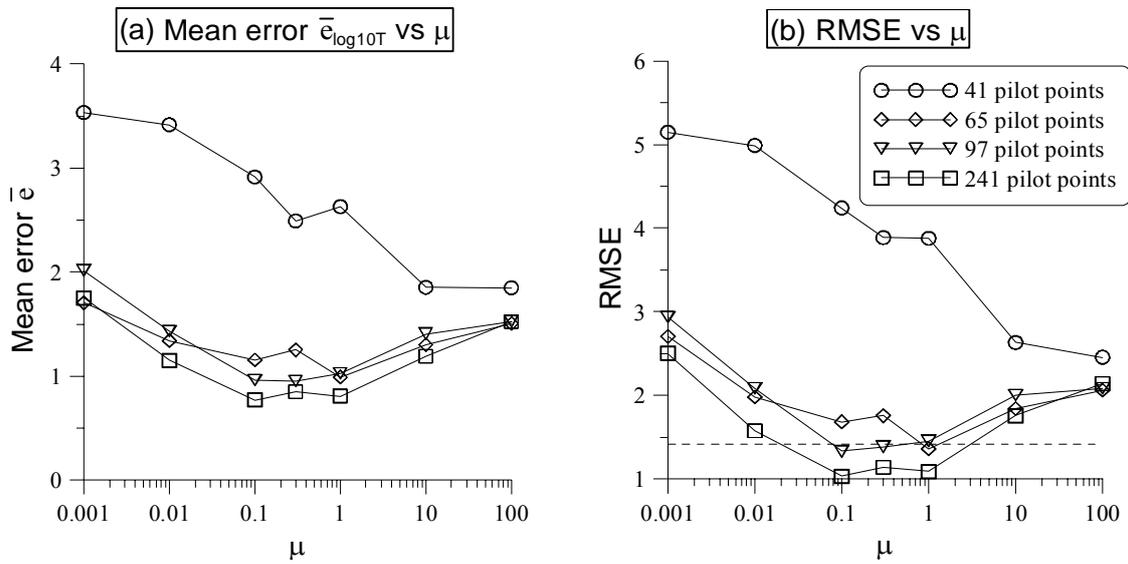


Figure 4. Log-Transmissivity estimation errors versus μ , weighting factor of the plausibility term: (a) Mean error $\bar{e}_{\log_{10}T}$, (b) Root Mean Square Error $RMSE_{\log_{10}T}$ (dashed horizontal line displays theoretical threshold value of $\sqrt{2}$)

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.

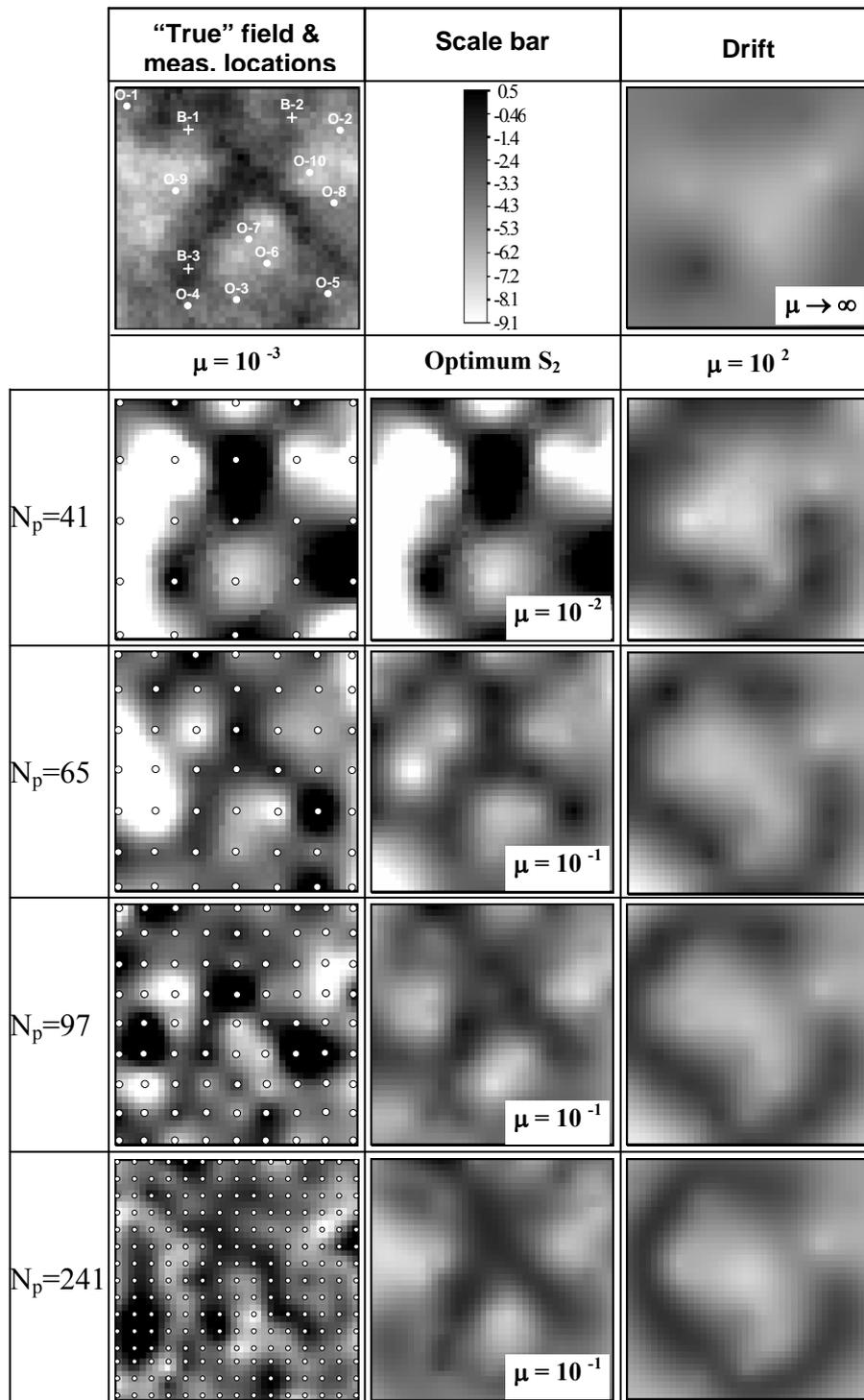


Figure 5. Qualitative comparison of results: Row 1. “True” $\log_{10}T$ field and measurements, common scale bar and the drift to be perturbed (obtained by ordinary kriging of the $\log_{10}T$ measurements). Rows 2-5 display log-transmissivities obtained after conditioning to $\log_{10}T$ and drawdown 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).

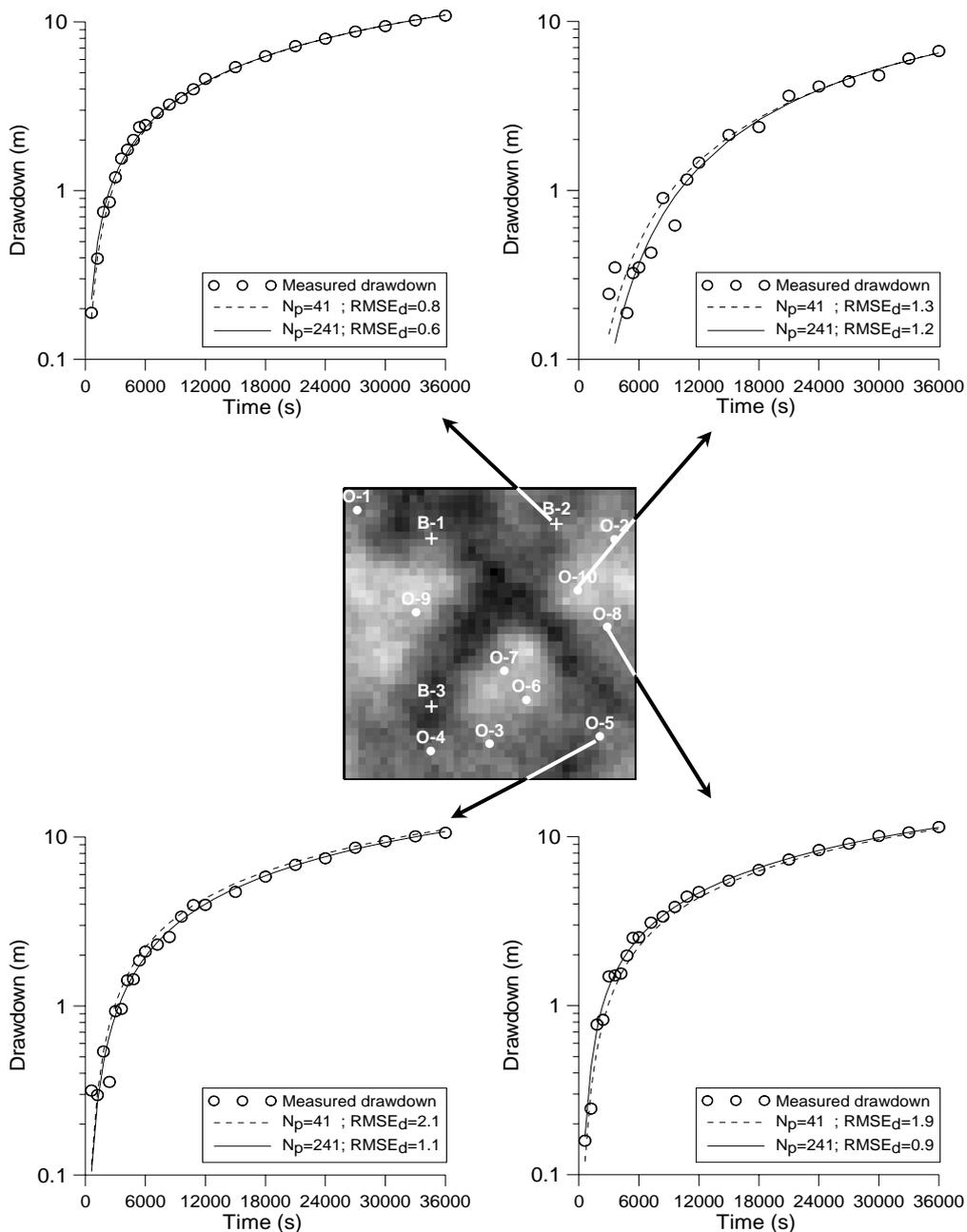


Figure 6. Time evolution of measured (circles) and computed (lines) drawdowns in response to pumping in B-3 at selected observation points. Weighting scalar of the plausibility term is 10^{-3} (best fit of measured drawdowns). The number of pilot points and the root mean square error of drawdowns, $RMSE_d$, 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 (Figure 5, column 1) and the calculated $RMSE_d$. 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.

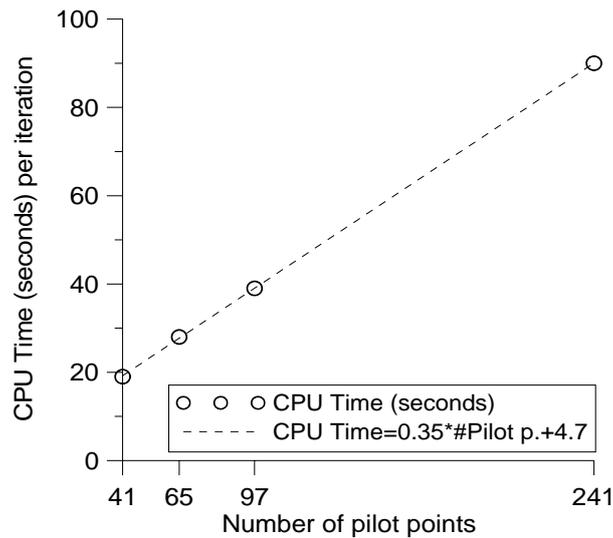


Figure 7. CPU time (in seconds) required for one iteration of the algorithm (steps 5 and 6). Results obtained in a Pentium IV platform.

6. 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 thirteen measurement points were available and they missed the channels of the true field (recall Figure 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.

Acknowledgements

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

- Alcolea A, Jodar J, Medina A, Carrera J. Geostatistical inverse problem: a modified technique for characterizing heterogeneous fields, in: Sanchez-Vila X, Carrera J and Gómez-Hernández JJ (ed) Proc GEOENV2002. Kluwer Academic Publishers, 2002:175-87.
- Bastin G, Duque C. Modelling of steady state groundwater flow systems, in Proc LASTED International symposium on Modelling, Identification and control, 1981:10-16.
- 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.
- Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions, 1. Maximum likelihood method incorporating prior information, Water Resour Res 1986a; 22(2): 199-210.
- Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions, 2. Uniqueness, Stability and Solution Algorithms, Water Resour Res 1986b; 22(2):211-27.
- 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 1987:549-85.
- Carrera J, Glorioso L. On geostatistical formulations of the groundwater flow inverse problem. Adv Water Resour 1991; 14(5):273-283.

- Carrera J, Medina A, Galarza G. Groundwater inverse problem. Discusión on geostatistical formulations and validation. *Hydrogeologie* 1993(4): 313-324
- Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ. Inverse problem in hydrogeology. *Hydrogeol J* 2005 (13): 206-222.
- 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.
- Cooley RL. Regression modeling of groundwater flow, USGS Open File Report 1985:85-180.
- 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.
- Cooley RL, Hill MC. Comment on RamaRao et al (1995) and Lavenue et al (1995), *Water Resour Res* 2000; 36 (9):2795-97.
- 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 et al. (ed). *Geostatistics for natural resources characterization. Part 2.*, D. Reidel Pub. Co 1984:831-49.
- de Marsily GH, Delhomme JP, Delay F, Buoro A. 40 years of inverse problems in hydrogeology, *Comptes Rendus de l'Academie des Sciences Series IIA Earth and Planetary Science* 1999, vol. 329(2):73-87. Elsevier Science.
- Doherty J. Groundwater model calibration using pilot points and regularization, *Ground Water* 2003; 41(2):170-77.
- Emsellem Y, de Marsily GH. An automatic solution for the inverse problem, *Water Resour Res* 1971; 7(5):1264-83.
- Gavalas GR, Shaw PC, Seinfeld JH. Reservoir history matching by Bayesian estimation. *J Soc Petrol Eng* 1976; 261:337-50.
- 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.
- Gutjahr AL, Wilson JR. Co-kriging for stochastic flow models, *Transp Porous Media*, 1989, 4(6) : 585-598
- 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.

- 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.
- Hendricks-Franssen HJ. Inverse stochastic modeling of groundwater flow and mass transport, Ph.D. thesis, Technical University of Valencia, Spain, 2001.
- 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.
- 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): 677-690
- 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.
- 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.
- 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.
- 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.
- Marquardt DW. An algorithm for least-squares estimation of non-linear parameters, *J Soc Ind Appl Math* 1963 ; 11:431-41.
- McLaughlin D, Townley LLR. A reassessment of the groundwater inverse problem, *Water Resour Res* 1996; 32(5):1131-61.
- 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 2000(9): 195-200.
- Medina A, Carrera J. Geostatistical inversion of coupled problems: dealing with computational burden and different types of data, *J Hydrol* 2003; 281:251–64.

- 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): 10-17.
- Nowak W, Cirpka A. A modified Levenverg-Marquardt algorithm for quasi-linear geostatistical inversing. *Adv Water Resour* 2004; 27(7):737-50.
- 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.
- 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.
- Rötting T, Carrera J, Bolzicco J, Salvany JM. Stream-stage response tests and their joint interpretation with pumping tests. *Ground Water*, 2006; 44 (3): 371-385.
- Rubin Y, Dagan G. Stochastic identification of transmissivity and effective recharge in steady groundwater flow: 1. Theory, *Water Resour Res* 1987, 23(7): 1185-1192
- Tikhonov AN. Solution of incorrectly formulated problems and the regularization method. *Sov Math Dokl* 1963a; 4:1035-38.
- Tikhonov AN. Regularization of incorrectly posed problems. *Sov Math Dokl* 1963b; 4: 1624-27.
- 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.
- 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.
- Yeh WWG. Review of parameter estimation procedures in groundwater hydrology: The inverse problem, *Water Resour Res* 1986; 22:95-108.
- Zimmerman DA, de Marsily GH, Gotway CA, Marietta MG, Axness CL, Beauheim RL, Bras RL, Carrera J, Dagan G, Davies PB, Gallegos DP, Galli A, Gómez-Hernández J, Grindrod P, Gutjahr AL, Kitanidis PK, Lavenue AM, McLaughlin D, Neuman SP, RamaRao BS, Ravenne C, Rubin Y. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling

advective transport by groundwater flow, Water Resour Res 1998; 34(6): 1373-1413.

PAPER II

**INVERSION OF HETEROGENEOUS PARABOLIC-TYPE EQUATIONS
USING THE PILOT POINTS METHOD**

Andrés Alcolea, Jesús Carrera and Agustín Medina

International Journal for numerical methods in fluids. *In press*. DOI:10.1002/flid.1213.

1. Abstract

The inverse problem (also referred to as parameter estimation) consists of evaluating the medium properties ruling the behavior of a given equation from direct measurements of those properties and of the dependent state variables. The problem becomes ill-posed when the properties vary spatially in an unknown manner, which is often the case when modeling natural processes. A possibility to fight this problem consists of performing stochastic conditional simulations. That is, instead of seeking a single solution (conditional estimation), one obtains an ensemble of fields, all of which honor the small scale variability (high frequency fluctuations) and direct measurements. The high frequency component of the field is different from one simulation to another, but a fixed component for all of them. Measurements of the dependent state variables are honored by framing simulation as an inverse problem, where both model fit and parameter plausibility are maximized with respect to the coefficients of the basis functions (pilot point values). These coefficients (model parameters) are used for parameterizing the large scale variability patterns. The pilot points method, which is often used in hydrogeology, uses the kriging weights as basis functions. The performance of the method (both its variants of conditional estimation / simulation) is tested on a synthetic example using a parabolic-type equation. Results show that including the plausibility term improves the identification of the spatial variability of the unknown field function and that the weight assigned to the plausibility term does lead to optimal results both for conditional estimation and for stochastic simulations.

2. Introduction

Parabolic equations represent natural diffusive phenomena and are used in many branches of engineering. Examples are the equations of heat conduction (industrial), groundwater flow (hydrogeology), or molecular diffusion (chemistry and contaminant transport), among others. Spatial variability of the properties (e.g. hydraulic conductivity for groundwater flow) entering those equations can be high but unknown, especially when they represent natural media. Moreover, it rules the performance of those equations (Freeze, 1975). Therefore, it has to be accounted for in meaningful

models. Identification of the spatial variability is carried out in the context of inverse modeling (Carrera et al, 2005; Carrera, 1987; Ouyang, 1992; Schnur and Zabararas, 1992; de Marsily et al, 1999; McLaughlin and Townley, 1996; Yeh, 1986) (also referred to as parameter estimation), which consists of estimating the properties as field functions from direct measurements of the properties (e.g. point values of thermal or hydraulic conductivity) and of dependent state variables (e.g. temperature and head for the heat transfer and groundwater flow equations, respectively).

The natural formulation of the inverse problem consists of assuming the state variable to be known and assume the field functions defining medium properties to be unknown. Such formulation is often ill-posed (i.e., a solution may not exist, it may not be unique, and it is usually unstable) for parabolic equations (Carrera and Neuman, 1986a, b and c). Moreover, it is not useful for practical purposes because the state variable is never known throughout the model domain. Therefore, one needs to parameterize the solution (i.e., to write the field functions in terms of a, hopefully small, number of parameters). Most parameterization techniques may be viewed as functional spaces where the parameters are the interpolation coefficients and the set of interpolation functions is a basis. A number of parameterization techniques have been used. Among them, the method of pilot points (de Marsily et al, 1984) has gained steam during recent years in hydrogeology because it is flexible and because it is formulated in a geostatistical context, so that it allows natural extensions to stochastic solutions of the governing equations. These are required when variability is important and unknown (Freeze, 1975; Narayanan, 1992).

The pilot points method consists of : (1) generating an initial spatially correlated field given a geostatistical model (i.e. measurements, if any, and correlation structure of the field function), (2) defining an interpolation method to obtain the value of the field functions over the model domain on the basis of their values at measurement and pilot point locations (model parameters) and (3) obtaining 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 state variables are considered). Thus, finding the optimum value of model parameters becomes an optimization problem. Notice that step 3 implies the perturbing the field generated in step 1.

As described above, the method suffers severe limitations. On the one hand, it is unstable, so that estimated pilot point values often reach non-plausible values. The inclusion of a regularization term to overcome this problem has been the subject of debate. Certes and de Marsily (1991) reject the use of such a term, questioning its performance, because it depends to a large extent on the reliability of the prior estimates. RamaRao et al (1995) argue that the plausibility is achieved inherently, given that the initial field to be perturbed already honours (1) the available measurements of the field function and (2) the covariance structure describing the spatial variability patterns as observed in the field. Similar arguments are used by other researchers for rejecting the plausibility term (Capilla et al, 1997; Gomez-Hernandez et al, 1997; Lavenue et al, 1995; Lavenue and de Marsily, 2001; Wen et al, 2002). Regularization has been used by Doherty (2003). Yet, his objective was to penalize non-homogeneity of the interpolated field rather than to include prior information about model parameters. Kowalsky et al (2004) used geophysical measurements (Ground Penetrating Radar) in a maximum a posteriori geostatistical context. Recently, Alcolea et al (2006) proposed adding a plausibility term to the objective function, so as to penalize departures of pilot point values from their prior estimates obtained by kriging. They showed that the use of such a regularization term improves (1) the identification of spatial variability and (2) the stability of the problem, allowing the use of larger number of pilot points, thus sharpening the resolution of the spatial variability. However, they also found that including the plausibility term may lead to worse results than simply interpolating from measurements (i.e. not inverting at all) if the plausibility term is not properly weighted. Fortunately, the use of a maximum likelihood statistical framework allows the identification of the optimum weight of the plausibility term.

A second limitation of the pilot points method is related to the definition of the initial spatially correlated field: the original method of de Marsily obtained this field through conditional estimation (variants of kriging, Gu, 2003; Krige, 1951). This yields a single “best” solution that minimizes the field variance and honors the available measurements of the field function. However, the resulting field is oversmoothed and does not allow a realistic representation of spatial variability. To overcome this problem, some authors (RamaRao et al, 1995; Capilla et al, 1997; Gomez-Hernandez et al, 1997; Hendricks-Franssen, 2001) included conditional simulations in the generation

of the initial field, yielding a set of equally probable realizations of the field functions conditioned to available measurements. That is, each of these simulations reproduces the expected variability and honors measurements. Therefore, they can be used to evaluate the uncertainty of predictions. Unfortunately, these methods do not allow the inclusion of a plausibility term, so that they are essentially unstable.

The objective of this paper is to overcome the above limitations. Specifically, we seek a formulation of the inverse problem capable of generating equally probable simulations of the field functions (e.g. transmissivity field) that are conditioned to measurements of the medium properties and the state variables (e.g. heads). To this end, we extend the method of Alcolea et al (2006) to the case of conditional simulation. That is, we explore the possibility of using a regularization term in the case of seeking stochastic simulations of the properties conditioned upon point measurements of both those properties and dependent state variables.

This paper is organized as follows. First, the methodology is outlined. Second, a synthetic example using the parabolic groundwater flow equation is presented. The paper ends with a discussion of the results and some conclusions about the use of the plausibility term in the context of the pilot points method.

3. 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 field functions at the pilot point locations) is updated through the optimization process. We assume that the (geostatistical) characteristics of the field functions are known. Here, the geostatistical model is defined by an autocorrelation function, but more sophisticated models may be used to represent complex heterogeneity patterns, geophysical data or known features. The procedure can be summarized as follows:

- **Step 1. Parameterization.** A field function f is expressed as the superposition of two fields: a known drift $f_D(\mathbf{x},t)$ and an uncertain residual $f_p(\mathbf{x})$, which is a linear combination of the model parameters p_j (Figure 1):

$$f(\mathbf{x}, t) = f_D(\mathbf{x}, t) + f_p(\mathbf{x}) \quad (1)$$

- Step 1.1. Calculation of $f_D(\mathbf{x},t)$. 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 (i.e. direct measurements of the field function \mathbf{f}^*) and possibly soft data \mathbf{g}^* (correlated with \mathbf{f}), that can be included as external drifts. In the case of conditional simulation, $f_D(\mathbf{x},t)$ reproduces the spatial variability patterns as observed in the field (honors the correlation structure as well), if the geostatistical model defined previously is informative enough. For the simple case of linear interpolation, it can be expressed as:

$$f_D(\mathbf{x}, t) = \sum_{i=1}^{\dim Z} \lambda_i^Z(\mathbf{x}) Z(\mathbf{x}_i, t) \quad (2)$$

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 (Krige, 1951), organized in the vector $\mathbf{Z} = (\mathbf{f}^*, \mathbf{g}^*)$. Our implementation of the methodology allows a large set of conditional estimation / simulation methods: simple kriging, residual kriging, kriging with locally varying mean, kriging with external drift, simple cokriging, ordinary cokriging, ordinary cokriging (standardized to the mean value of the primary variable f), among the methods for conditional estimation (Gu, 2003), plus a sequential simulation algorithm for conditional simulations (Liu and Journel, 2004).

- Step 1.2. Parameterization of the uncertain residual $f_p(\mathbf{x})$. It can be viewed as the perturbation that the drift requires to honor measurements of dependent state variables. It is expressed as a linear combination of the model parameters (value of the field function at the pilot point locations):

$$f_p(\mathbf{x}) = \sum_{j=1}^{N_p} \lambda_j^{pp}(\mathbf{x}) p_j \quad (3)$$

where N_p is the number of pilot points used to parameterize $f_p(\mathbf{x})$ (this number does not need to be the same for all field functions representing properties) and $\lambda_j^{pp}(\mathbf{x})$ are the (co-) kriging weights for the model parameters p_j , calculated in the same way as λ_i^Z for measurements. In our implementation, the location of the pilot points can be fixed or vary randomly as the optimization process proceeds.

- **Step 2. Prior estimation.** Prior estimates of the pilot point values \mathbf{p}^* and the corresponding a priori error covariance matrix \mathbf{V}_p are obtained by conditional estimation / simulation to measurements in vector \mathbf{Z} . Notice that correlation is included during the estimation process. In fact, pilot point values should be close (i.e. highly correlated) when pilot point locations are close. Therefore \mathbf{V}_p is a full matrix.
- **Step 3. Objective function.** Using Maximum Likelihood Estimation (Medina and Carrera, 2003), the optimum set of model parameters minimize the objective function:

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

where “nstat” denotes number of state variables \mathbf{u}_i with available measurements \mathbf{u}_i^* (i.e., in groundwater, $i=1$ for heads / drawdowns, $i=2$ for concentrations, etc.); “ntypar” is the number of types of model parameters being optimized, with prior information \mathbf{p}_j^* (i.e., $j=1$ for pilot points linked to transmissivities, $j=2$ for storativities, etc.). Matrices \mathbf{V}_{u_i} and \mathbf{V}_{p_j} represent our best guess of the error covariance matrices of state variables and models parameters, respectively, and β_i , μ_j are weighting scalars correcting errors in the specification of \mathbf{V}_{u_i} and \mathbf{V}_{p_j} . In our synthetic example, we solve the groundwater flow equation using only drawdown data as state variable for identifying the spatial variability of the transmissivity field. Thus, we will term hereinafter F_d the term of state variables

and F_p the term of model parameters. Assuming that error covariance matrix of drawdown is correct ($\beta_1=1$), the simplified objective function can be written as:

$$F = F_d + \mu F_p = (\mathbf{s} - \mathbf{s}^*)^t \mathbf{V}_s^{-1} (\mathbf{s} - \mathbf{s}^*) + \mu (\mathbf{p} - \mathbf{p}^*)^t \mathbf{V}_p^{-1} (\mathbf{p} - \mathbf{p}^*) \quad (5)$$

While the objective function stated in equation 4 (or its particularization in 5) was originally based on favoring the best match of state variables (F_d), while ensuring stability and plausibility of the model parameters (F_p), it can also be derived in a statistical framework. Gavalas et al (1976) derived it by maximizing the posterior pdf (probability density function) of the model parameters, MAP, while Carrera and Neuman (1986a) arrived to it by maximizing the likelihood of the parameters given the data, MLE. Instead, Medina and Carrera (2003) prefer working with the expected value of the likelihood function, as it allows the most stable estimation of statistical parameters, i.e., β_i and μ_j . Here, we use the same formulation.

- **Step 4. Minimization.** The minimization of equation 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 (Marquardt, 1963; Cooley, 1985; Beraux and Clermont, 1995):

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

where \mathbf{H}^k is an approximation of the Hessian matrix of F (equation 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 (the so-called Marquardt's parameter).

When the objective function takes a form similar to equation 5, second order derivatives of the state variable with respect to the parameters are often neglected, and the approximation of the Hessian matrix can be written as:

$$\mathbf{H}^k = 2 \mathbf{J}_s^t \mathbf{V}_s^{-1} \mathbf{J}_s + 2 \mu \mathbf{V}_p^{-1} \quad (7)$$

where \mathbf{J}_s is the jacobian matrix (i.e. derivatives of drawdowns with respect to model parameters at iteration k). The latter can be calculated by direct derivation of the PDE or by the adjoint state method (Galarza et al, 1999). The gradient of the objective function can be written as:

$$\mathbf{g}^k = 2 \mathbf{J}_s^t \mathbf{V}_s^{-1} (\mathbf{s} - \mathbf{s}^*) + 2 \mu \mathbf{V}_p^{-1} (\mathbf{p} - \mathbf{p}^*) \quad (8)$$

- **Step 5. Updating the vector of model parameters.** After each iteration, the vector of model parameters is updated as:

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \Delta \mathbf{p}^k \quad (9)$$

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 4 and 5 are repeated until convergence, which is checked using the criteria of Medina et al (2000): (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 numbers 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.

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 a good convergence. On the contrary, starting from values that are too low often leads to poor convergence.

- **Step 6. 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 (Medina and Carrera, 2003):

$$S_2 = N + \ln |\mathbf{H}| + N \ln \left(\frac{F}{N} \right) - \sum_{j=1}^{n_{ypar}} k_j \ln \mu_j \quad (10)$$

Here, N is the total number of data, n_i and k_j are the number of measurements of state variable 'i' and the number of prior information data of the j-th parameter type, respectively and \mathbf{H} is the approximation of the Hessian matrix at the end of the optimization process.

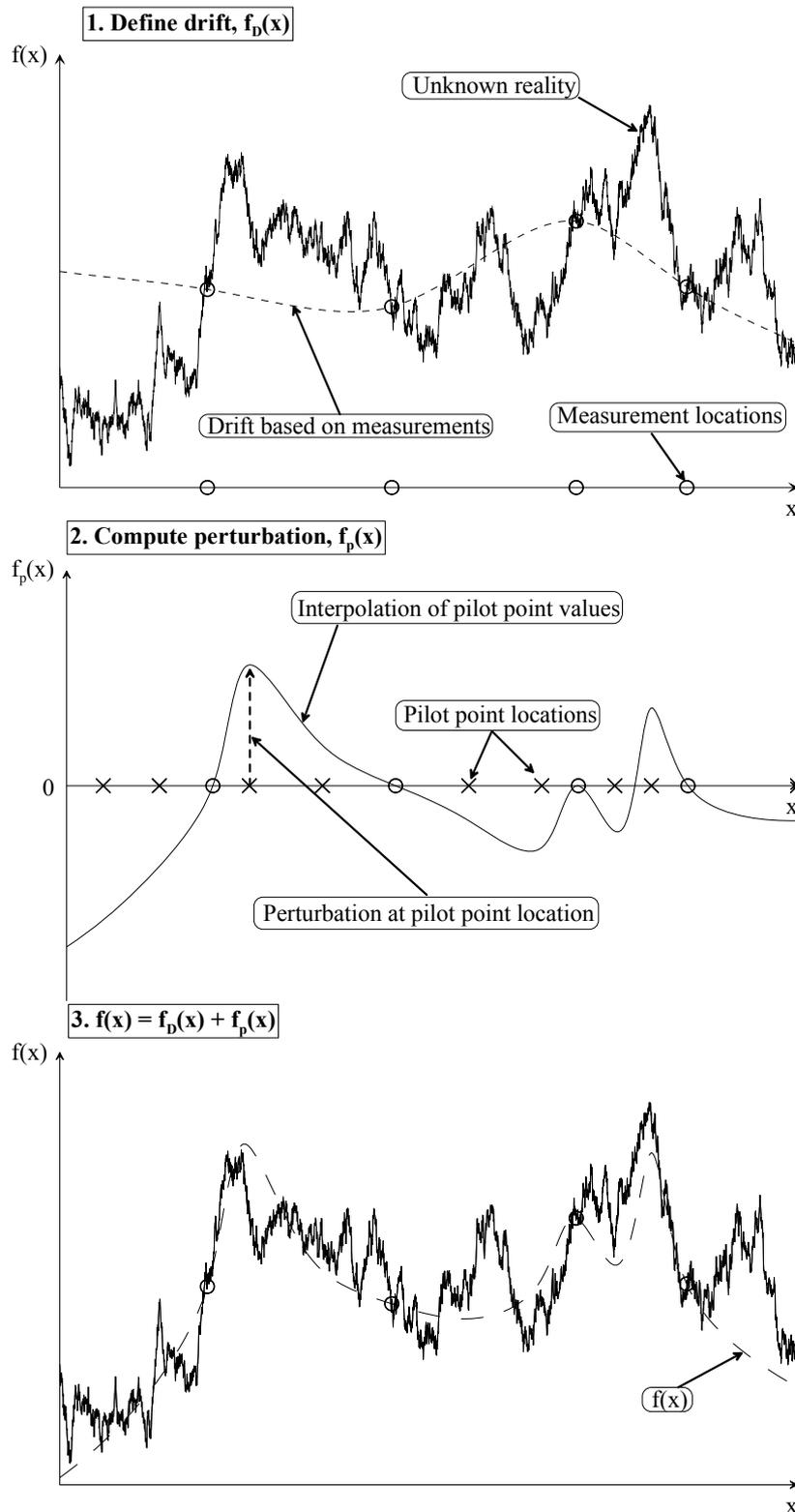


Figure 1. Schematic description of the pilot points method for defining a spatial random function $f(x)$, as the sum of a drift $f_D(x)$ and a perturbation $f_p(x)$. The drift is defined by conditional estimation (the smooth drift in the figure) or simulation (a sharper drift; not depicted) on available 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 (e.g. measurements of the state variable).

Notice that the methodology for variants of conditional estimation and conditional simulation only differs in the generation of the initial drift (step 1.1). This drift is unique in the case of conditional estimation and there are many realizations for conditional simulation. In the latter case, steps 2 to 6 must be repeated for each realization of the initial drift.

4. Application

The objective of this example is to extend the results of the previous work of Alcolea et al (2006) to the case of conditional simulation, exploring the possibility of using a plausibility term.

Results are explored on the basis of a synthetic example, consisting of the simultaneous interpretation of three pumping tests ruled by the parabolic groundwater flow equation:

$$\nabla(\mathbf{T}\nabla h) = S \frac{\partial h}{\partial t} \quad \text{on } \Omega \quad (11)$$

where Ω is the flow domain, \mathbf{T} is the transmissivity tensor, S is storativity and h is head (the state variable). Initial and boundary conditions can be written as:

$$\begin{aligned} h(t=0) &= h_0(\mathbf{x}) \quad \text{on } \Omega \\ \mathbf{T}\nabla h \mathbf{n} &= \alpha(H - h) + Q \quad \text{on } \Gamma \end{aligned} \quad (12)$$

where Γ denotes the boundary of the flow domain, \mathbf{n} is a unit vector normal to Γ and pointing outwards, H and Q are prescribed heads and flows, respectively and α is a coefficient controlling the type of boundary condition ($\alpha=0$ for prescribed flow, $\alpha \rightarrow \infty$ for prescribed head and a mixed condition otherwise). When pumping tests are interpreted, it is useful to work with drawdowns (difference between head in presence of pumping and head in absence of pumping), denoted as 's' hereinafter. This leads to homogeneous (zero) initial and boundary heads, as well as boundary flow rates. Equation (11) is solved applying the Galerkin method in space and forward finite differences in time. Elements are quadrangular bilinear.

The flow domain is a square of $400 \times 400 \text{ m}^2$, despite it is enlarged to avoid spurious boundary effects to a global domain of $3600 \times 3600 \text{ m}^2$ (Figure 2a). The finite element grid is more refined the central part (zone of interest, Figure 2). There, the finite element mesh is structured. Outside, the element size increases as the mesh progresses towards the boundary domain (Figure 2a).

The “true” log transmissivity field ($\log_{10}T$ hereinafter) was generated with the code TRANSIN (Medina et al, 2000) by sequential simulation (Figure 2a) conditional to a set of measurements defining two channels of high transmissivity. The “true” variogram (field variance minus autocorrelation function) 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 [$\log_{10} (\text{m}^2/\text{s})$]. In this work, only the 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. Measurement locations were purposefully selected in such a way that the initial drift of equation 2 (calculated by ordinary kriging or by sequential simulation, for the cases of conditional estimation / simulation, respectively) was radically different from the “true” field (Figure 2b). 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 . This setup was chosen to ensure that the plausibility term, which biases the solution 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 $10^{-2} \text{ m}^3/\text{s}$ at wells B1, B2 and B3 in Figure 2). Transient drawdowns were simulated at grid nodes (Figure 3), assuming a zero drawdown as initial condition and prescribed at the boundaries of the global domain. Drawdown measurements were calculated at the thirteen points where $\log_{10}T$ measurements are available (total of 936 drawdown data). A Gaussian white noise was added to those measurements, simulating acquisition errors, with 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).

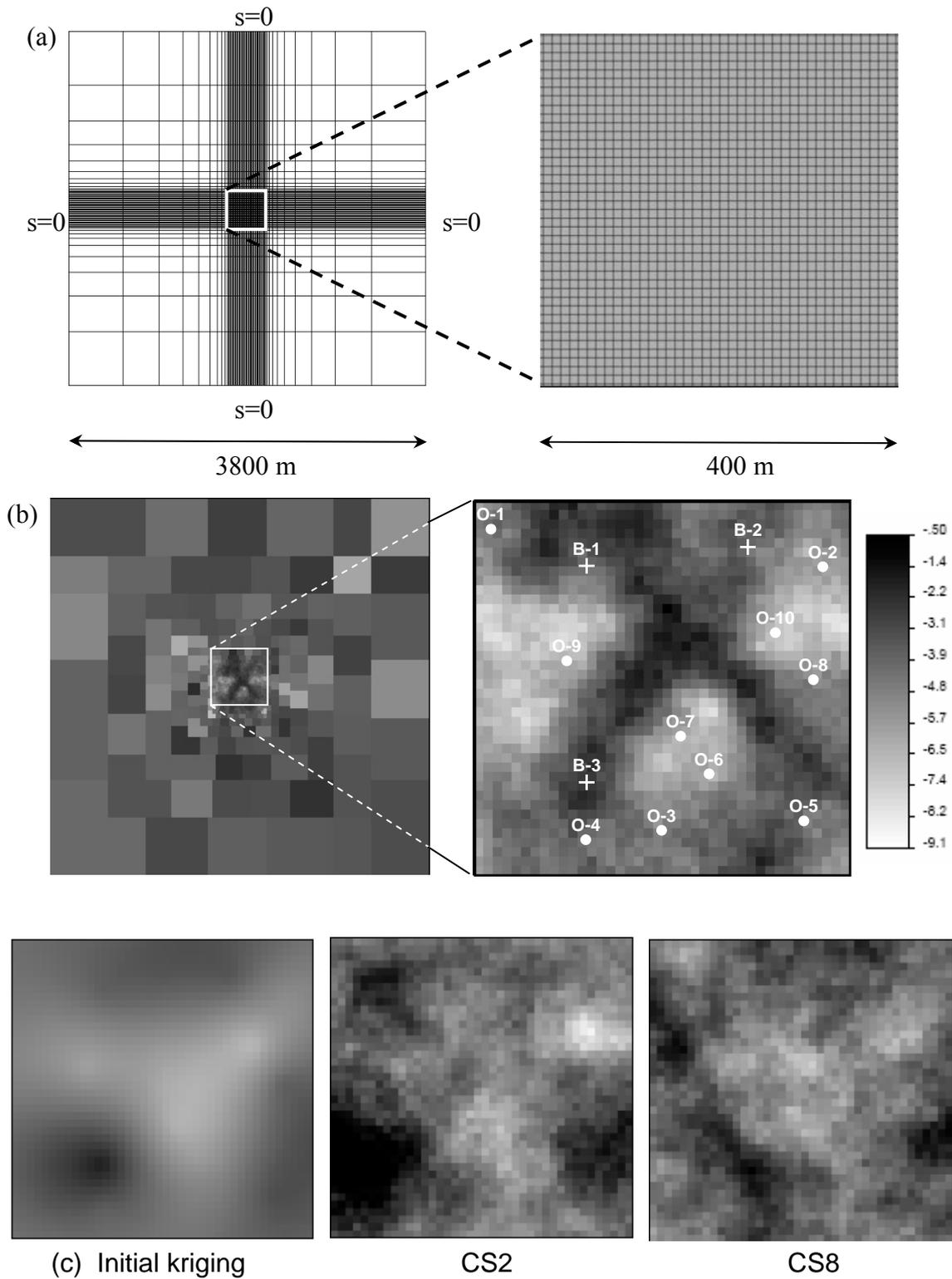


Figure 2. Test problem description. a) Finite element mesh and flow domain, b) true $\log_{10}T$ field and location of conditioning measurements. All boundaries have a prescribed drawdown condition (zero). White square limits the zone of interest, where three pumping tests are performed independently at points B-1, B-2 and B-3. Below, initial drifts, obtained by kriging of the thirteen $\log_{10}T$ measurements (c) and by conditional simulation (cases listed at Table 1). Notice that they are radically different from the “true” field depicted above.

In the previous work of Alcolea et al (2006), the optimum weight of the plausibility term was found for the case of conditional estimation. For the purpose of this paper, we take as starting point the conditional estimations using 97 pilot points located in a regular grid (Figure 5; row 1, column 3) and explore the optimum weight of the plausibility term when the initial correlated fields are drawn by conditional simulations (a total of 10 realizations). We use values of the weighting factor ranging from 10^{-1} to 10^2 . 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 the spatial variability, as the field would be biased to the initial drift (Figure 2b). On the contrary, small values of μ tend to disregard the plausibility term, thus risking instability.

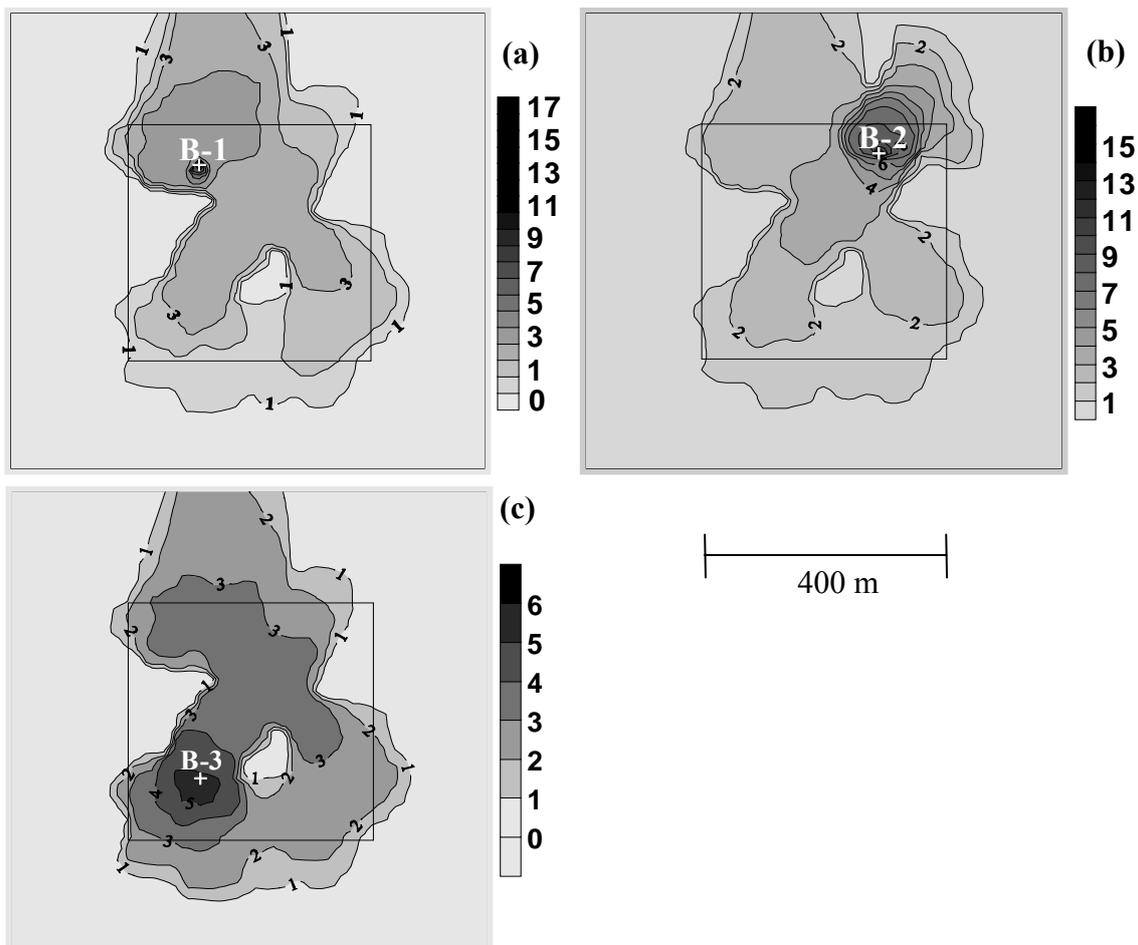


Figure 3. “True” drawdown (m) after pumping ($t=7200$ seconds) at wells B-1 (a), B-2 (b) and B-3 (c). The zone of interest (central square of $400 \times 400 \text{ m}^2$) has been enlarged two hundred meters each side.

5. Results

Results are explored in the same way as in the previous work [23]: qualitatively ($\log_{10}T$ maps and drawdown fits) and quantitatively. For the latter, an error vector \mathbf{e} is defined as the difference between calculated and “true” values of $\log_{10}T$ at the central part of the domain (1600 blocks of $10 \times 10 \text{ m}^2$). We analyze the following statistics:

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

$$\bar{e}_{\log_{10}T} = \frac{1}{1600} \sum_{i=1}^{1600} |e_i| = \frac{1}{1600} \sum_{i=1}^{1600} |\log_{10}T^{\text{calc}} - \log_{10}T^{\text{true}}| \quad (13)$$

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 on 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.

$$\text{RMSE}_{\log_{10}T} = \left(\frac{1}{1600} \mathbf{e}^t \mathbf{e} \right)^{1/2} \quad (14)$$

As will be discussed later, mean error and root mean square error are very sensitive to the location and extreme values of the zones of high / low transmissivity.

Table 1 displays a comparison between the evolution of the statistics with the weight of the plausibility term for the conditional estimation and two out of ten conditional simulations. Table 2 summarizes the quantitative comparisons and contains the value of μ for which the estimation statistics reach their optimum value. For instance, the minimum value of $\text{RMSE}_{\log_{10}T}$ is attained at $\mu=0.1$ in simulation 5. Figure 4 displays the quantitative comparison in terms of the support function of the expected likelihood S_2 and the estimation errors, $e_{\log_{10}T}$ and $\text{RMSE}_{\log_{10}T}$. Identifications of $\log_{10}T$ are presented in Figure 5. Figure 6 displays the best matching of drawdown.

The first observation that becomes apparent from Table 1 is the strong effect of the dependence of the plausibility term, as occurred in the previous work. The relative importance given to this term is measured by the value of the weighting factor μ . Using small values for this factor (little importance of the plausibility term, disregarding prior estimates in the optimization process and therefore, prior information) consistently leads to the best fit of drawdowns (minimum value of F_d) and to the worst fit of model parameters (maximum value of F_p), as expected. Identifications of the spatial variability (Figure 5; last row, column 2) using a weighting factor of 10^{-1} offer a somewhat “lumpy” appearance, with large jumps in the calibrated transmissivity over small distances. In fact, when the drift is generated by conditional estimation, Alcolea et al (2006) found the optimum identification when μ equals 0.1 (the minimum value tested in this example), that yields the worst qualitative identification of the $\log_{10}T$ field in this work. In fact, values of μ lower than 10^{-1} were excluded from this study due to instability problems.

Similarly, large values of the weighting factor also yield poor results. The final solution tends to be close to the initial drift (Figure 5, first row), which contains little information about the actual variability of the “true” field. However, estimation errors are sometimes smaller when μ equals 10^2 than the ones in the case of 10^1 (see CS2 in Table 1 and Figures 4b,c). We attribute this effect to the sensitivity of the estimation errors to the geometrical definition and extreme values of the high transmissivity channels (i.e. a small error in the position or the inclination of the channels may lead to large values of the estimation). As displayed in Figure 5 (column 2), the identification of the $\log_{10}T$ field in row 2 ($\mu=10^2$) is worse than the one in row 3 ($\mu=10^1$), although its

Table 1. Summary of results of the sensitivity analysis to the weighting factor μ , for the conditional estimation CE (from Alcolea et al, 2006) and two out of ten conditional simulations CS. Minimum values for each set are written in bold characters.

Test problem	Objective function (Equation 5)				Estimation errors		
	Weighting factor μ	Total obj. func. (F)	Drawdown obj. func. (F _d)	Parameters obj. func. (F _p)	S ₂ (Eq. 8)	$\bar{e}_{\log_{10}T}$ (Eq. 11)	RMSE _{$\log_{10}T$} (Eq.12)
CE	$\mu \rightarrow \infty$	1.156·10 ⁶	1.156·10 ⁶	---	---	1.390	1.831
	10 ²	17426	5566	119	4018	1.525	2.081
	10 ¹	3070	1318	175	2267	1.408	2.001
	10 ⁰	1033	829	203	1205	1.025	1.456
	3·10 ⁻¹	875	784	302	1074	0.950	1.386
	10 ⁻¹	787	753	348	1007	0.961	1.331
	10 ⁻²	759	744	1501	1075	1.431	2.080
	10 ⁻³	741	737	3690	1214	2.016	2.938
CS 2	$\mu \rightarrow \infty$	7.677·10 ⁵	7.677·10 ⁵	---	---	1.71	2.20
	10 ²	8467	2856	56	3193	1.37	1.82
	10 ¹	1621	917	70	1536	1.73	2.29
	10 ⁰	901	784	117	998	1.03	1.42
	10 ⁻¹	773	747	253	958	1.38	1.89
CS 8	$\mu \rightarrow \infty$	1.233·10 ⁶	1.233·10 ⁶	---	---	1.64	2.12
	10 ²	9294	2670	66	3286	1.36	1.89
	10 ¹	2117	1095	102	1804	1.40	1.94
	10 ⁰	912	783	129	1013	1.15	1.57
	10 ⁻¹	774	750	244	933	1.22	1.67

estimation errors are smaller. This effect is not reproduced for the case of conditional estimation.

Optimum identifications, as measured by criteria S₂, are obtained when μ equals 10⁻¹ in the ten conditional simulations, the same result attained by Alcolea et al (2006) for conditional estimation. This is important because it suggests that the modeler does not need to identify the optimum weight for each conditional simulation (usually a large number), but to obtain it just once using the method in its variant of conditional estimation.

Table 2. Values of the weighting factor μ for which the estimation errors are minima (CE and CS denote conditional estimation and conditional simulation, respectively). The tested values of μ were 10^{-1} , 10^0 , 10^1 and 10^2 . The optimum identification of the spatial variability, as measured by the support function of the expected likelihood S_2 (Equation 10) is always attained when μ equals 0.1.

Test problem	$\bar{e}_{\log_{10}T}$ (Equation 13)	RMSE $_{\log_{10}T}$ (Equation 14)
CE	10^{-1}	10^{-1}
CS 1	10^0	10^0
CS 2	10^0	10^0
CS 3	10^{-1}	10^{-1}
CS 4	10^0	10^0
CS 5	10^{-1}	10^{-1}
CS 6	10^2	10^2
CS 7	10^{-1}	10^{-1}
CS 8	10^0	10^0
CS 9	10^0	10^0
CS 10	10^2	10^2

Another result shared by conditional estimation and simulation is that, if the plausibility term is not properly weighted, the identification of the spatial variability is worse than the initial drift, as measured by estimation errors (Table 1). However, the use of the methodology in a maximum likelihood framework allows the estimation of the weighting factor μ . Therefore, the use of a plausibility term is advisable, independently of how the drift was calculated.

Figure 6 displays the best drawdown fit (μ equals 0.1) for the conditional simulations at Table 1 and the conditional estimation. Calculated drawdowns are very similar in all cases and fit the data. In fact, drawdown objective functions were very similar in all cases (Table 1). Despite the best match to drawdown data is obtained when the plausibility term is neglected, the drawdown fits for the optimum identification of the $\log_{10}T$ field (optimum weighting scalar μ) where nearly as good as the best ones ($\mu \rightarrow 0$) and the simulation was stable.

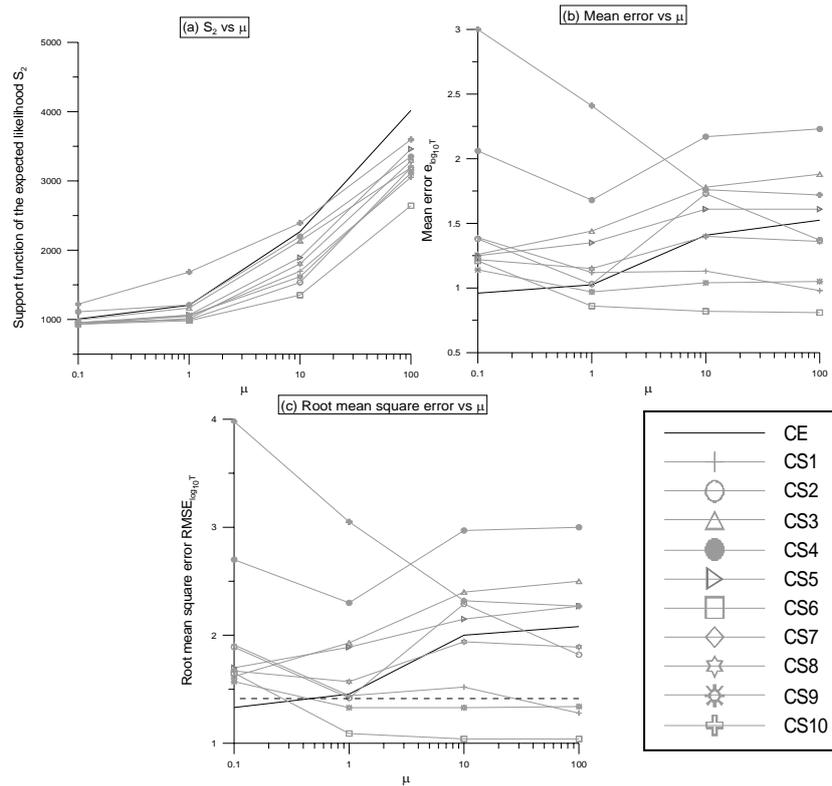


Figure 4. Support function of the expected likelihood and estimation errors versus μ and: (a) Support function S_2 , (b) Mean error $\bar{e}_{\log_{10} T}$, (c) $\text{RMSE}_{\log_{10} T}$ (dashed horizontal line displays theoretical threshold value of $\sqrt{2}$)

6. Conclusions

The pilot points method provides a powerful tool for the identification of the field functions ruling the behavior of a PDE. The suggested approach includes a plausibility term in the optimization process and two ways for calculating the initial drift, by conditional estimation or simulations conditioned to the direct measurements of the field function. Conditional estimation leads to optimal results (i.e., minimum estimation errors) but fails to reproduce small scale variability, which may be important when using the model for predictions. Instead, conditional simulation seeks a set of equally likely realizations of the field conditioned to all available information. Both variants have been tested on a synthetic example using the parabolic groundwater flow equation, examining the role of the plausibility term.

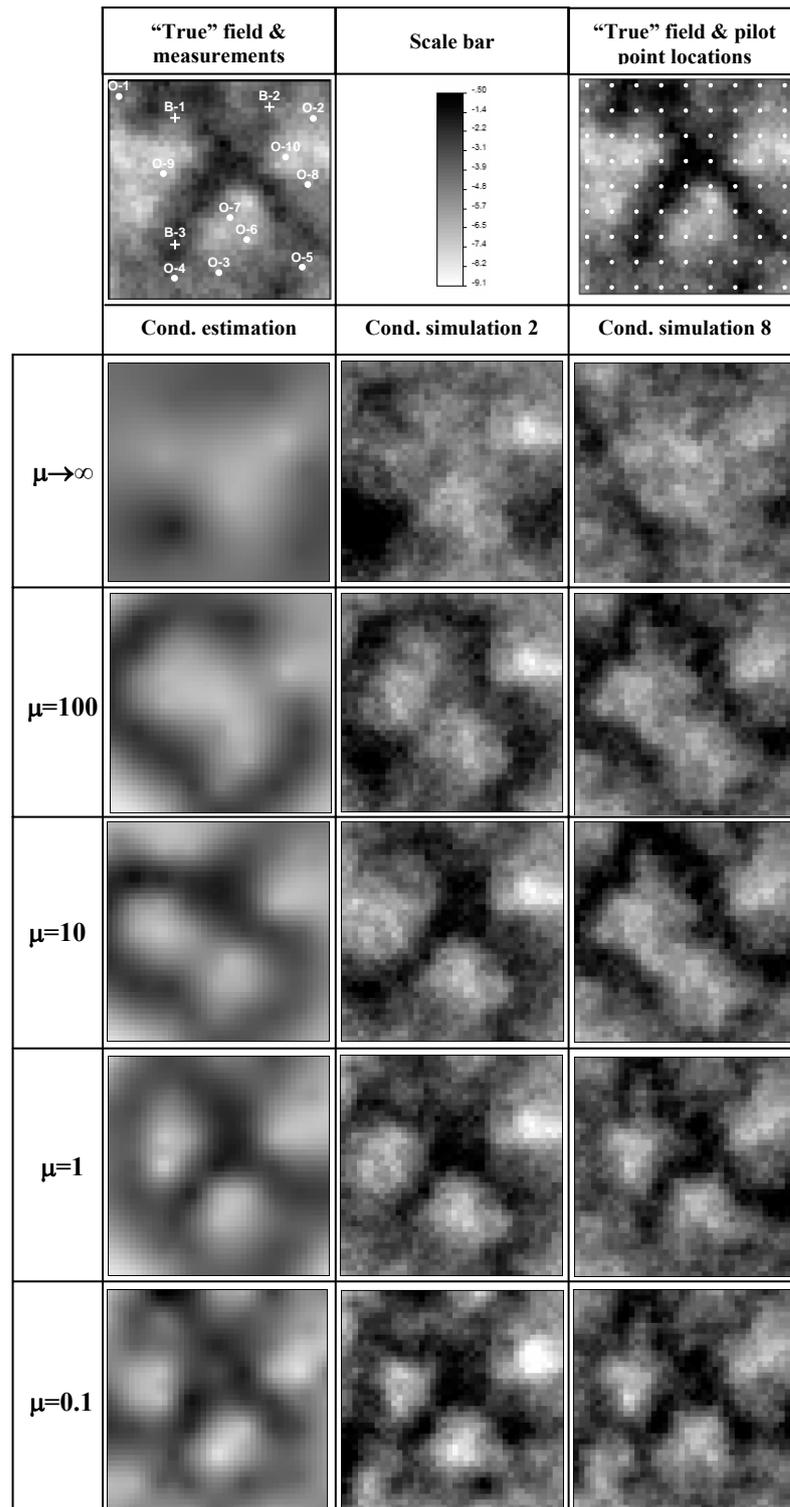


Figure 5. Qualitative comparison of the results: Row 1. “True” field, $\log_{10}T$ conditioning measurements, common scale bar and situation of pilot points. Row 2: drifts to be perturbed (column 1 obtained by ordinary kriging of the $\log_{10}T$ measurements; columns 2 and 3 by sequential simulation). Rows 3-6 display the results after conditioning to $\log_{10}T$ and drawdown measurements with varying weight μ . The “true” field is resembled when optimum weight is assigned, as measured by S_2 (in all cases, when μ equals 0.1). Conditional estimation resembles the large scale patterns of the “true” field, despite the identifications of the spatial variability are oversmoothed.

We have found that, neglecting the plausibility term, which is the standard approach in the context of pilot points, favors the best match of drawdown data, but often leads to an unstable identification of the model parameters. Large values give too much importance to the plausibility term, which biases the solution towards the drift. If the geostatistical model contains little information of the variability patterns (as in our synthetic example), the solution yields poor identifications of the spatial variability. In fact, a disturbing finding is that, in most cases, conditioning the fields to state variable data worsens the results if the plausibility term is not weighted properly. Fortunately, the use of a statistical framework (maximum likelihood in this case) allows the estimation of the optimum weight of the plausibility term, and therefore, its use is recommended. However, to search this optimum weight for each conditional simulation can be tedious.

A key finding of this work is that the optimum value of the weighting factor (as measured by the support function of the expected likelihood) was the same as the one obtained using conditional estimation. This frees the modeler of the burden of having to seek the optimum weight at each conditional simulation (usually a large number).

Good fits to measured state variable were obtained when neglecting (assigning low weight to) prior information. Still, fits nearly as good were obtained with stable simulations when moderate weights were assigned to prior information. We stress that the prior information provides a valuable data of aquifer heterogeneity, even when it is poorly informative of the actual variability. Thus, the use of a plausibility term including it (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).

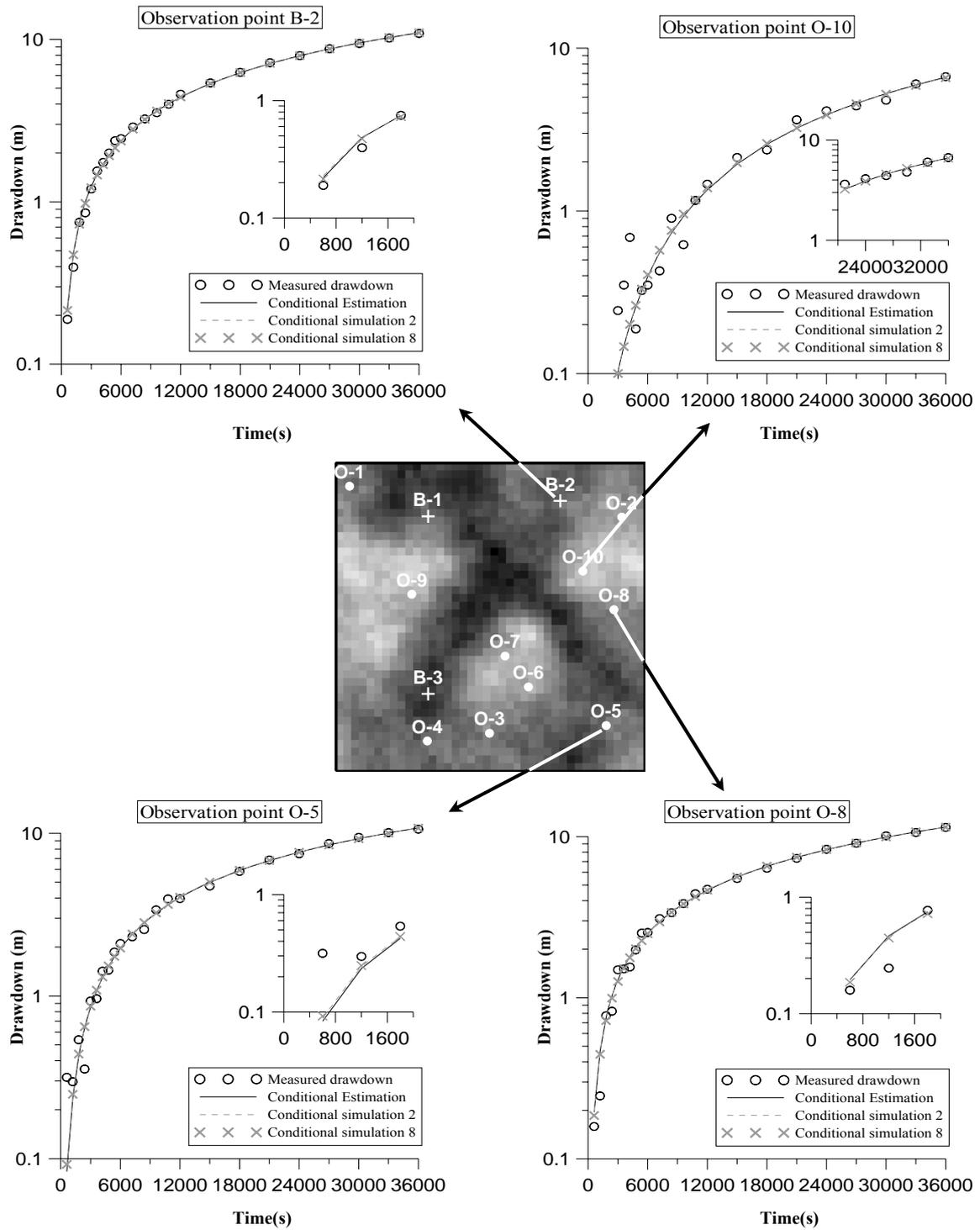


Figure 6. Time evolution of measured (circles) and computed drawdowns in response to pumping at B-3 at selected observations points. Results of conditional estimation (black line) and two of the conditional simulations are presented. Notice that the fit of drawdown data is almost the same in the three cases.

References

- Alcolea A, Carrera J, Medina A. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Advances in Water Resources*. In press. DOI:10.1016/j.advwatres.2005.12.009.
- Bereaux Y, Clermont JR. Numerical-simulation of complex flows of non-newtonian fluids using the stream tube method and memory integral constitutive-equations. *International Journal for Numerical Methods in Fluids*, 1995; 21(5): 371-389.
- 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. *Journal of Hydrology*, 1997; 203, 175-188.
- Carrera J, Neuman SP Estimation of aquifer parameters under transient and steady-state conditions, 1. Maximum likelihood method incorporating prior information. *Water Resources Research* 1986a; 22(2), 199-210.
- Carrera J, Neuman SP Estimation of aquifer parameters under transient and steady-state conditions, 2. Uniqueness, stability and solution algorithms. *Water Resources Research* 1986b; 22(2), 211-227.
- Carrera J, Neuman SP Estimation of aquifer parameters under transient and steady-state conditions, 3. Application to synthetic and field data. *Water Resources Research* 1986c; 22(2), 228-242.
- Carrera J. State of the art of the inverse problem applied to the flow and solute transport problems. *Groundwater Flow and Quality Modeling*, NATO ASI Series, 1987; 549-585.
- Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ. Inverse problem in hydrogeology. *Hydrogeology Journal*, 2005; 13(1), 206-222.
- Certes C, de Marsily GH. Application of the Pilot Point Method to the identification of aquifer transmissivities. *Advances in Water Resources*, 1991 ; 14(5), 284-300.
- Cooley RL. Regression modeling of groundwater flow. *USGS Open Repor*, 1985: 85-180.
- 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: *Geostatistics for natural resources characterization*. Part 2. Verly et al. Eds, D. Reidel Pub. Co, 1984; 831-849.

- de Marsily G, Delhomme JP, Delay F, Buoro A. 40 years of inverse problems in hydrogeology. *Comptes Rendus de l'Academie des Sciences Series IIA Earth and Planetary Science*, 1999; 329(2), 73-87. Elsevier Science.
- Doherty J. Groundwater model calibration using pilot points and regularization. *Ground Water* 2003; 41(2), 170-177.
- Freeze RA. A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media, *Water Resources Research*, 1975; 11(5),725-741.
- Gavalas GR, Shaw PC, Seinfeld JH. Reservoir history matching by Bayesian estimation. *Society of Petroleum Engineers journal*, 1976; 261, 377-350.
- Gómez-Hernández JJ, Sahuquillo A, Capilla JE. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 1. Theory. *Journal of Hydrology*, 1997, 204(1-4), 162-174.
- Galarza GA, Carrera J, Medina A. Computational techniques for optimization of problems involving non-linear transient simulations. *International Journal for Numerical Methods in Engineering*, 1999; 45(3): 319-334.
- Gu L. Moving kriging interpolation and element-free Galerkin method. *International Journal for Numerical Methods in Engineering*, 2003; 56 (1): 1-11.
- Hendricks-Franssen HJ. Inverse stochastic modelling of groundwater flow and mass transport, PhD Thesis, Technical University of Valencia, Spain. 2001.
- Kowalsky MB, Finsterle S, Rubin Y. Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone. *Advances in Water Resources*, 2004; 27, 583-599.
- Krige DG. A statistical approach to some mine evaluation and allied problems on the Witwatersrand. PhD Thesis, University of the Witwatersrand, Johannesburg. 1951.
- 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 Resources Research*, 1995; 31(3), 495-516.
- Lavenue M, de Marsily GH. Three-dimensional interference test interpretation in a fractured aquifer using the pilot-point inverse method. *Water Resources Research* 2001; 37(11), 2659-2675.
- Liu YH, Journel A. Improving sequential simulation with a structured path guided by information content. *Mathematical geology*, 2004; 36(8), 945-964.
- Marquardt DW. An algorithm for least-squares estimation of non-linear parameters. *J. Soc. Indust. Appl. Math.*, 1963; 11(2).

- McLaughlin D, Townley LLR. A reassessment of the groundwater inverse problem. *Water Resources Research*, 1996; 32(5), 1131-1161.
- 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: 195-200.
- Medina A, Carrera J. Geostatistical inversion of coupled problems: dealing with computational burden and different types of data. *Journal of Hydrology*, 2003; 281, 251–264.
- Narayanan VAB, Zabarar N. Stochastic inverse heat conduction using a spectral approach. *International Journal for Numerical Methods in Engineering*, 1992; 60 (9): 1569-1593.
- Ouyang T. Analysis of parameter-estimation heat-conduction problems with phase-change using the finite-element method. *International Journal for Numerical Methods in Engineering*, 1992; 33 (10): 2015-2037 .
- 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 Resources Research*, 1995; 31 (3), 475-493.
- Schnur DS, Zabarar N. An inverse method for determining elastic-material properties and a material interface. *International Journal for Numerical Methods in Engineering*, 1992; 33(10):2039-2057.
- Wen XH, Deutsch CV, Cullick AS. Construction of geostatistical aquifer models integrating dynamic flow and tracer data using inverse technique. *Journal of Hydrology*, 2002; 255, 151-168.
- Yeh WWG. Review of parameter estimation procedures in groundwater hydrology: The inverse problem. *Water Resources Research*, (1986); 22(2), 95-108.

PAPER III

**REGULARIZED PILOT POINTS METHOD FOR REPRODUCING THE
EFFECT OF SMALL SCALE VARIABILITY. APPLICATION TO
SIMULATIONS OF CONTAMINANT TRANSPORT**

Andrés Alcolea, Jesús Carrera and Agustín Medina

1. Abstract

Small scale variability of hydraulic conductivity is relevant for properly simulating transport through heterogeneous media. Accepting that this scale of variability cannot be characterized, we aim at evaluating whether the presence of high frequency fluctuations, which define small scale variability, impedes the characterization of large scale variability patterns. In parallel, we investigate whether including small scale variability allows us to reproduce tailing in breakthrough curves. To this end, we apply the regularized pilot points method for simulating fields conditioned to available hydraulic information. Calibrated fields are applied to the prediction of a transport problem. Heterogeneity of hydraulic conductivity is represented by two nested variograms simulating small scale variability (short range variogram) and large scale patterns (long range). Application to four synthetic examples (with different importance of the small scale variability) show that, first, the calibrated fields reproduce the statistics of the “true” ones and, second, small scale variability is not critical for flow problems. More important, small scale variability leads to increased tailing in solute breakthrough curves and needs to be acknowledged for proper transport prediction.

2. Introduction

Characterization of heterogeneity is essential for contaminant transport. Solutions obtained using the advection-dispersion equation (ADE) while ignoring spatial variability display numerous departures from field observations. These include the well known scale effects of dispersivity (Gelhar, 1986; Lallemand and Peaudecerf, 1978; Neuman, 1990), but also directional effects of porosity (Sanchez-Vila et al, 1997; Neuman, 2005) and tailing in the breakthrough curves (Kennedy and Lennox, 2001; Fernandez-Garcia et al, 2005). The latter becomes critical for the design of remediation systems or the migration of contaminants from a geological waste disposal. The literature shows consistent discrepancies between the measured breakthrough curves (BTCs) predicted by the ADE and the measured ones (Vallochi, 1985; Carrera, 1993; Cortis and Berkowitz, 2004; Kosakowsky, 2004).

Tailing in BTCs is usually simulated by adding terms representing the exchange of solute between mobile and immobile regions (multirate mass transfer) to the ADE (Rasmuson, 1985; Haggerty and Gorelick, 1995). These terms can be represented by means of a memory function (Carrera et al, 1998; Haggerty et al, 2000). The concept has been generalized by the continuous time random walk approach (Kosakowsky et al, 2001; Dentz and Berkowitz, 2003; Margolin and Berkowitz, 2004), which allows for a systematic study of the transition from normal (Fickian) to anomalous transport behavior. However, none of these approaches links explicitly the additional terms to actual variability patterns, so that it is not possible to define them a priori.

While tailing may be attributed to heterogeneity, stochastic research has concentrated on explaining the scale growth of observed dispersivities (Dagan, 1986). In this context, the universal scaling theory of Neuman (1990) is particularly relevant to our work. According to this theory, hydraulic conductivity displays many scales of heterogeneity at any given sample size. In fact, it is this superposition what explains the observed scale dependence of dispersivity. For simplicity, we consider only two scales of variability. The large scale is comparable to the domain size. Variability at this scale can be characterized using geological maps, geophysics, model calibration, etc. Small scale variability is defined by (high frequency) fluctuations at length scales smaller than typical distances between boreholes. Its characterization is difficult with usually available data. As a result, small scale variability is often disregarded in hydrogeological modeling. In fact, Rubin et al (2003) developed an approach to define dispersivity as a function of the scale of variability truncated by modeling. Unfortunately, it is not known if the superposition of variability scales reproduces tailing. Certainly, ignoring small scale variability does not help.

Methods devoted to the identification of heterogeneity can be classified in two groups, termed conditional estimation and conditional simulation methods. The first group seeks a deterministic, though uncertain, optimum characterization in the sense of minimum estimation error, honoring all available data (typically, hydraulic conductivity and head measurements). This group includes linearized cokriging (Kitanidis and Vomvoris, 1983), conditional expectation (Dagan, 1985) and maximum likelihood

estimation (Carrera and Neuman, 1986), among others. While formulations of this group are different, they do not vary from each other in the essence (Carrera et al, 2005). All of them neglect the effect of small scale variability. This limitation can be overcome by conditioning moment equations, where one seeks estimates of mean parameters while acknowledging the effect of small scale variability (Hernandez et al, 2003, 2006). Conditional simulation methods are explicitly stochastic. They yield a number of equally likely realizations of the unknown field conditioned to all available information [Sahuquillo, 1992; Gomez-Hernandez et al, 1997; Capilla et al, 1999; Hendricks-Franssen, 2001, Hendricks-Franssen et al, 2003].

The pilot points method (de Marsily et al, 1984; Alcolea et al, 2006a and b; RamaRao et al, 1995; Lavenue and Pickens, 1992) is a flexible parameterization technique, that can be used both for conditional estimation and for conditional simulation. It allows reproducing the effect of small scale variability. It has been successfully applied to a number of problems (RamaRao et al, 1995; Vesselinov et al, 2001; Hernandez et al, 2003). Yet, it suffers a number of limitations, including instability. Alcolea et al (2006a and b) extended the method by adding a regularization term. This allowed them to use a large (of the order of 200) number of pilot points, thus being able to properly resolve the large scale trends of variability. This approach should help in realizing the hope of simulating hydraulic conductivity fields that are consistent with available large scale data and yet contain high frequency fluctuations. We argue that doing so in a consistent and reproducible manner is required for proper simulation of contaminant transport in hydrological practice.

The objective of this paper is to present a step in this direction. Specifically, we aim at evaluating whether the presence of high frequency fluctuations impedes the characterization of large scale variability trends. In the negative case, whether including small scale fluctuations allows us to reproduce tailing in BTCs.

3. Inversion methodology

The inversion technique used in this work is a modification of the pilot points method (de Marsily et al, 1984; Lavenue and de Marsily, 2001; Gomez-Hernandez et al, 1997), including the use of a plausibility term. Algorithmic details of this methodology can be found in the work of Alcolea et al (2006a and b). The procedure can be summarized as follows:

3.1) Parameterization. The unknown hydraulic property (typically $\log_{10}K$) is expressed as the superposition of two fields: a drift and an uncertain residual. The latter is a linear combination of the model parameters (value of the property at the pilot points locations). The drift can be calculated by conditional estimation or conditional simulation, in which case the drift is a random function. In both cases, all available information (direct measurements, geophysics, geological data, etc) can be used for conditioning.

3.2) Optimization of the model parameters. The optimum set of model parameters minimizes an objective function F that quantifies the misfit between calculated and measured data:

$$F(\mathbf{p}) = \sum_{i=1}^{nstat} \beta_i (\mathbf{s}_i - \mathbf{s}_i^*)^t \mathbf{V}_{s_i}^{-1} (\mathbf{s}_i - \mathbf{s}_i^*) + \sum_{j=1}^{ntypar} \mu_j (\mathbf{p}_j - \mathbf{p}_j^*)^t \mathbf{V}_{p_j}^{-1} (\mathbf{p}_j - \mathbf{p}_j^*) \quad (1)$$

where the first term measures the misfits between calculated and measured ‘nstat’ types of state variables (\mathbf{s}_i and \mathbf{s}_i^* respectively) and the second is a plausibility term which measures the departure of the model parameters (\mathbf{p}_j) from their prior information (\mathbf{p}_j^* ; ‘ntypar’ denotes the number of types of parameters). \mathbf{V}_s and \mathbf{V}_p (block matrices containing \mathbf{V}_{s_i} and \mathbf{V}_{p_j} , respectively) are the best guess of the corresponding covariance matrices and β_i and μ_j are weighting scalars correcting the specification of \mathbf{V}_{s_i} and \mathbf{V}_{p_j} , respectively. Prior estimates of model parameters can be calculated by conditional estimation or simulation to available measurements. For the case of conditional estimation, \mathbf{V}_p is the kriging error covariance matrix (\mathbf{V}_k). \mathbf{V}_p is corrected if conditional simulation is performed ($\mathbf{V}_p = 2\mathbf{V}_k$; see appendix 1).

3.3) Finding the optimum weighting scalars β_i and μ_j (a posteriori statistical analysis). The optimization process is repeated using different values of the weighting scalars, whose optimum values are the ones leading to the maximum of the expected likelihood of the parameters given the data (Medina and Carrera, 2003). Assigning low weights (μ_j) to model parameters disregards their prior information, but leads to the best fit of the measured state variables. On the contrary, assigning large weights to model parameters disregards the measured state variables, biasing the solution towards prior information.

4. Procedure for representing small scale variability

The objective of this work is to test the ability of the above inversion methodology for reproducing the effect of the small scale variability of hydraulic conductivity (although the procedure can be applied to any other parameter type). Accepting that the high frequency fluctuations cannot be characterized, we aim at evaluating if their presence impedes the characterization of large scale variability patterns. To this end, estimated (CE) and simulated (CS) drifts are conditioned to $\log_{10}K$ measurements (Y hereinafter) and to Y and drawdown measurements ('s' hereinafter). The outcomes are compared both qualitatively (drawdown fits, Y maps and histograms) and quantitatively (estimation errors). In parallel, we investigate the effect of the small scale variability of hydraulic conductivity on BTCs tailing. To this end, we apply the estimated/simulated fields to the prediction of a transport problem.

The procedure is tested using four synthetic cases on a single domain with increasing level of small scale variability. In essence, the procedure follows the steps of Meier et al (2001), Hendricks-Franssen (2001), Alcolea et al (2006a). It consists of generating data from a synthetic problem and then using this data to test the effectiveness of the inversion methodology. Specifically, our work consists of the following steps:

- 1) Define the problem setup.
- 2) Generate four "true" Y fields with increasing small scale variability.

- 3) Generate the available drawdown data.
- 4) Calibration of Y fields using the regularized pilot points method.
- 5) Application of the calibrated Y-fields to a transport prediction.

These steps are outlined below.

4.1 Problem domain

The problem domain is a $400 \times 400 \text{ m}^2$ square that is discretized using elements of $10 \times 10 \text{ m}^2$. This is enlarged to a global domain of $3800 \times 3800 \text{ m}^2$ (Figure 1) to avoid spurious boundary effects. Hydraulic tests and transport prediction take place in the central portion. Outside, the element size increases as the mesh progresses towards the boundaries. The whole domain is used for flow calibration, while only the central portion is used for transport predictions.

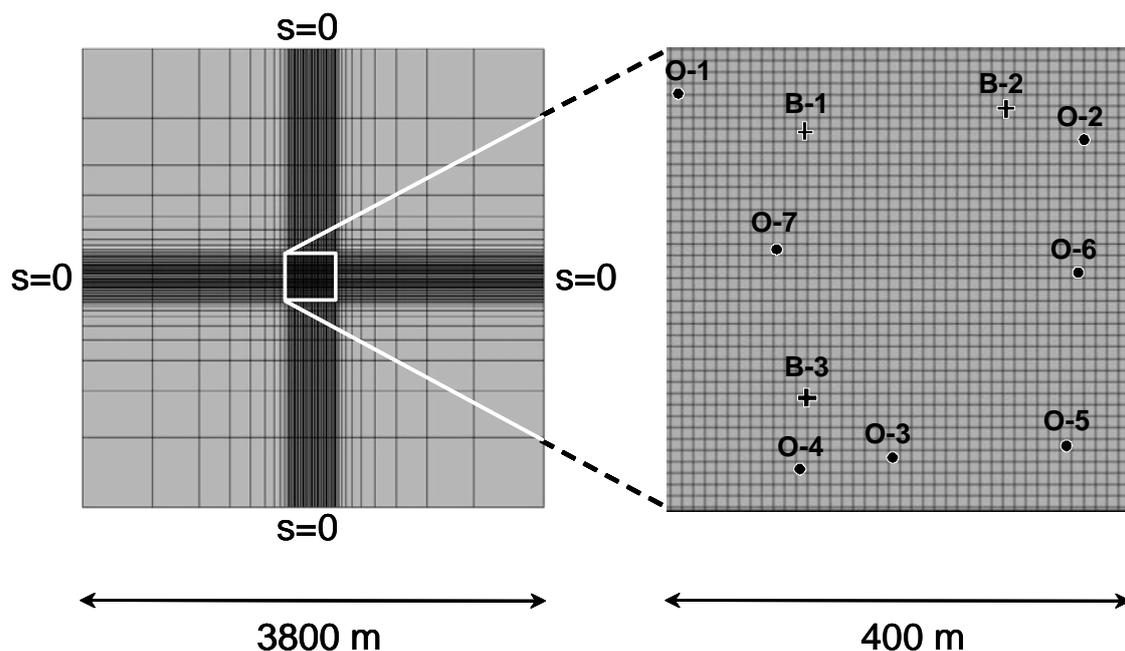


Figure 1. Flow domain and location of conditioning measurements. The inset bounds the zone of interest (model domain for transport prediction).

4.2 Generation of the “true” conductivity fields

First, we select 10 measurement locations at the central part of the domain and fix a value of Y at those points (common for all four tests). Then, we generate four “true” Y fields (Figure 2) by sequential simulation using the code TRANSIN (Deutsch and Journel, 1992; Alcolea et al, 2006c), conditioned to the ten measurements and the geostatistical models presented in Table 1. The models are stationary with a variance of 2. Their spatial variability is simulated by two nested spherical variograms of 40 and 200 m range, representing the small and large scales of variability, respectively. The four structures differ on the weight assigned to the short range variogram, ranging from none (NH) to 75% (HH). The procedure is such that the measurements are slightly biased with respect to the “true” fields.

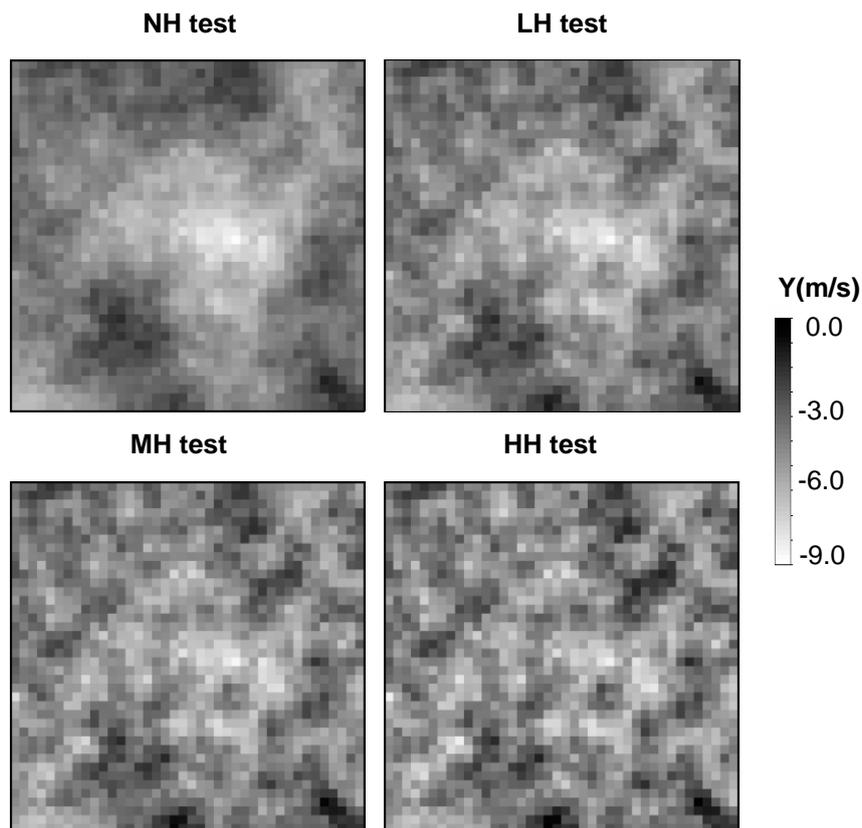


Figure 2. Zone of interest of the “true” Y fields. Notice that the large scale trends of NH (low Y in the middle, high Y in the lower two corners, etc) are reproduced in all realizations. Still, small scale variability, as reflected by the granularity of the fields, increases from NH to HH.

Table 1. Statistical parameters of nested structures defining variograms used for the generation of the four “true” log-conductivity fields.

Synthetic test	Nested structure defining				
	Small scale variability	Small scale heterogeneity		Large scale heterogeneity	
		Range (m)	Sill ($\log_{10}[\text{m/s}]^2$)	Range (m)	Sill ($\log_{10}[\text{m/s}]^2$)
NH	None	---	---	200	2
LH	Low	40	0.5	200	1.5
MH	Medium	40	1.0	200	1.0
HH	High	40	1.5	200	0.5

4.3 Generate drawdown data

Drawdown data comes from three independent pumping tests in the most productive wells of the central part (pumping rates of 10^{-3} m³/s at wells B1, B2 and B3 in Figure 1). “Real” steady-state drawdowns were simulated at grid nodes (Figure 3a) using the four “true” Y fields and prescribing a zero drawdown as initial conditional and at the boundaries. Drawdowns were calculated at the ten measurement locations of Figure 1. A Gaussian white noise was added to those data to simulate measurement errors, with a standard deviation of 0.25 m for the pumping test at well B1 and 0.15 m for the pumping tests at wells B2 and B3 (1% of the maximum drawdown at each test).

4.4 Calibration of the Y fields using the regularized pilot points method

Conditional estimation and twenty conditional simulations of the Y field are obtained for each test (a total of 84 calibrations), by conditioning the model to the available Y and drawdown measurements, as well as their initial drifts (conditioned to the geostatistical model and Y measurements only). The geostatistical models are considered known and error-free (Table 1). Regarding pilot points, they are arranged on a regular network of 81 points within the zone of interest (4.5 pilot points per correlation range in each direction). Sixteen additional points are located beyond the zone of interest (coarse discretization at Figure 1). This number of pilot points is larger

than traditionally used in the pilot points method (de Marsily et al, 1984; Gomez-Hernandez et al, 1997). However, the use of a plausibility term adds stability to the formulation of the inverse problem and therefore, permits us to use as many pilot points as computationally feasible (Alcolea et al, 2006a).

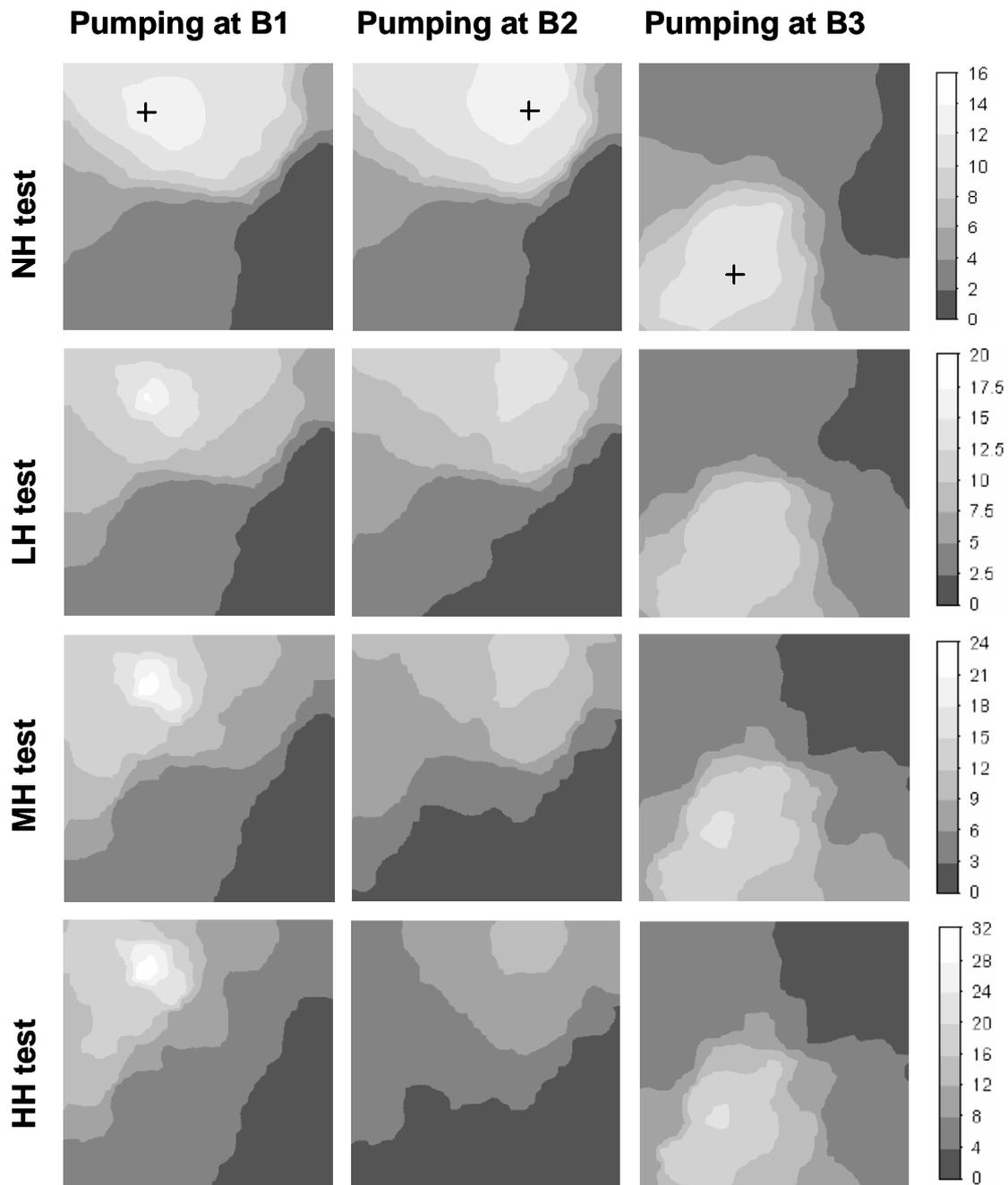


Figure 3. “True” steady state drawdowns (in meters) at the zone of interest. Crosses at pictures in row 1 depict the location of pumping wells.

Following the methodology of Medina and Carrera (2003) and Alcolea et al (2006b), the a posteriori statistical analysis (step 3 in the inversion methodology) was performed for each realization of the Y field, finding the optimum weights for both conditional estimation and simulations. These weights were always 1.0, both for the term of state variables and for the plausibility term.

4.5 Application of the calibrated Y-fields to a transport prediction

The transport problem used for predictions consists of the instantaneous invasion of a solute in the zone of interest. To this end, we first simulate a steady state flow field in the finely gridded central domain (Figure 1) by imposing no flow at the left and right segments and a head gradient of 4/400 between the lower and upper segments. The solute enters the domain with a concentration of 100 ppm through the lower segment. Transport parameters are constant, with longitudinal and transversal dispersivities of 8 and 1 m, respectively, a 0.1 porosity and a 0.01 m saturated thickness. Concentrations are displayed in Figure 4. Integrated flux averaged BTC at the upper segment will be used for evaluating the transport predictions.

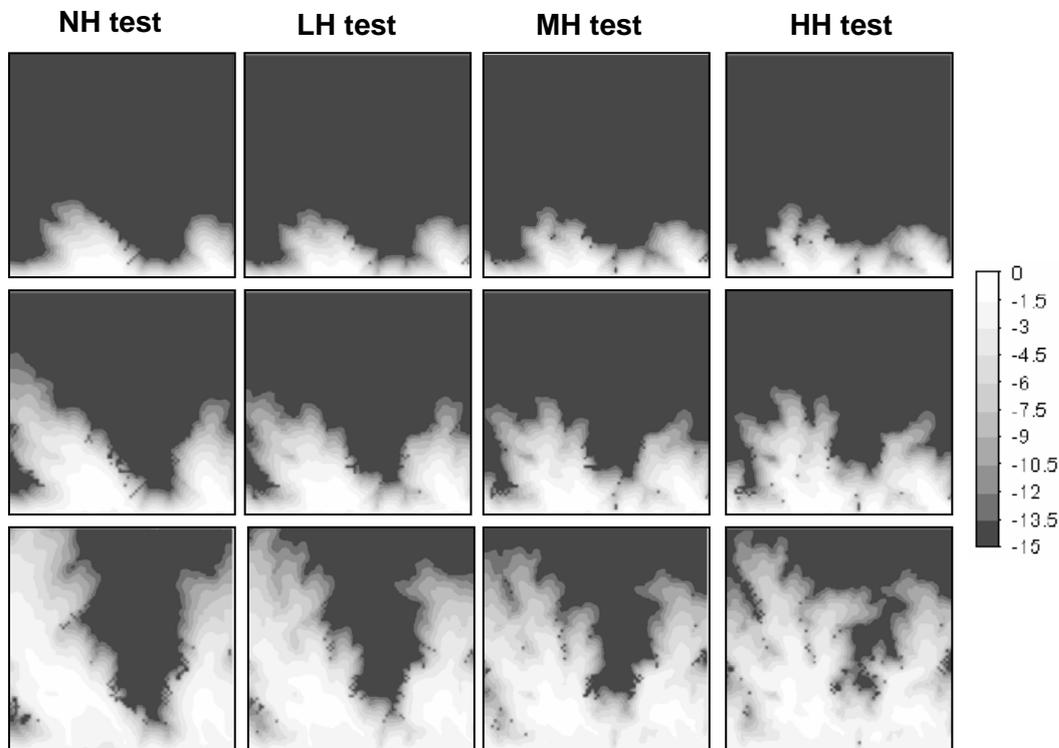


Figure 4. “True” \log_{10} concentrations after 9900 seconds (row 1), 39000 seconds (row 2) and $1.4 \cdot 10^5$ seconds (row 3).

5. Results

Results are evaluated both qualitatively (Y maps and histograms and concentration fits) and quantitatively. Quantitative analysis is performed using an error vector \mathbf{e} , defined as the difference between calculated and “true” values of the Y field at each block of the zone of interest (1600 blocks of $10 \times 10 \text{ m}^2$). We use the following statistics for analyzing results from the estimation process:

- 1) Total objective function (F in equation 1).
- 2) Mean absolute error of the Y field:

$$e_Y = \frac{1}{1600} \sum_{i=1}^{1600} |e_i| = \frac{1}{1600} \sum_{i=1}^{1600} |Y^{\text{true}} - Y^{\text{calc}}| \quad (2)$$

- 3) Root mean square error of the Y field (RMSE_Y) and the analogous magnitudes for drawdowns and concentrations (RMSE_s and RMSE_c , respectively)

$$\text{RMSE}_Y = \left(\frac{1}{1600} \sum_{i=1}^{1600} e_i^2 \right)^{1/2} \quad (3)$$

We test the performance of the calibrated fields for predicting transport BTCs, using the peak concentration and arrival time as well as the slope of the tail. The latter is obtained through regression of late time concentrations. The late time portion of the BTC often displays a “bumpy” aspect due to the development of preferential flow paths. As a result, the definition of this slope is somewhat arbitrary. We chose as “late time” the portion with concentrations three orders of magnitude below the peak.

Calibration results are presented first. The Y fields and estimation errors are presented in Figure 5. The histograms of the Y values and the corresponding statistical moments are depicted in Figure 6 and Table 2, respectively. The total objective function is depicted in Figure 7 and the matching to drawdown data is summarized in Table 3. Regarding the transport prediction, the BTCs obtained with the calibrated Y fields are presented in Figure 8. The corresponding flow and transport mass balances are depicted at Figure 9. Peak concentrations, arrival times and the slope characterizing the late time behavior of the BTC are summarized in Table 4.

The first observation that becomes apparent from Figure 5 is the strong effect of the conditioning to drawdown data. For any given test and realization (CE or CS), the field conditioned only to Y measurements is qualitatively worse than the one obtained by conditioning to drawdown and Y data. Therefore, calibration to drawdown data using the optimum weight of the plausibility term reduces estimation errors. The reduction is not dramatic in terms of estimation errors (both e_Y and $RMSE_Y$ are reduced by about 30%). Conditioning to drawdowns improves the overall look of the computed field (it resembles vaguely the “true” field), but not so much the actual point values. In fact, in terms of point estimation errors, conditional estimation is consistently better than conditional simulation.

The main difference between conditional estimation and simulation stems from the variability, which is best analyzed through the histograms (Figure 6). Histograms of the conditional estimation to Y measurements, $CE(Y)$, are almost symmetric around the mean value of the Y measurements (Figure 6), as measured by the small skewness of the distribution (Table 2). This effect becomes increasingly apparent with increasing relevance of small scale variability, as measured by the variances (decrease), the skewness (tend to zero) and the kurtosis coefficients (increase) at Table 2. Thus, histograms of $CE(Y)$ (for any given relevance of the small scale variability) are centered on the mean and are sharper as the relevance of the small scale variability increases. The same effect, though less notable, can be observed in the fields characterized by conditional estimation to Y and drawdown data, $CE(Y, s)$. However, these outcomes are better as these fields are biased towards the mean value of the corresponding “true” field. In summary, conditional estimation to Y measurements only leads to fields that are too homogeneous (narrow histogram) and centered around direct measurements. Adding drawdowns to conditioning data broadens the histogram and displaces it towards the “true” mean, but not sufficiently.

On the other hand, simulations (regardless the type of conditioning data) yield more realistic results (Figure 5), even though their estimation errors are larger than the ones obtained by conditional estimation. Simulations conditioned only to Y data, $CS(Y)$, overestimate conductivities (Figure 6) because measurements present a positive bias. Fortunately, conditioning to drawdown data helps alleviating this problem and the ensembles of simulations conditioned to Y and drawdown data, $CS(Y,s)$, resemble (for

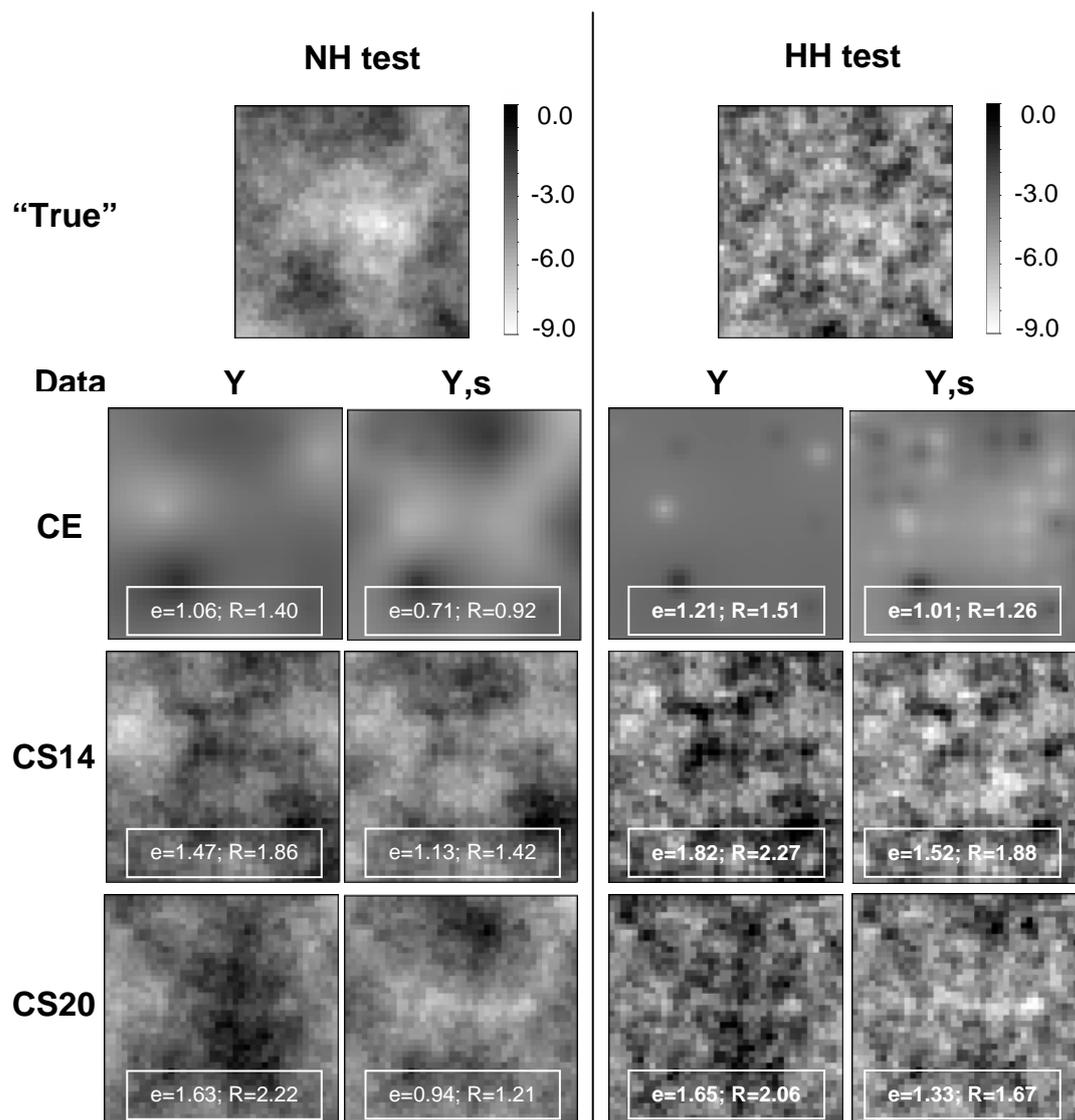


Figure 5. Y fields obtained for the largest relevance of the small scale variability (HH test; subset of pictures at the right of the Figure) and when this variability is neglected (NH test). Two columns are displayed for each subset. Left column displays the Y field conditioned to Y measurements only. Right column displays the Y field conditioned to Y and drawdown measurements. Row 1 contains the “true” Y fields. Row 2 contains the estimated fields and rows 3 and 4 contain the results of two out of twenty simulations.

any given relevance of the small scale variability) the histogram of the “true” fields (Figure 6) and the first two moments of the corresponding distributions (Table 2).

Analyzing the role of small scale variability on calibration is complex. We use drawdown fits ($RMSE_s$ at Table 3) and the total objective function F (Figure 7) to analyze the improvement caused by conditioning. Best results are often obtained when the relevance of small scale variability is negligible (NH test). This case yields the

smallest values of F and $RMSE_s$ in most cases, because the “true” field is smoothest (Figure 2) and, thus, easiest to characterize. However, results do not degrade monotonically with increasing relevance of small scale variability. In any case, $RMSE_s$ are very small for all realizations (even smaller than the standard deviation of measurement errors, which suggests a certain overestimation) and display little dependence on the level of small scale variability. In summary, small scale variability does not control the behavior of steady-state heads, which is ruled by the large scale heterogeneity patterns. This implies that hydraulic data will not suffice for characterizing high frequency fluctuations. Yet, small scale variability becomes important for modeling contaminant transport. The late time behavior of the BTCs depends to a large extent on small scale variability. Regardless the type and number of conditioning data, conditional estimation, which yields smooth Y fields not capable of reproducing the small scale variability, does not match the slope of the late-time portion of BTCs (Figure 8). In fact, the error in the calculated slope increases as the relevance of the small scale variability becomes more important (Table 4). On the contrary, most conditional simulations reproduce this slope, even when only Y measurements are used for conditioning. In this case, breakthroughs are too fast and peak concentrations too high, reflecting the high bias of K measurements. Moreover, simulated BTCs span a very broad range. Yet, all simulations display a tail similar to the “true” BTC.

Table 2. Statistical moments of the distribution of measurements, “true” and calibrated fields (in bold, values of mean and variance which are closest to the “true” ones).

		Mean	Variance	Skewness	Kurtosis
Measurements		-3.94	1.50	0.04	0.69
	True field	-4.60	1.57	-0.32	-0.14
	CE (Y)	-3.90	0.47	-0.36	0.10
NH test	CE (Y,s)	-4.33	0.73	0.29	-0.31
	CS (Y)	-3.95	1.50	-0.05	-0.04
	CS (Y,s)	-4.53	1.52	-0.16	0.16
	True field	-4.65	1.47	-0.11	-0.01
LH test	CE (Y)	-3.93	0.25	-0.28	1.06
	CE (Y,s)	-4.48	0.53	0.18	-0.32
	CS (Y)	-3.99	1.73	-0.04	0.01
	CS (Y,s)	-4.68	1.75	-0.11	0.14
MH test	True field	-4.68	1.55	0.05	-0.04
	CE (Y)	-3.94	0.14	-0.14	3.78
	CE (Y,s)	-4.58	0.42	0.15	-0.10
	CS (Y)	-4.01	2.02	-0.04	0.03
	CS (Y,s)	-4.74	1.88	-0.10	0.07
HH test	True field	-4.71	1.71	0.12	-0.08
	CE (Y)	-3.95	0.07	-0.02	12.87
	CE (Y,s)	-4.63	0.33	0.23	0.32
	CS (Y)	-4.01	2.33	-0.05	0.04
	CS (Y,s)	-4.79	2.07	-0.07	0.04

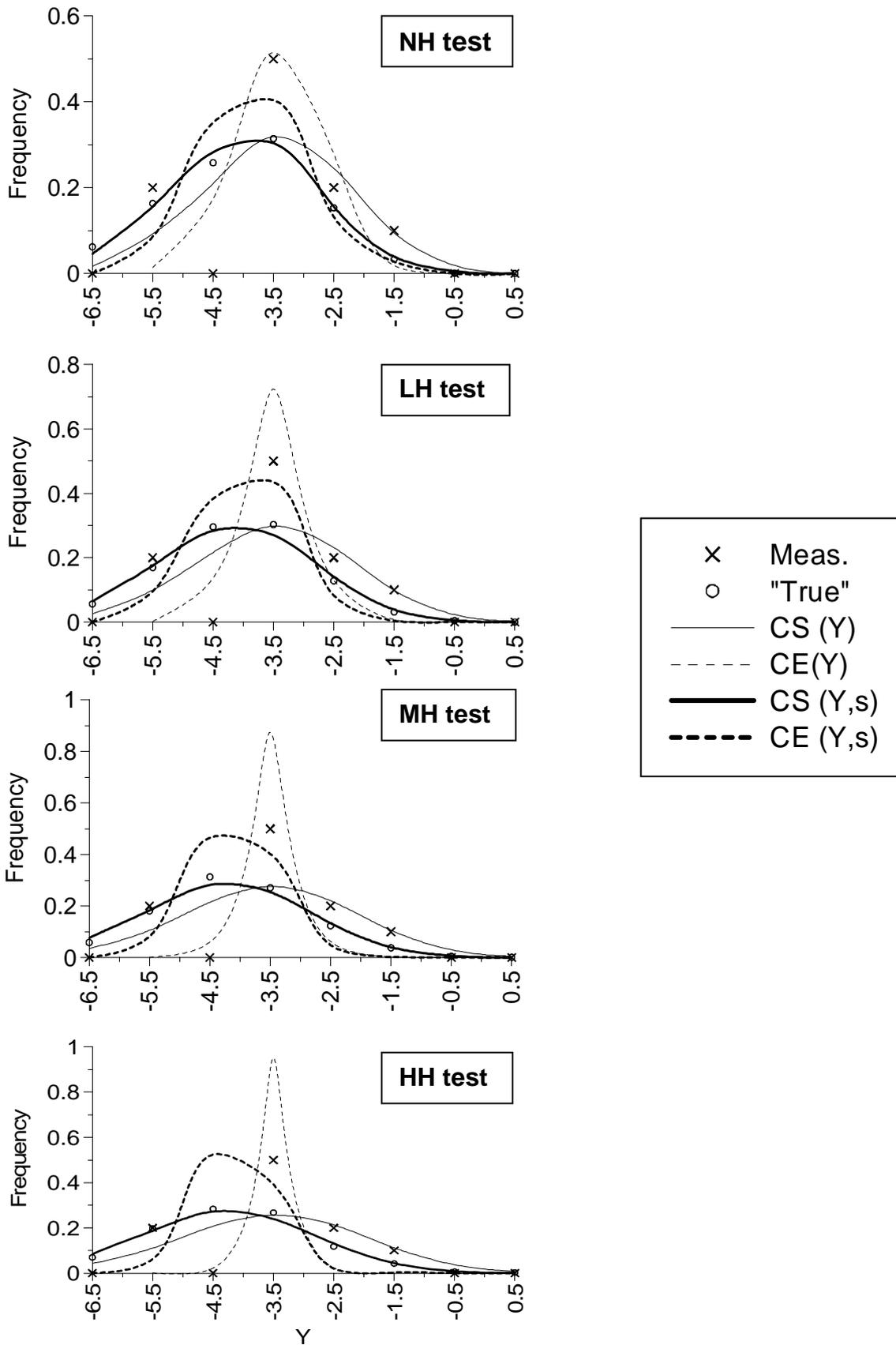


Figure 6. Histogram of Y fields. Results of conditional simulation are the ensemble of 20 realizations.

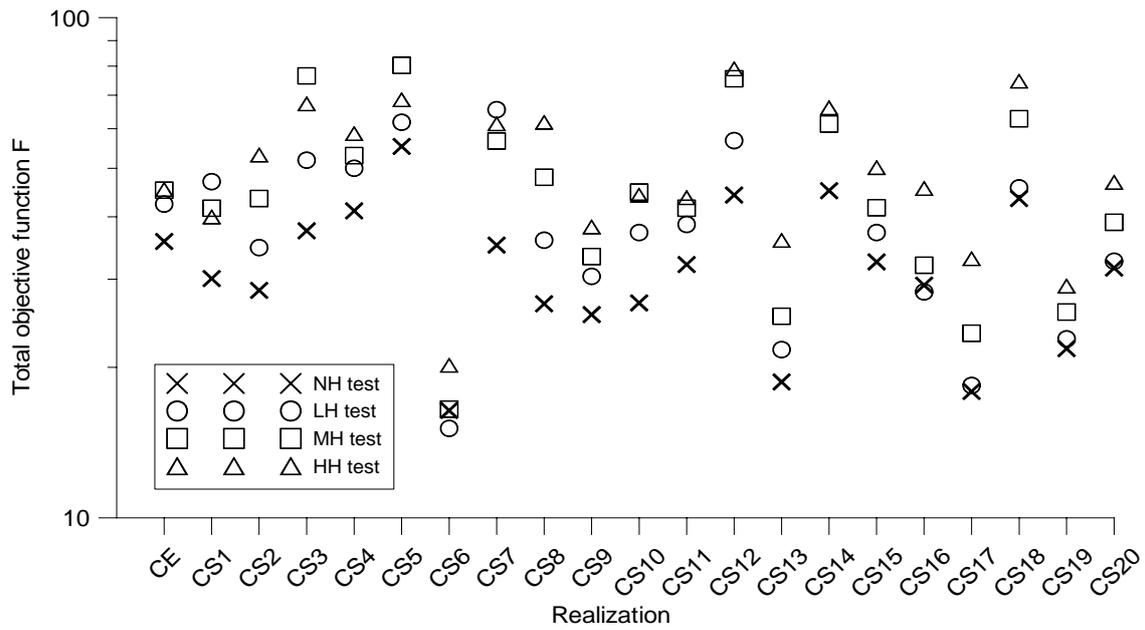


Figure 7. Total objective function for all realizations, when both Y and drawdown measurements are used for conditioning.

The impact of using also drawdowns for conditioning is manifested again in Figure 8 and Table 4. First, errors in peak time, peak concentration, final slope and $RMSE_c$ decrease as expected (both for conditional estimation and simulation). Second, the uncertainty of the predicted BTCs using conditional simulations is reduced substantially, as measured by the statistics listed at Table 4. Qualitatively, one can observe how the true BTCs are bounded by the set of predicted BTCs obtained with the simulated fields, for any given relevance of the small scale variability. This is corroborated by the reduction in the variances of the statistics listed in Table 4. It is also relevant to point out the non monotonic dependence of uncertainty with the degree of small scale variability. Uncertainty, as measured by the breath of the envelope of simulated BTCs, is smallest for the two extreme cases. In the NH case, this reflects that the field is sufficiently smooth to be quite well estimated with hydraulic data. In the case of high local heterogeneity (HH), small uncertainty reflects that the precise details are unimportant. BTCs may be more controlled by the presence of small variability patterns than by their precise details.

Table 3. Root mean square error of drawdowns (RMSE_s) for all realizations. In bold, conditional simulations yielding worse drawdown fits than the corresponding conditional estimation.

	HH test	MH test	LH test	NH test
CE (Y,s)	9.09E-02	8.52E-02	8.27E-02	6.55E-02
CS1 (Y,s)	9.57E-02	1.03E-01	9.32E-02	4.60E-02
CS2 (Y,s)	6.15E-02	6.11E-02	5.94E-02	5.28E-02
CS3 (Y,s)	7.51E-02	1.49E-01	1.11E-01	5.90E-02
CS4 (Y,s)	6.95E-02	6.95E-02	5.87E-02	6.26E-02
CS5 (Y,s)	6.76E-02	1.49E-01	1.15E-01	1.59E-01
CS6 (Y,s)	3.45E-02	3.42E-02	2.48E-02	4.12E-02
CS7 (Y,s)	7.07E-02	8.00E-02	1.55E-01	6.18E-02
CS8 (Y,s)	5.89E-02	6.69E-02	5.30E-02	4.55E-02
CS9 (Y,s)	6.64E-02	6.98E-02	5.23E-02	4.60E-02
CS10 (Y,s)	9.96E-02	1.01E-01	8.40E-02	6.57E-02
CS11 (Y,s)	7.80E-02	6.54E-02	6.57E-02	6.18E-02
CS12 (Y,s)	9.42E-02	9.44E-02	5.87E-02	6.22E-02
CS13 (Y,s)	4.37E-02	3.94E-02	3.71E-02	4.95E-02
CS14 (Y,s)	9.96E-02	9.01E-02	2.76E-01	1.40E-01
CS15 (Y,s)	6.76E-02	4.80E-02	4.42E-02	5.55E-02
CS16 (Y,s)	8.16E-02	5.60E-02	4.26E-02	4.14E-02
CS17 (Y,s)	4.90E-02	3.82E-02	3.47E-02	3.79E-02
CS18 (Y,s)	1.13E-01	6.40E-02	5.37E-02	5.24E-02
CS19 (Y,s)	4.99E-02	4.37E-02	3.61E-02	3.87E-02
CS20 (Y,s)	6.72E-02	6.23E-02	4.77E-02	4.33E-02
Mean CS	7.22E-02	7.43E-02	7.51E-02	6.11E-02
Variance CS	4.18E-04	1.01E-03	3.11E-03	9.52E-04

Flow mass balance is somewhat independent of the relevance of the small scale variability (Figure 9a) and both conditional estimation and conditional simulation methods yield flow mass balances close to the “true” ones (for any given relevance of the small scale variability). This result confirms that large scale patterns of heterogeneity rule the flow behavior. On the contrary, transport mass balance depends to a large extent on, first, the type of conditioning (estimation / simulation), and second, the relevance of small scale variability. Comparing the mass balances of conditional estimation and any of the realizations of conditional simulation, one can see that conditional simulation resembles better the “true” mass balance. The smoothness of the conditional estimation Y fields favors that most solute mass has been washed away by the end of the simulation. As variability increases, increasing portions of mass remain

retained in low permeability areas. This is what explains the huge differences in solute mass in Figure 9b as well as the differences in tailing slope.

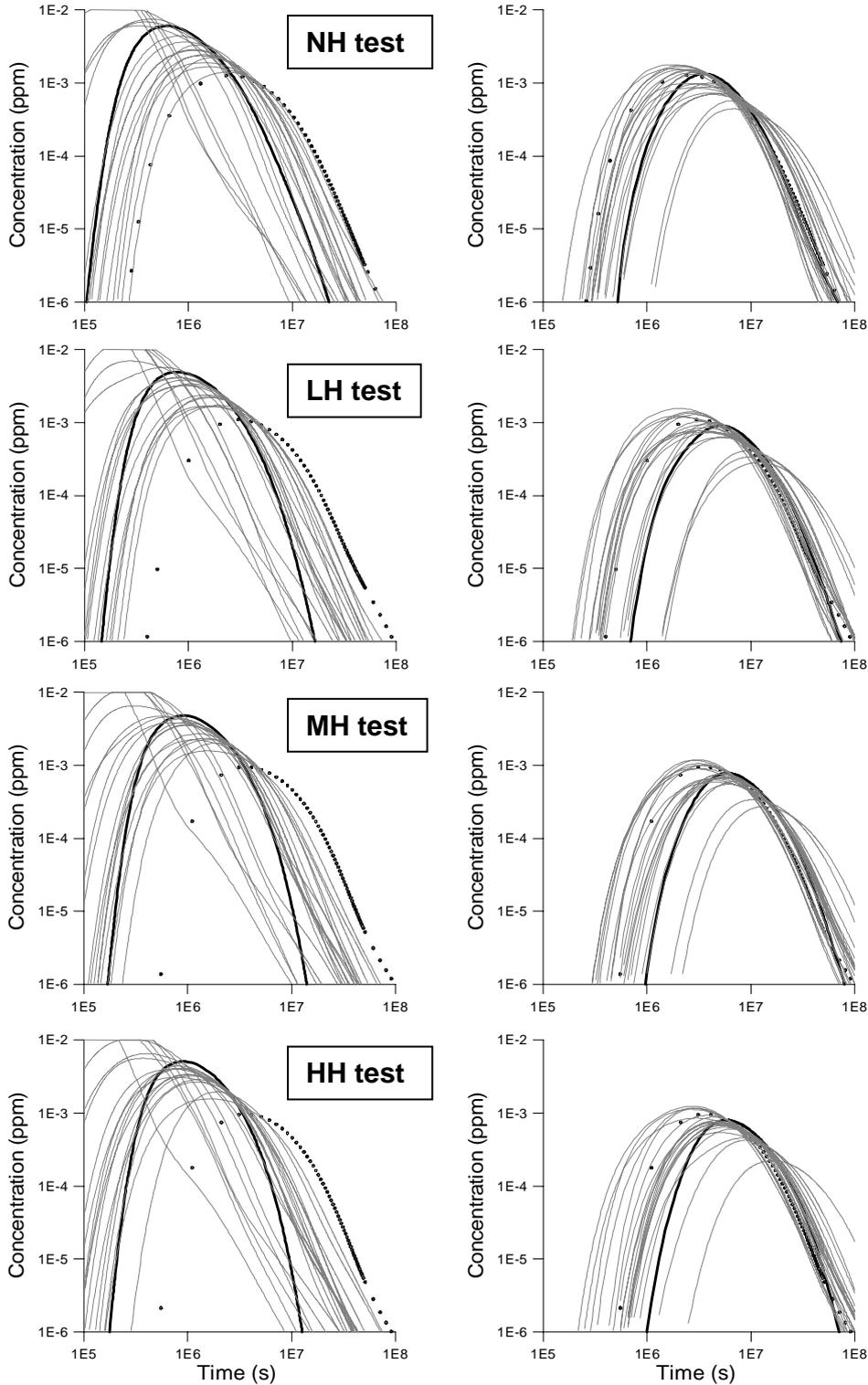


Figure 8. BTCs of the transport prediction for the Y fields conditioned to Y measurements only (left column) and Y and drawdown measurements (right column). “True” BTCs are depicted with dots, the BTC corresponding to the conditional estimation with thick line and the 20 BTCs corresponding to conditional simulation with thin grey line.

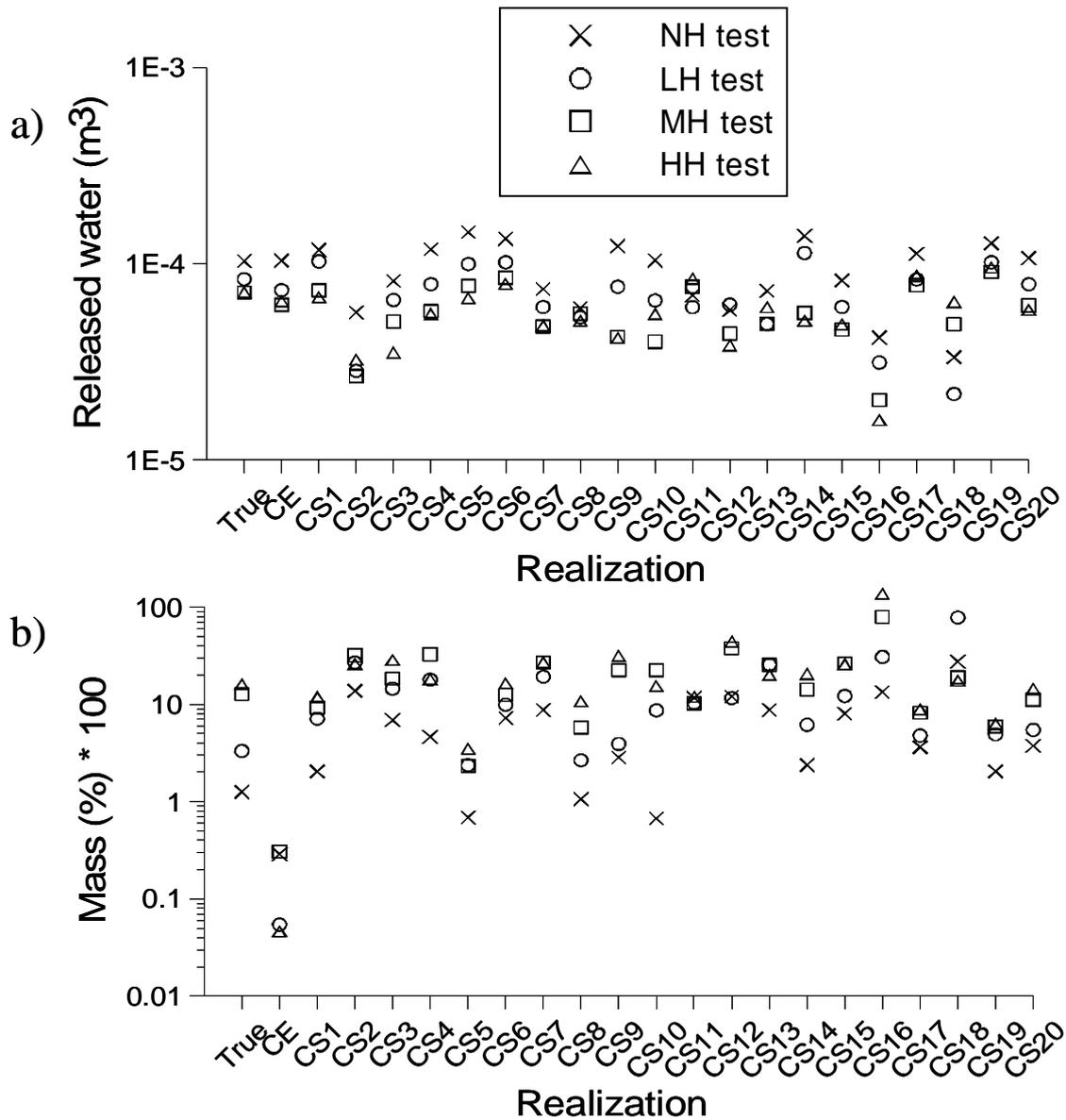


Figure 9. Mass balance of the transport prediction. a) Released volume of water through the upper contour of the zone of interest, b) Tracer mass in the aquifer at the end of the simulation.

6. Conclusions

This work was motivated by the concept that one may be able to identify the large scale (low frequency) trends of spatial variability, but not the high frequency components. Yet, these are relevant for properly simulating and understanding solute transport through heterogeneous media. Accepting that small scale variability cannot be

Table 4. Peak time, peak concentration and late-time slope of the BTCs for the different realizations (mean value and variance for conditional simulations)

		Conditioned to Y				Conditioned to Y, drawdown			
		log10 Peak time	log10 Peak conc.	Slope	RMSE _c	log10 Peak time	log10 Peak conc.	Slope	RMSE _c
NH test	“True”	6.40	-2.90	-2.96	---	6.40	-2.90	-2.96	---
	CE	5.78	-2.22	-7.85	9.34 10 ⁻⁴	6.54	-2.88	-4.16	8.16 10 ⁻⁵
	CS (mean)	5.75	-2.26	-3.27	3.55 10 ⁻³	6.49	-2.96	-2.91	8.88 10⁻⁵
	CS (variance)	0.25	0.24	1.64	2.61 10 ⁻³	0.03	0.03	0.37	2.57 10 ⁻⁹
LH test	“True”	6.48	-2.96	-3.24	---	6.48	-2.96	-3.24	---
	CE	5.90	-2.31	-16.39	6.73 10 ⁻⁴	6.71	-3.04	-5.53	9.21 10 ⁻⁵
	CS (mean)	5.77	-2.28	-3.54	3.35 10 ⁻³	6.59	-3.08	-2.97	8.64 10⁻⁵
	CS (variance)	0.25	0.22	8.61	2.37 10 ⁻³	0.05	0.04	0.93	3.25 10 ⁻⁹
MH test	“True”	6.56	-3.02	-2.44	---	6.56	-3.02	-2.44	---
	CE	5.95	-2.32	-27.78	6.29 10 ⁻⁴	6.77	-3.11	-4.62	9.26 10 ⁻⁵
	CS (mean)	5.75	-2.27	-3.80	3.40 10 ⁻³	6.68	-3.14	-2.76	7.49 10⁻⁵
	CS (variance)	0.27	0.22	28.77	2.70 10 ⁻⁵	0.03	0.02	0.66	2.27 10 ⁻⁹
HH test	“True”	6.56	-3.02	-2.29	---	6.56	-3.02	-2.29	---
	CE	5.95	-2.29	-39.59	6.51 10 ⁻⁴	6.76	-3.10	-5.27	9.24 10 ⁻⁵
	CS (mean)	5.73	-2.25	-4.23	3.55 10 ⁻³	6.68	-3.14	-2.74	7.90 10⁻⁵
	CS (variance)	0.27	0.22	62.43	3.29 10 ⁻⁵	0.04	0.03	0.75	2.25 10 ⁻⁹

identified, we follow on the steps of Gomez-Hernandez et al (1997), Hendricks-Franssen (2001), RamaRao et al (1995). That is, we first simulate fields conditioned to all available direct measurements and conceptual constraints. Here, direct measurements were exact point measurements of log₁₀K and the only conceptual constraint was the assumption that the “true” field was a stationary random field with two nested variograms. The resulting, random, drift is then perturbed so as to ensure that observations (here, drawdowns) are well fitted by the model, using the regularized pilot points method. The question is whether this approach does indeed allow accurate transport simulations. The application leads to the following conclusions:

- 1) Adding a component of small scale variability (i.e. simulating log₁₀K with two nested variograms) leads to increased tailing in transport simulations. The tail slope is much larger than that observed in practice (receding limb too steep). Yet, our results suggest that the slope may be decreased by adding more nested variograms. This lends support to the universal scaling theory of Neuman. This theory was developed to explain scale effects in dispersivity. The fact that it also explains tailing can be viewed as an independent confirmation of the theory.
- 2) Simulated fields reproduce the statistics of the “true” field. This confirms the results of Gomez-Hernandez et al (1997), Hendricks-Franssen (2001), RamaRao et al (1995).

- 3) Simulated fields reproduce the main features of “true” BTCs (arrival time, peak concentration and tail slope). This confirms the main conjecture that motivated this work, namely that one does not need to identify small scale variability, but to simulate its presence.
- 4) When small scale variability is ignored, simulated BTCs reproduce arrival time and peak concentration, but not the tail. This is not critical when the small scale component (NH test) is negligible. In such cases, optimal (smooth) estimation of hydraulic conductivity yields good results. However, in view of the ubiquity of tailing, we feel that that this will be rarely, if ever, the case in practice.
- 5) When calibration is not performed, simulated breakthrough curves reproduce the shape of “true” BTCs (in $\log t - \log c$ scale), but they may be biased if direct measurements are biased, which we fear is often the case.

Much remains to be done. Here we assumed that the structure of variability is known (known variograms). Moreover, the adopted structure is relatively simple (stationary random field defined by only two nested variograms). Dealing with more complex structures will be needed and will require overcoming several difficulties. In this context, the results presented here should be viewed as a hopeful step in the direction of simulating transport through heterogeneous media in a realistic manner.

Acknowledgements

This work was funded by ENRESA (Spanish Agency for Nuclear Waste Disposal) and MEC (Spanish Ministry for Education and Science). Authors gratefully acknowledge X. Sanchez-Vila (UPC) for his help.

References

- Alcolea A, Carrera J, Medina A. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Adv Water Res*, 2006a. DOI: 10.1016/j.advwatres.2005.12.009.
- Alcolea A, Carrera J, Medina A. Inversion of heterogeneous parabolic-type equations using the pilot points method. *International Journal of Numerical Methods for Engineering*, 2006b. DOI: 10.1002/flid.1213.
- Alcolea A, Medina A, Carrera J, Slooten LJ, Hidalgo J, Transin V. *Fortran Code for solving the coupled non linear flow and transport inverse problem*. In preparation.
- 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.
- Carrera J. An overview of uncertainties in modelling groundwater solute transport. *Journal of Contaminant Hydrology*, 1993, 13: 23-48.
- 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.
- Carrera J, Sanchez-Vila X, Benet I, Medina A, Galarza G, Guimera J. On matrix diffusion: formulations, solution methods and qualitative effects. *Hydrogeology Journal* 1998; 6: 178-190.
- Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ. Inverse problem in hydrogeology. *Hydrogeology Journal*, 2005; 13: 206-222.
- Cortis A, Berkowitz B. Anomalous Transport in “Classical” Soil and Sand Columns. *Soil Sci. Soc. Am. J.* 2004, 68:1539–1548.
- Dagan G. Stochastic modelling of groundwater flow by unconditional and conditional probabilities: the inverse problem. *Wat Resour Res* 1985; 21(1): 65-72
- Dagan G. Statistical theory of groundwater flow and transport: Pore to laboratory, laboratory to formation and formation to regional scale. *Wat Resour Res* 1986; 22(9): 120S-134S.
- 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 et al. (ed). *Geostatistics for natural resources characterization. Part 2.*, D. Reidel Pub. Co 1984:831-49.

- Dentz M, Berkowitz B. Transport behaviour of a passive solute in continuous time random walks and multirate mass transfer. *Water Resources Research*, 2003; 39(5), 1111, doi:10.1029/2001WR001163.
- Deutsch CV, Journel AG. *GSLIB, Geostatistical Software Library and User's Guide*. Oxford University Press, New York. 1992.
- Fernandez-Garcia D, Rajaram H, Illangasekare TH. Assessment of the predictive capabilities of stochastic theories in a three-dimensional laboratory test aquifer: Effective hydraulic conductivity and temporal moments of breakthrough curves. *Water Res Res*, 2005; 41, W04002.
- Gelhar LW. Stochastic subsurface hydrology from theory to applications. *Water Resour Res*, 1986; 22(9):135S-145S.
- 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.
- Haggerty R, Gorelick SM. Multi-rate mass transfer for modelling diffusion and surface reactions in media with pore-scale heterogeneity. *Water Resour Res* 1995; 31(10): 2383-2400.
- Haggerty R, McKenna SA, Meigs LC. On the late-time behavior of tracer test breakthrough curves. *Water Resour Res*, 2000; 36(12): 3467-3479.
- Hendricks-Franssen HJ. Inverse stochastic modeling of groundwater flow and mass transport, Ph.D. thesis, Technical University of Valencia, Spain, 2001.
- Hendricks Franssen HJ, Gómez-Hernández JJ, Sahuquillo A. Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data. *Journal of Hydrology*, (2003); 281(4), 281-295.
- 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.
- Hernandez AF, Neuman SP, Guadagnini A, Carrera J, Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media, *Water Resour. Res*, 2006; 42, W05425, doi:10.1029/2005WR004449.
- Hoeksema RJ, Kitanidis PK. Comparison of gaussian conditional mean and kriging estimation in the geostatistical solution to the inverse problem. *Wat Resour Res* 1984; 21(6): 337-350.

- Kennedy CA, Lennox WC. A stochastic interpretation of the tailing effect in solute transport. *Stochast Environ Res Risk Assess*, 2001; 15: 325-340.
- Kitanidis P.K., Vomvoris EG. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one dimensional simulations, *Water Resour Res*, 1983; 19(3): 677-690
- Knudby C, Carrera J. On the relationship between indicators of geostatistical, flow and transport connectivity. *Advances in Water Resources*, 2005; 28: 405-421.
- Kosakowsky G, Berkowitz B, Scher H. Analysis of field observations of tracer transport in a fractured till. *Journal of contaminant Hydrology* 2001; 47: 29-51.
- Kosakowski G. Anomalous transport of colloids and solutes in a shear Zone. *Journal of Contaminant Hydrology* 2004, 72(1-4), 23-46
- Lalemand-Barres P, Peaudecerf P. Recherche des relations entre la valeur de la dispersivité macroscopique d'un milieu aquifère, ses autres caractéristiques et les conditions de mesure- Étude bibliographique. 1978. *Bull. Bur. Rech. Géol. Min., Sér. 2, Sect III,4: 277-284*
- 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.
- 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.
- Margolin G, Berkowitz B. Continuous time random walks revisited: First passage time and spatial distributions. *Physica A* 2004,334:46 – 66.
- Medina A, Carrera J. Geostatistical inversion of coupled problems: dealing with computational burden and different types of data, *J Hydrol* 2003; 281:251–64.
- 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): 10-17.
- Neuman SP. Universal scaling of hydraulic conductivities and dispersivities in geological media. *Water Resour Res*, 1990; 22(8): 1479-1758.
- Neuman SP. On the tensorial nature of advective porosity. *Adv Water Resour*, 2005; 28: 149-159.
- 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.
- Rasmuson A. The influence of particle shape on the dynamics of fixed beds. *Chem Eng Sci*, 1985; 40: 1115-1122.
- Rubin Y, Bellin A, Lawrence AE. On the use of block-effective macrodispersion for numerical simulations of transport in heterogeneous formations. *Water Resour Res*, 2003; 39(9), 1242, doi:10.1029/2002WR001727.
- Sahuquillo A., Capilla J., Gómez-Hernández JJ, Andreu J. Conditional simulation of transmissivity fields honouring piezometric data. In: Blain WR, Cabrera E. *Fluid Flow Modelling, Comput. Mech.*, Billerica, Mass. 1992, pp. 201-212.
- Sanchez-Vila X, Carrera J. Directional effects on convergent flow tracer tests. *Math Geol* 1997; 29(4):551–569.
- Valocchi AJ. Validity of the local equilibrium assumption for modeling sorbing solute transport through homogeneous soil. *Water Resour Res*, 1985; 21(6): 808-820.
- 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.

Appendix. A priori covariance matrix of parameters

The formulation of the regularized pilot points method requires the specification of the a priori covariance matrix of parameters. If kriging is used for defining the drift (step 1 of the inversion methodology), this matrix is the kriging covariance matrix. If simulation is performed, the covariance matrix is calculated as follows.

Let \mathbf{Y} be the vector of true values of the field at the n_e points/blocks to be estimated. Let \mathbf{Y}^K and \mathbf{Y}^{CS} be the vectors of kriged and simulated values, respectively. Let \mathbf{V}_K and \mathbf{V}_{CS} be the kriging error and conditional simulation covariance matrices, respectively. \mathbf{M} is a matrix such that $\mathbf{V}_K = \mathbf{M} \cdot \mathbf{M}^t$ and \mathbf{u} is a vector of independent variables $u_i \sim N(0,1)$.

Conditional simulation can be expressed as:

$$\mathbf{Y}_{CS} = \mathbf{Y}^K + \mathbf{M} \cdot \mathbf{u} \quad [\text{A1}]$$

The m-n-th component of the covariance matrix of CS errors is:

$$(\mathbf{V}_{CS})_{m,n} = \text{cov} [Y_m^{CS} - Y_m, Y_n^{CS} - Y_n] \quad [\text{A2}]$$

Substituting equation [A1] in [A2]:

$$\begin{aligned} (\mathbf{V}_{CS})_{m,n} &= \mathbf{E} \left[\left(Y_m^K - Y_m + \sum_{j=1}^{ne} \mathbf{M}_{mj} \mathbf{u}_j \right) \left(Y_n^K - Y_n + \sum_{k=1}^{ne} \mathbf{M}_{nk} \mathbf{u}_k \right) \right] \\ &= \mathbf{E} \left[(Y_m^K - Y_m)(Y_n^K - Y_n) \right] + \sum_{j=1}^{ne} \sum_{k=1}^{ne} \mathbf{M}_{mj} \mathbf{M}_{nk} \mathbf{E} [\mathbf{u}_j \mathbf{u}_k] \\ &\quad + \sum_{j=1}^{ne} \mathbf{M}_{mj} \mathbf{E} [\mathbf{u}_j (Y_n^K - Y_n)] + \sum_{k=1}^{ne} \mathbf{M}_{nk} \mathbf{E} [\mathbf{u}_k (Y_m^K - Y_m)] \end{aligned} \quad [\text{A3}]$$

The last two terms are zero because \mathbf{u} is independent of \mathbf{Y} . The first term in the right hand side is the definition of kriging errors covariance matrix. Finally, the second term equals $\mathbf{M}\mathbf{M}^t = \mathbf{V}_K$ because the components of \mathbf{u} are independent with unit variance. Therefore,

$$\mathbf{V}_{CS} = 2\mathbf{V}_K \quad [\text{A4}]$$

PAPER IV

**GEOSTATISTICAL INVERSE MODELLING OF A COASTAL AQUIFER
FROM TIDAL RESPONSE AND HYDRAULIC TEST DATA**

Andrés Alcolea, Eduardo Castro, Manuela Barbieri, Jesús Carrera and Sergio A. Bea

1. Abstract

Remediation of contaminated aquifers demands a reliable characterization of hydraulic connectivity patterns. Hydraulic diffusivity is possibly the best indicator of connectivity. It can be derived using the tidal response method (TRM), which is based on fitting observations to a closed-form solution. Unfortunately, the conventional TRM assumes homogeneity. The objective of our work is to overcome this limitation and use tidal response to identify preferential flow paths. Spatial variability is characterized using the regularized pilot points method. The procedure requires joint inversion with pumping test data to resolve diffusivity into transmissivity and storage coefficient. Actual application is complicated by the need to filter tidal effects from the response to pumping and by the need to deal with different types of data, which we addressed using maximum likelihood methods. Application to a contaminated artificial coastal fill leads to flow paths that are consistent with the materials used during construction and to solute transport predictions that compare well with observations. We conclude that tidal responses can be used to identify connectivity patterns.

2. Introduction

The aim of this work is to characterize a contaminated site near the coast in Eastern Spain. Remediation requires identifying preferential flow paths (i.e., connectivity patterns) in the study area. Knudby and Carrera (2005) showed that hydraulic diffusivity, D ($D=T/S$ where T and S are transmissivity and storativity, respectively) is possibly the best indicator of hydraulic connectivity. Therefore, one would expect that characterization methods leading to reliable estimation of diffusivity should also contain valuable information about connectivity of high hydraulic conductivity paths. Point values of effective hydraulic diffusivity, D_{eff} , are easily obtained from the interpretation of tidal response at a borehole (Erskine, 1991; Schultz and Ruppel, 2002; Jhan et al, 2003; Fakir, 2003; Shih and Lin 2004; Trefry and Bekele, 2004). The tidal response method TRM estimates D_{eff} from the amplitude and / or the time lag of the tidal response at an observation borehole (Ferris, 1951; Hvorslev, 1951). This analytical solution assumes a one dimensional flow in a homogeneous and infinite confined

aquifer, which is subjected to a sinusoidal perturbation at its boundary (assumed vertical). Estimated D_{eff} is often validated in parallel with the interpretation of hydraulic tests (Drogue et al, 1984; Millham and Howes, 1995).

Efforts have been devoted to relaxing the TRM hypotheses. Li et al (2002) present an analytical solution in a confined and L-shaped aquifer, extending the solution to leaky aquifers (Li and Jiao, 2001). Anisotropy of hydraulic conductivity is assessed by Pandit et al (1991). One dimensional modelling of tidal propagation in a coastal aquifer with complex heterogeneity is explored by Trefry (1999). The effect of boundary heads representing tidal fluctuations is studied by Wang and Tsay (2001) and Jhan et al (2003), who use a superposition of harmonics.

Estimating D_{eff} by the TRM and complementing it with hydraulic tests suffers a number of shortcomings. First, TRM yields point values of D_{eff} , but does not acknowledge heterogeneity, which may affect the aquifer response to tides. Second, hydraulic test data may not be suitable for standard analysis due to the superposition of pumping and tidal effects (Trefry and Johnston, 1998; Chen and Jiao, 1999). TRM can be used for filtering the tidal effects, but requires a known hydraulic diffusivity. An additional shortcoming is that TRM does not represent real tide accurately (sinus function in the work of Chapuis et al, 2006 or a superposition of harmonics, Wang and Tsay, 2001; Jhan et al, 2003).

We conjecture that geostatistical inversion may allow us to overcome the above drawbacks. First, strict hypotheses of the TRM are relaxed. For instance, complex model geometries and heterogeneities can be accommodated. Second, joint interpretation of tidal response at all boreholes of the observation network should yield the connectivity structure rather than a set of point values of diffusivity. Third, explicit numerical modelling of tidal fluctuation enables us to accommodate observed fluctuations of sea level, which contains both deterministic and random components. Fourth, numerical inversion facilitates explicit incorporation of hydraulic test data. This should lead to not only improve the identification of connectivity patterns (Carrera and Neuman, 1986a and c; Weiss and Smith, 1998), but also allows us to resolve diffusivity into transmissivity and storativity (Carrera and Neuman, 1986b, Rötting et al, 2006).

The objective of this work is to test the above conjecture. We present a procedure to integrate tidal response and injection test data so as to characterize connectivity patterns at coastal aquifers. The procedure is based on the regularized pilot points method (Alcolea et al, 2006a and b). The method is applied to a polluted site that consists largely of an artificial fill. This allows us to compare “as-built” maps with preferential flow paths derived from inversion.

3. Site description

The study area (Figure 1) is located at the edge of an unconfined coastal aquifer. It is made up of an anthropogenic fill lying on top of Quaternary conglomerates (zone 1). The anthropogenic fill consists of several zones (2 through 7 in Figure 1) where different materials were deposited down to 10 m deep. The mean depth of the water table is 5 m. Thus, zones with a saturated thickness of around 5 m, some filled with highly conductive material can be found. These zones become preferential flow paths.

The site was occupied by a factory. A review of the factory construction project reveals an approximate location of zones of anthropogenic heterogeneity (Figure 1):

- Seawater pipelines (zone 2). An underground concrete structure (3.5 m thick) is located on top of the conglomerate base in this zone. This structure contains seawater pipelines and was covered with a gravel fill. While the structure reduces the saturated thickness, which is approximately 1 m in this zone, large transmissivity values are expected.
- Seawater pipelines filling (zone 3). Conglomerates in zones 2 and 3 were dug to accommodate the seawater pipelines. Thereafter, anthropogenic material was used to fill up this excavation. Presumably, this highly conductive material and the elongated shape of this zone render it the most suitable preferential flow path.

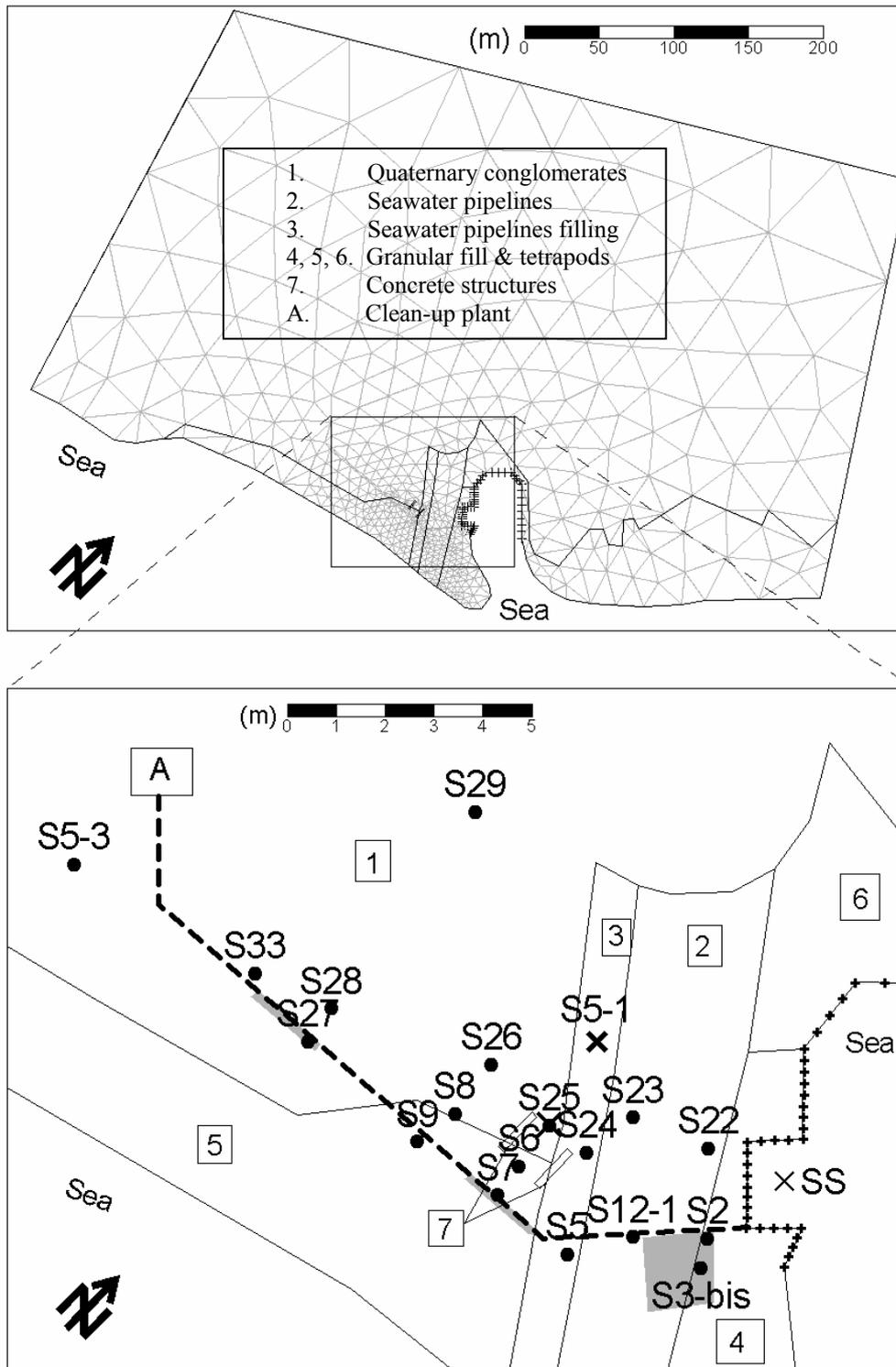


Figure 1. Site description state. The study area (below) lays on top a Quaternary conglomerate (zone 1). Zone 2 accommodates two pipelines for sea water pumping. Zones 4, 5 and 6 represent land gained from the sea, covered with tetrapod marine defences. They cover the sea-shore except in its middle part (depicted by crosses), where a concrete wall panel was built for protection. Zones indicated by '7' are concrete structures that cover almost the whole saturated thickness. Contamination has been detected in the three shaded zones, where the discharge pipeline (dashed line) was presumably broken. Observation boreholes are depicted by black circles.

- Land gained from the sea (zones 4, 5 and 6). The conglomerate base in these areas is covered with granular fill and tetrapod marine defences. These structures protected the factory from the sea. A concrete wall panel protected the factory in the middle part of the shore (depicted by crosses in Figure 1). Presumably, this wall covers almost the whole saturated thickness.
- A discharge pipeline (dashed line in Figure 1) was placed in the non saturated zone of the aquifer. Two concrete structures (zones 7) were built to support this pipeline. The construction project reveals these structures cover the whole saturated thickness.

Prior field-work in the study area detected a contamination in the shaded zones of Figure 1. Presumably, the discharge pipeline was broken in these zones. Although prior studies estimated that the breach occurred in 1975, it was not detected until 1992. Thus, the contaminant spilled and accumulated in the non saturated zone for a long time. The ultimate objective of this work is to characterize the study area to design a remediation system.

4. Methodology

4.1 Tidal fluctuation response

Absolute pressure was automatically recorded at the sea shore (SS in Figure 1) and at twenty boreholes using TD-Diver (Van Essen Instruments, Schlumberger). These measurements were transformed into relative pressures by subtracting barometric pressure (measured at borehole S22 using BaroDiver, Van Essen Instruments, Schlumberger). Next, heads were obtained as the sum of pressure head and diver elevation.

The very high frequency fluctuations of sea-level (i.e., due to wind and waves) were filtered out as they are assumed not to propagate far within the aquifer. To simplify boundary and initial conditions, we express tidal response in terms of variations with respect to natural heads. Thus, we only need to simulate head changes

induced by sea-level fluctuations, but not the regional flow in the aquifer. To this end, head measurements at every borehole were corrected by subtracting their mean value.

Unbiasedness in the calculation of mean head at a borehole requires a long measurement period. This long record was not available at most boreholes because we did not have enough pressure sensors (only boreholes S5, S9 and SS were continuously monitored for 41 days). In addition, measured heads during injection periods were suppressed as tidal and injection effects were superimposed. We used kriging with external drift for filling these gaps (lack of monitoring and injection periods) using the records at boreholes S5 and S9 (not affected by injections) as external drifts. The procedure is outlined in Appendix 1.

4.2 Hydraulic tests

Two injection tests were performed in the area of interest. Relevant data about these tests are summarized in Table 1. Even though injection rates were very high, the observed response to injections at monitored boreholes reached a maximum of only 4 cm. Thus, it was masked by tidal effects (amplitude of tide is ~ 40 cm). Reconstruction of head evolution during injection periods, as described in previous section, allows us to filter the tidal effect. Simply, we subtract the kriged values (in response to tidal effect) from the actual measured heads (Figures 2 and 3).

Table 1. Description of injection tests in the study area.

	Injection type	Injection rate (m ³ /d)	Injection interval (d)	Tot. inj. volume (m ³)	Recovery period (d)	Available measurements
Injection at S5-1	In two steps	199	0 - 0.26	198	0.90 – 1.40	S5, S6, S9 (no response) S12-1, S24, S25, S26, S29
			0.46 - 0.90			
Injection at S25	Continuous	130	0 - 0.5	66	0.5 – 1.05	S5, S7, S8,S9 S26, S29 (no response) S24

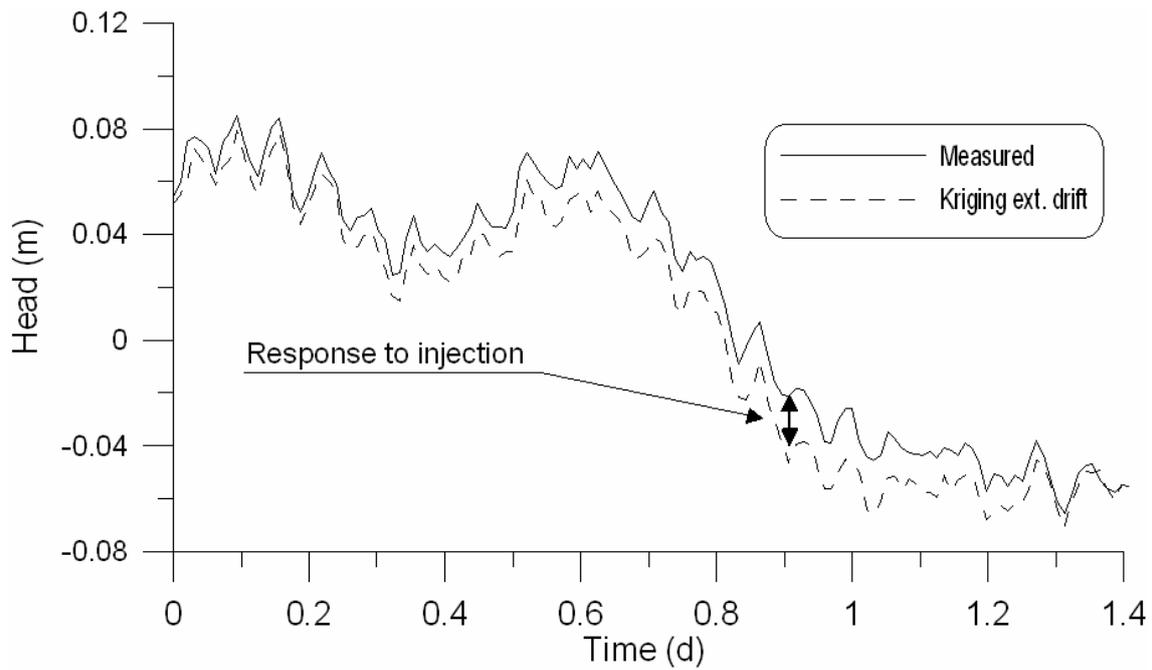


Figure 2. Kriged vs measured heads at borehole S24 during injection test 1 at borehole S5-1.

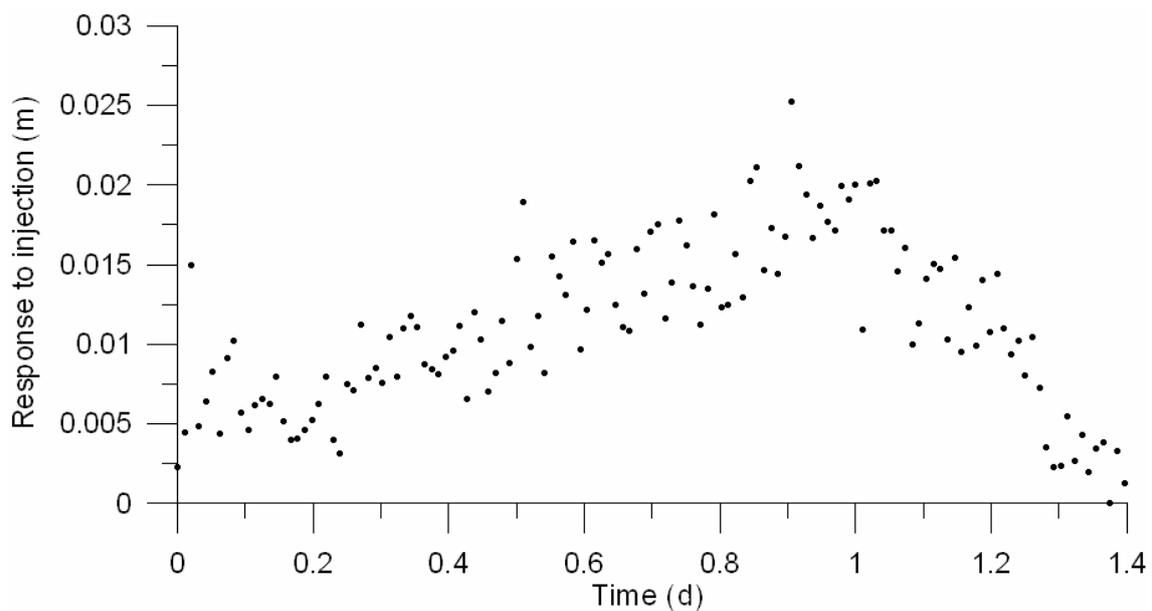


Figure 3. Filtered response to injection test 1 at borehole S24.

A preliminary interpretation of these injection tests was performed using the code EPHEBO (UPC, 2002; Table 2). Estimated storativities were about 0.1 for the Quaternary conglomerates and 0.3 for the anthropogenic fill in most cases. Estimated transmissivities ranged from 15 to 350 m²/d for the conglomerates. Analysis of data at boreholes S12-1 (seawater pipeline) and S24 (seawater pipeline filling) yields transmissivities of 150 m²/d and 225 m²/d, respectively. However, this interpretation

displayed a large uncertainty (i.e., large confidence intervals in the estimation). We attribute this to the fact that all boreholes are partially penetrating (few centimetres in the saturated zone), which may lead to an underestimation of transmissivity values.

Table 2. Available transmissivity measurements in the study area. This information arises from the standard analysis (e.g., homogeneous medium) of the injection tests at boreholes S5-1 and S25.

Zone	Borehole	Transmissivity (m²/d)
	S7	75
	S8	25
Quaternary conglomerate	S25	15
	S28	350
	S29	50
Seawater pipeline	S12-1	150
Tetrapods (zone 5 in Figure 1)	S24	225

3.3 Model calibration

Inversion methodology follows roughly the procedure described by Meier et al (2001) and Rötting et al (2006) to identify preferential flow paths. The main differences stem from (1) the use of the regularized pilot points method (Alcolea et al 2006a), (2) the use of tidal response data and (3) the anthropogenic nature of the site. This allows us to compare the flow paths obtained using only hydraulic data to those revealed by construction records.

The regularized pilot points method parameterizes a hydraulic property (typically $\log_{10}T$) as the sum of a deterministic drift and an unknown residual. The drift is calculated by conditional estimation (ordinary kriging in this case) to available direct measurements, if any, assuming a known correlation structure defined by a variogram. The residual can be viewed as the perturbation needed by the drift to honour measurements of dependent variables (heads, concentrations, etc.). The optimum set of model parameters (value of the hydraulic property at the pilot point locations depicted in Figure 4) minimises an objective function F which accounts for matching measurements of dependent variables and parameter plausibility:

$$F = \sum_{i=1}^{nstat} (\mathbf{u}_i(\mathbf{p}) - \mathbf{u}_i^*)^t \mathbf{V}_{u_i}^{-1} (\mathbf{u}_i(\mathbf{p}) - \mathbf{u}_i^*) + \sum_{j=1}^{ntypar} \mu_j (\mathbf{p}_j - \mathbf{p}_j^*)^t \mathbf{V}_{p_j}^{-1} (\mathbf{p}_j - \mathbf{p}_j^*) \quad (1)$$

where “nstat” denotes number of state variables \mathbf{u}_i with available measurements \mathbf{u}_i^* and covariance matrix \mathbf{V}_{u_i} . In this case, two subsets of state variables were used ($i=1$ for tidal responses and $i=2$ for injection test data). We assumed \mathbf{V}_{u_i} to be diagonal ($\mathbf{V}_{u_i} = \sigma_{u_i}^2 \mathbf{I}$, where σ_{u_i} is the unknown standard deviation of the corresponding measurement type and \mathbf{I} is the identity matrix). “ntypar” is the number of types of model parameters \mathbf{p}_j being calibrated (organized in vector \mathbf{p}), with prior information \mathbf{p}_j^* and covariance matrix \mathbf{V}_{p_j} (in this case, $j=1$ for pilot points related to transmissivities, $j=2$ for those of storativities, etc). \mathbf{V}_{p_j} is the kriging error covariance matrix. \mathbf{p}_j^* is calculated in the same way as the deterministic drift. μ_j are weighting scalars correcting errors in the specification of the covariance matrices.

Two model structures are used for representing the relation between dependent variables \mathbf{u} and model parameters \mathbf{p} . They differ in the specification of the geological zonation. First, we neglected such information, so that actual patterns of connectivity (as revealed by the construction project) were ignored. This first model structure is termed “hydraulic information-based model” hereinafter. Second, this information was explicitly stated in the model (“geology-based model”). For the latter, heterogeneity of $\log_{10}T$ field in each zone was defined by an isotropic spherical variogram. The ranges were 50 m and 25 meters for the conglomerates and for the zones of anthropogenic heterogeneity, respectively. Corresponding sills were 1 and 0.5 (i.e., $\log_{10}T$ can vary one order of magnitude – or half an order- within a correlation range). When the geological zonation was ignored, a single transmissivity zone encompassed the whole model domain. In this case, the variogram is spherical with range 50 m and a sill of 1. The few (and uncertain) $\log_{10}T$ available measurements complicated the specification of the aforementioned variograms. This affects the calculations of \mathbf{p}_j^* and the corresponding covariance matrices and will be discussed later. Prior interpretation of injection tests (section 3.2) yielded almost constant values of storativity for the conglomerates and for the anthropogenic fill. Thus, regardless the zonation of transmissivity we assumed storativity to be constant (modelled by a single pilot point), but unknown, in these zones.

Boundary and initial conditions were homogeneous (i.e. zero head variations and fluxes) because we seek head variations. Only the boundary conditions governing the test (i.e., sea level fluctuation for tidal response and flow rates for the injection tests) must be expressed as time functions. All other boundary conditions are zero. The concrete wall in the middle part of the sea shore (dashed line in Figure 1 connecting zones 5 and 6) is modelled by a mixed boundary condition. The leakage coefficient was assumed to be constant and known ($\sim 10^{-7} \text{ d}^{-1}$). This small value led to a negligible flux through the wall panel as it was assumed to cover almost the whole saturated thickness. Boundary conditions for the three tests interpreted are summarized in Table 3. Initial head variations are also zero given that before the start of the test, heads are defined by “natural” conditions of the system. Likewise, areal recharge does not need to be evaluated.

Table 3. Summary of boundary conditions of the flow characterization model.

Boundary	Type	Problem 1. Tidal response	Problem 2. Injection at S5-1	Problem 3. Injection at S25
Left	Presc. flow	$Q=0$	$Q=0$	$Q=0$
Right	Presc. flow	$Q=0$	$Q=0$	$Q=0$
Upper	Presc. flow	$Q=0$	$Q=0$	$Q=0$
Sea shore	Presc. head	$\Delta h = \Delta H_{\text{sea}}$	$\Delta h = 0$	$\Delta h = 0$
Middle part of the sea shore (concrete wall panel)	Mixed	$Q = \alpha(\Delta H_{\text{sea}} - \Delta h)$	$Q = \alpha(0 - \Delta h)$	$Q = \alpha(0 - \Delta h)$
S5-1	Presc. flow	---	$Q = 281 \text{ m}^3/\text{d}$	---
S25	Presc. flow	---	---	$Q = 134 \text{ m}^3/\text{d}$

A finite element mesh of 1039 elements (Figure 1) was used. Element size increased as the mesh progressed outside the area of interest. Forward in time finite differences were used to model temporal behaviour. The time step was 0.01 days (15 minutes). This was chosen equal to the frequency of sampling. Simulation times

spanned the intervals [0, 20.8], [0, 1.4] and [0, 0.9] (units in days) for tidal response and for injections at S5-1 and S25, respectively.

Three sources of information were included as conditioning data. On the one hand, tidal response (section 3.1) and data arising from injection tests (section 3.2) were analysed simultaneously. Available measurements of T and S (section 3.2 and Table 2) were used to calculate prior information of the model parameters. The key point of the inversion methodology is the specification of the statistical parameters in equation (1). The statistical unknowns are (σ_1, σ_2) , the standard deviations of tidal response and injection test data, respectively, and (μ_1, μ_2) , the weights of the plausibility terms of $\log_{10}T$ and $\log_{10}S$, respectively. Optimum values of statistical parameters maximize the expected likelihood of the parameters given the data (Medina and Carrera, 2003).

4.4 Transport prediction

Estimated $\log_{10}T$ fields were validated in the prediction of a transport model. This is aimed at reproducing the movement of the contaminating solute under “natural” steady-state flow conditions. “Natural” flow conditions consist of prescribing no flow along the left and right boundaries (they represent regional streamlines), a prescribed flow along the upper boundary simulating regional flow, and a prescribed head on the sea shore simulating mean sea-level (zero). A mass flux is prescribed along the upper boundary (regional flow times baseline concentration). The contaminating episode was modelled by prescribing mass fluxes in the contaminated areas (shaded zones in Figure 1). These fluxes were the outcomes of a reactive transport model (Bea et al, 2004) simulating the mobilization of the solute from the non saturated zone to the saturated one. Flow and transport boundary conditions and parameters are summarised in Table 4. The geometry of model 1 remains unaltered. However, the finite element mesh was refined to avoid numerical dispersion (16624 elements). The contaminating solute was measured at boreholes S2, S3-bis, S5, S8, S12-1, S23, S24, S25, S26, S27, S28.

Table 4. Summary of boundary conditions and parameters of the transport model. $f(t)$ denotes the time function simulating the mass of solute shifted from the non-saturated zone towards the saturated zone. This arises from a reactive transport model.

Zone	Equation	Type		Value
Left boundary	Flow	Prescribed flow		$Q=0 \text{ m}^3/\text{d}$
Right boundary	Flow	Prescribed flow		$Q=0 \text{ m}^3/\text{d}$
Upper boundary	Flow	Prescribed flow		$Q=2.19 \text{ m}^3/\text{d}$
	Transport	Mass flux		$c=30 \text{ Bq/m}^3$
Lower boundary	Flow	Prescribed head		$H=0 \text{ m}$
Contamin. zones	Flow	Prescribed flow		$Q=0.73 \text{ m}^3/\text{d}$
	Transport	Mass flux		$c=f(t) \text{ Bq/m}^3$
Whole domain	Flow	Areal recharge		$q_r=2.7\text{E-}4 \text{ mm/d}$
	Transport	Mass flux		$c=30 \text{ Bq/m}^3$
Concrete wall	Flow	Leakage		$\alpha=1\text{E-}7 \text{ d}^{-1}; H=0 \text{ m}$
Long. dispersivity	Transport	---		3 m
Trans. dispersivity	Transport	---		0.3 m
Porosity	Transport	---		0.2 m
Saturated thickness	Transport	---		2.5 m
Retard. coeff. (zone 1)	Transport	---		250
Retard. coeff. (zone 2)	Transport	---		80
Retard. coeff. (zone 3)	Transport	---		80
Retard. coeff. (zone 4)	Transport	---		60
Retard. coeff. (zone 5)	Transport	---		100
Retard. coeff. (zone 6)	Transport	---		100

5. Results

Results are evaluated in terms of estimation plausibility and fits of measured state variables (head variations and concentrations). $\log_{10}T$ fields obtained by the hydraulic information-based and geology-based models are depicted in Figure 4. A quantitative comparison of $\log_{10}T$ fields obtained by both models is summarized in Table 5. Fits of measured head variations are presented in Figures 5 (tidal response) and 6 (injection data). Only fits of the hydraulic information-based model are presented given that the ones using the geology-based model are very similar. State variable residuals are summarized in Table 6. Results of transport prediction are displayed in Figure 7.

An important point of the inversion methodology is how to weigh the importance of the different data sets in the calibration. This is controlled by the statistical parameters in Equation 1, which control the contribution of each data set in the calibration process. For instance, if the standard deviation of tidal response data (σ_1)

is large compared to the standard deviation of injection test data (σ_2), the large volume of tidal response data may dominate the objective function F , hiding the information contained in data arising from injections and viceversa. Likewise, assigning large plausibility weights biases the solution towards the deterministic drift which in this case was poorly informative, due to the large uncertainty of available direct measurements. On the contrary, assigning small values leads to the best match of dependent variables, but to an unstable characterization of the unknown properties. Fortunately, setting the regularized pilot points method in a maximum likelihood framework allows us to obtain the optimum values of the statistical parameters. These maximize the expected likelihood of the parameters given the data (Medina and Carrera, 2003). We tested 36 cases with values of $(0.5 \cdot 10^{-3}, 10^{-3}, 2.5 \cdot 10^{-3}, 5 \cdot 10^{-3}, 7.5 \cdot 10^{-3}, 10^{-2})$ m and $(10^1, 10^2, 10^3, 10^4, 10^5, 10^6)$ for σ_1 and μ_1 , respectively, assuming $\mu_1 = \mu_2$ and a value of $5 \cdot 10^{-4}$ m for σ_2 . The latter leads to a contribution of injection test data of 30% of the total objective function. The optimum values of σ_1 and μ_1 obtained for both models were $5 \cdot 10^{-3}$ m and 1000, respectively. This value of μ_1 (and μ_2), the second smallest among the tested set, gives little importance to prior information of parameters (10% of the total objective function), what manifests the large uncertainty of the prior interpretation of injection tests. Thus, we allow large departures of model parameters from their prior information.

Estimated $\log_{10}T$ fields (Figure 4) identify preferential flow paths defining the hydraulic connectivity structure and compare well to those revealed by the “as-built” maps. Unfortunately, this information is rarely available but becomes a valuable tool for the verification of our methodology. As expected, these flow paths are defined by the zones of the seawater pipeline and its filling. The similarity of $\log_{10}T$ fields is an important finding of this work. The hydraulic information-based model is capable of reproducing the “anthropogenic” geological contacts although these were not explicitly accounted for. A quantitative analysis of this similarity was performed. We calculated the mean transmissivity of mesh elements within a zone (Table 5). These values are similar in both models. They differ significantly mainly in zones 2 and 3, representing the recharge pipelines and their accommodation. The geology-based model concentrates the largest transmissivities in the seawater pipeline filling, whereas the hydraulic information-based model extends this preferential flow path to zone 5 (i.e. $\log_{10}T$ is

smaller in the seawater pipeline filling and larger in its vicinity within zone 5). Mean transmissivities are also different in zone 4, where no hydraulic data are available. As regards storativities, these were assumed to be constant though unknown. The estimated values are very similar to the prior information (0.09 vs 0.1 for the conglomerates and 0.25 vs 0.3 for the anthropogenic fill).

Table 5. Summary of average residuals (mean difference between calculated and measured state variables at a given borehole) of the calibration. P1, P2 and P3 denote the flow problems of tidal response and injections at boreholes S5-1 and S25, respectively. For instance, P3-S24 denotes measurements at borehole S24 corresponding to injection at borehole S25.

	Hyd info-based	Geology based
P1-S5	1.30E-02	1.33E-02
P1-S6	1.17E-02	1.38E-02
P1-S7	1.12E-02	1.25E-02
P1-S9	1.12E-02	1.27E-02
P1-S22	1.65E-02	1.59E-02
P1-S24	1.48E-02	1.52E-02
P1-S25	1.06E-02	1.12E-02
P1-S26	1.73E-02	1.86E-02
P1-S27	1.29E-02	1.31E-02
P1-S29	1.30E-02	1.47E-02
P1-S5-3	1.04E-02	1.04E-02
P2-S12-1	4.96E-02	3.77E-02
P2-S24	5.39E-02	5.00E-02
P2-S25	7.06E-02	6.85E-02
P2-S26	4.15E-02	3.62E-02
P2-S29	4.81E-02	4.91E-02
P3-S24	4.97E-02	3.52E-02

Fits of measured hydraulic data were very similar and very satisfactory for both models (Figures 5 and 6). Only the fits obtained using the hydraulic information-based model are presented. Average residuals (mean difference between calculated and measured values at a given observation borehole, Table 6) of tidal response are close to zero in both cases. As expected, the hydraulic information-based model yielded slightly larger average residuals for injections given that the geological zonation was not accounted for explicitly. Surprisingly, residuals of tidal response are smaller.

Table 6. Mean estimated transmissivities (m^2/d) in each anthropogenic zone using the hydraulic information-based and the geology-based models.

Model	Hydr. info-based	Geology-based
Quaternary conglom. (zone 1)	2091	2479
Seawater pipelines (zone 2)	4866	7805
Seawater pipel. Fill. (zone 3)	9876	175797
Tetrapod defences (zone 4)	1437	86
Tetrapod defences (zone 5)	4341	1324
Tetrapod defences (zone 6)	904	1194

In addition, several calibrations of $\log_{10}T$ (assuming storage coefficients to be known) using only tidal response data and prior information were performed (i.e. neglecting injection test data). Those runs (beyond the scope of this paper) yielded $\log_{10}T$ fields similar to those depicted in Figure 4 and excellent fits of measured tidal response data. Thus, tidal response data is by itself a powerful tool for identifying preferential flow paths.

As for transport prediction, estimated $\log_{10}T$ fields compare well to available concentration measurements (not used in the calibration) as displayed in Figure 7. The use of zonation as hard data (geology-based conceptual model) yielded slightly better predictions.

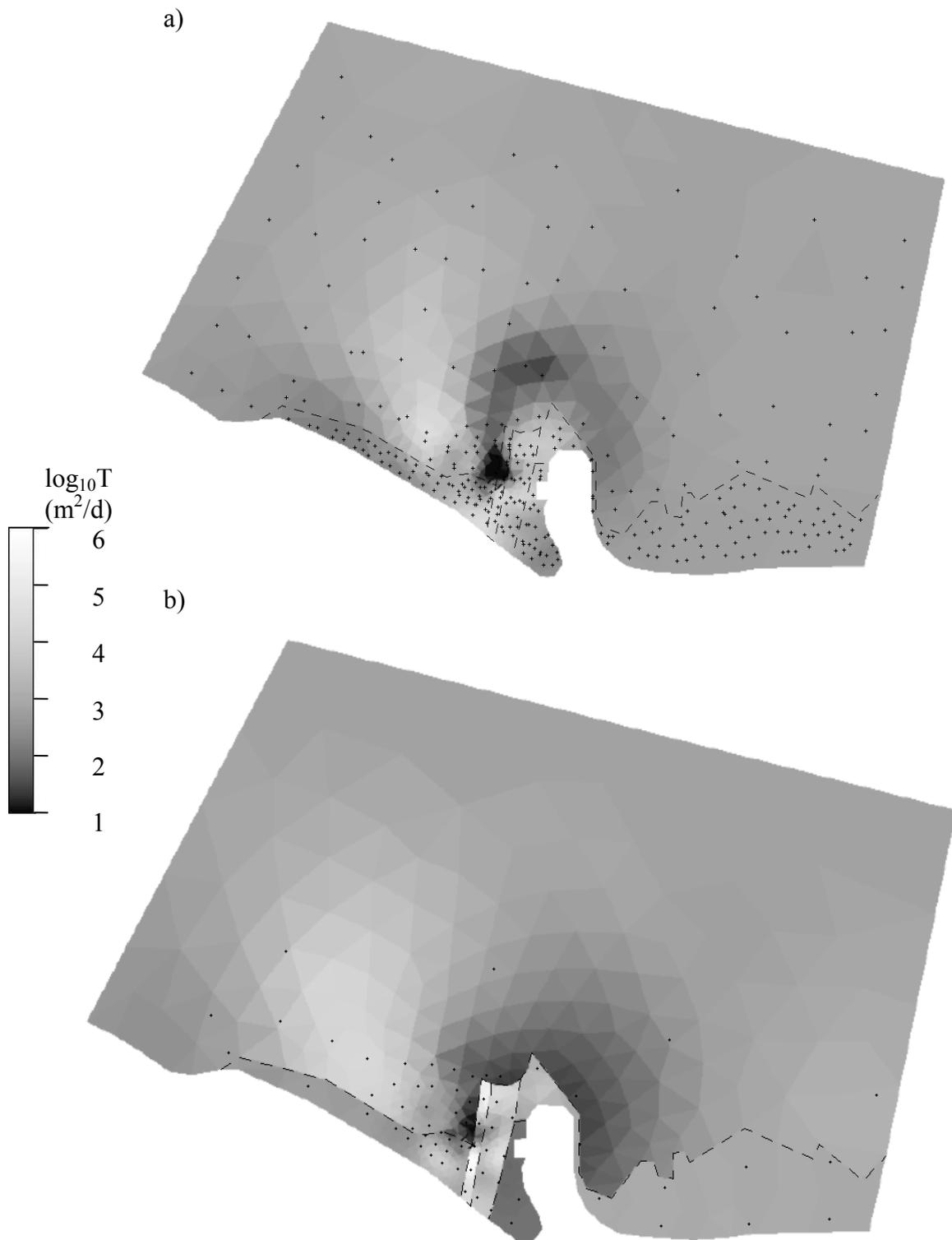


Figure 4. Estimated log-transmissivities using tidal response and injection tests as calibration data. a) hydraulic information-based model (i.e., zonation not accounted for explicitly). b) geology-based model. Dashed lines depict the contact between anthropogenic zones. Pilot points are depicted by dots.

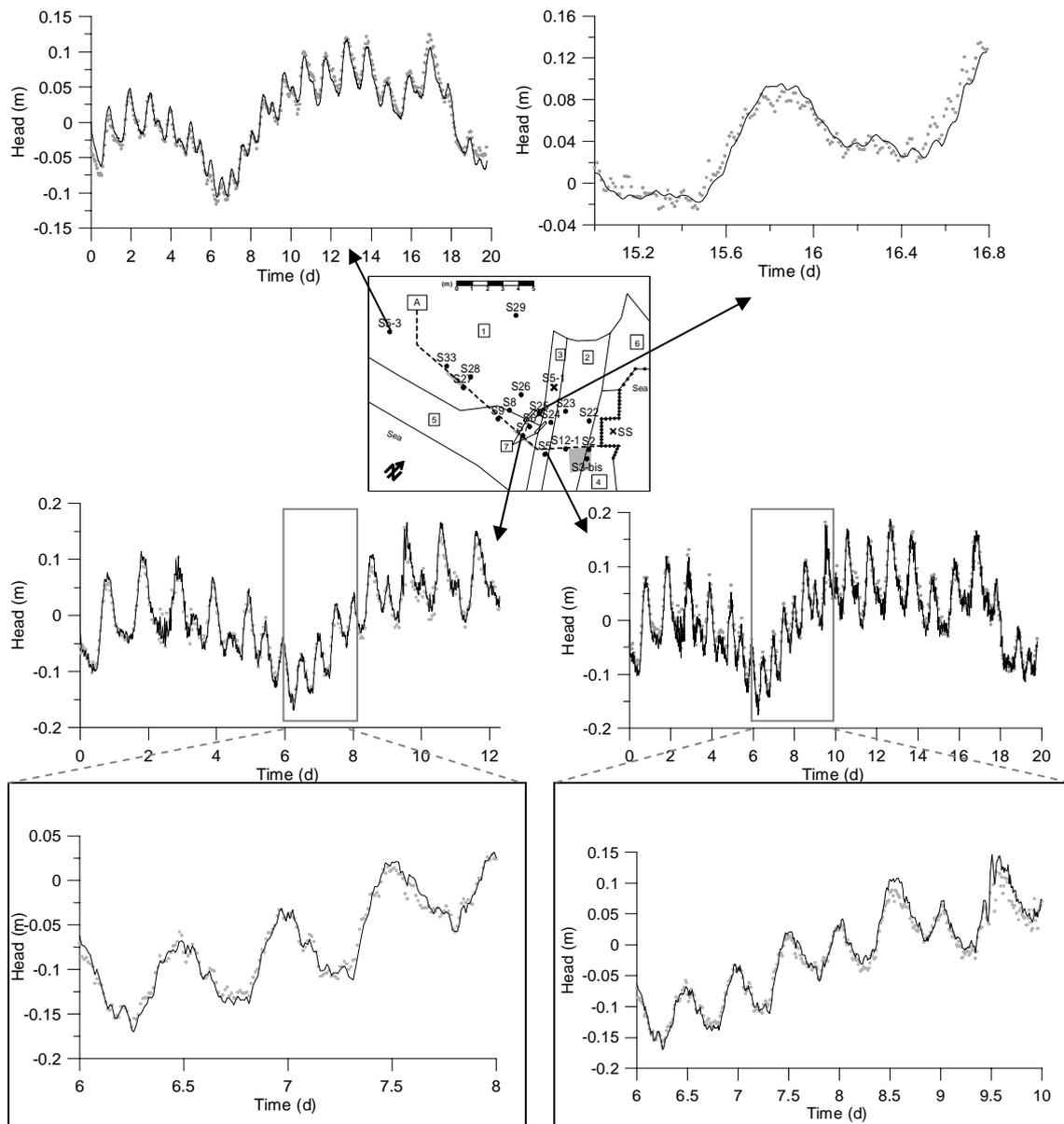


Figure 5. Calculated (line) vs measured (circle) tidal response using $\log_{10}T$ field calibrated by the hydraulic information-based model (Figure 4a). Some measurements at boreholes S5, S7 and S5-3 are not depicted to clarify the figure.

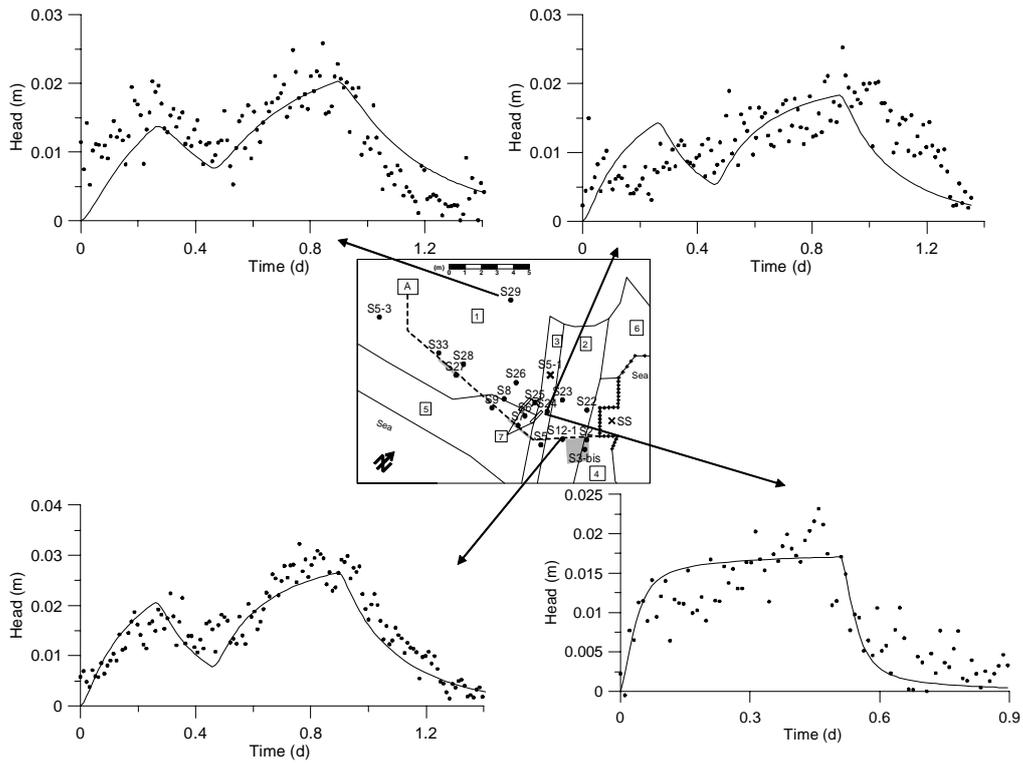


Figure 6. Calculated (line) and measured (circles) response to injections at boreholes S5-1 (boreholes S24 –on top right-, S29 and S12-1) and S25 (borehole S24 on bottom right) using $\log_{10}T$ field calibrated by the hydraulic information-based model (Figure 4a).

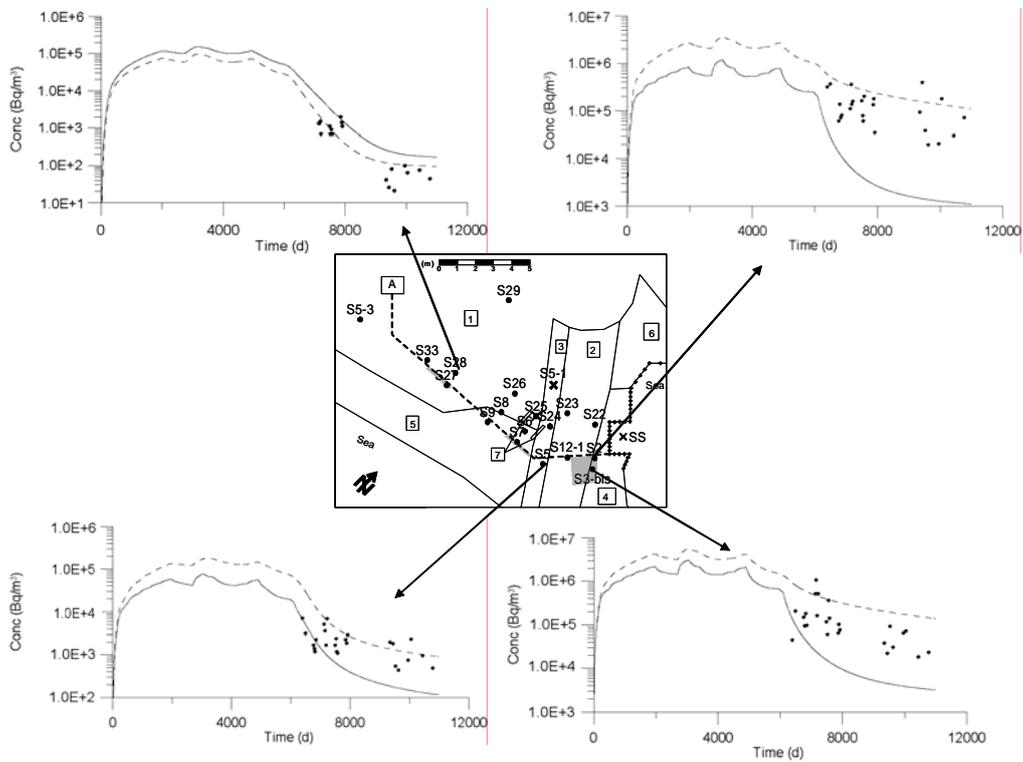


Figure 7. Measured (dots) and calculated concentrations obtained with the geology-based model (continuous line) and the hydraulic information-based (dashed line) models.

6. Conclusions

Tidal response is widely used to obtain point values of hydraulic diffusivity by means of the tidal response method, which assumes homogeneity. The objective of this paper was to overcome this limitation and use the tidal response to identify preferential flow paths. To this end, we applied an integrated methodology, which ranges from data gathering and filtering to geostatistical inversion. Spatial variability of transmissivity and storativity is characterized using the regularized pilot points method. The procedure demands the joint calibration of tidal response and injection test data, which allows us to resolve diffusivity into transmissivity and storativity. Application to real field data is complicated by the need to filter tidal effects from the response to injection and by the need to deal with different types of data. The first shortcoming demands the reconstruction of the tidal effect during the injection period. This is achieved by kriging with external drift. Framing the regularized pilot points method in a maximum likelihood context allows us to balance the importance of different types of data in the calibration.

This methodology is applied to a contaminated artificial coastal fill. The construction project revealed the location of a several zones of anthropogenic heterogeneity. In order to validate the methodology, we tested two model structures. First, information about anthropogenic zonation was ignored (“hydraulic-information based model”). We included this information in the second structure, termed “geology-based model”. Results are summarized next:

- (1) The flow paths identified by the hydraulic information-based model are consistent with those revealed by the “as-built” maps. This lends support to the robustness of the methodology.
- (2) The hydraulic information-based model is capable of identifying the anthropogenic zonation, which was accounted for only in the geology-based model. This may be of great help for identifying contacts between different “geological” formations.

- (3) Excellent fits of measured tidal response and injection test data were obtained with both model structures.
- (4) Calculated concentrations of a transport prediction compare well with the observed ones

We conclude that tidal response is a useful and economical tool for identifying preferential flow paths in coastal aquifers and that the presented methodology which includes the regularized pilot points method is, indeed, robust.

Acknowledgments

This work was funded by ENRESA (Spanish Agency for Nuclear Waste Disposal) and MEC (Spanish Ministry for Education and Science).

References

- Alcolea A, Carrera J, Medina A. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Adv Water Res*, 2006a. *In press*. DOI:10.1016/j.advwatres.2005.12.009.
- Alcolea A, Carrera J, Medina A. Inversion of heterogeneous parabolic-type equations using the pilot points method. *International Journal of Numerical Methods for Engineering*, 2006b. *In press*. DOI: 10.1002/flid.1213
- Bea SA, Carrera J, Soler JM, Ayora C, Saaltink M. Simulation of remediation alternatives for a Cs-137 contaminated soil. *RADIOCHIMICA ACTA*, 2004. 92 (9-11): 827-833
- Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions, 1. Maximum likelihood method incorporating prior information, *Water Resour Res* 1986a; 22(2): 199-210.
- Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions, 2. Uniqueness, Stability and Solution Algorithms, *Water Resour Res* 1986b; 22(2):211-227.

- Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions, 3. Applications, *Water Resour Res* 1986c; 22(2):228-242.
- Chapuis RP, Belanger C, Chenaf D. Pumping test in a confined aquifer under tidal influence. *Groundwater*, 2006; 44(2): 300-305.
- Chen C, Jiao JJ. Numerical simulation of pumping tests in multilayer wells with non-Darcian flow in the well-bore. *Ground Water*, 1999; 37(3): 465-474.
- Drogue C, Razack M, Krivic P. Survey of a coastal karstic aquifer by analysis of the effect on the sea-tide: Example of the Kras of Slovenia, Yugoslavia. *Environmental Geologic Water Sciences*, 1984;6(2): 103-109.
- Erskine AD. The effect of tidal fluctuation on a coastal aquifer in the UK. *Ground Water*, 1991; 29(4): 556-562.
- Fakir Y. Hydrodynamic characterization of a Sahelian coastal aquifer using the ocean tide effect (Dridrate Aquifer, Morocco). *Hydrological Sciences Journal*, 2003; 48(3): 441-454.
- Ferris JG. Cyclic fluctuations of water level as a basis for determining aquifer transmissibility. *International Association of Hydrological Sciences*, 1951; Pub. 33, vol. 2, 148-155.
- Hvorslev MJ. Time-lag and soil permeability in groundwater observations. Bulletin 36. US Army Corps of Engineers, 1951, Waterways Experiment Station, Vicksburg, Mississippi.
- Jhan MK, Kamii Y, Chikamori K. On the estimation of phreatic aquifer parameters by the tidal response technique. *Water Resour Management*, 2003; 17(1): 69-88.
- Knudby C, Carrera J. On the relationship between indicators of geostatistical, flow and transport connectivity. *Adv Water Resour*, 2005; 28(4): 405-421.
- Li HL, Jiao JJ. Analytical studies of groundwater-head fluctuation in a coastal confined aquifer overlain by a semi-permeable layer with storage. *Adv Water Resour*, 2001; 24(5): 565-573.
- Li HL, Jiao JJ, Luk M, Cheung KY. Tide-induced groundwater level fluctuation in coastal aquifers bounded by L-shaped coastlines. *Water Resour Res*, 2002; 38(3): 1024.
- Medina A, Carrera J. Geostatistical inversion of coupled problems: dealing with computational burden and different types of data. *J Hydrol* 2003; 281(4):251-64.

- 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): 10-17.
- Millham NP, Howes BL. A comparison of methods to determine K in shallow coastal aquifer. *Ground Water*, 1995; 33(1): 49-57.
- Pandit A, Elkhazen CC, Sivaramapillai SP. Estimation of hydraulic conductivity values in a coastal aquifer. *Ground Water*, 1991; 29(2): 175-180.
- Rötting T, Carrera J, Bolzicco J, Salvany JM. Stream-stage response tests and their joint interpretation with pumping tests. *Ground Water*, 2006; 44 (3): 371-385.
- Schultz G, Ruppel C. Constraints on hydraulic parameters and implications for groundwater flux across the upland-estuary interface. *Journal of Hydrology*, 2002; 260 (1-4): 255-269.
- Shih DCF, Lin GF. Application of spectral analysis to determine hydraulic diffusivity of a sandy aquifer (PingTung County, Taiwan). *Hydrological processes*, 2004; 18(9): 1655-1669.
- Trefry MG, Johnston CD. Pumping test analysis for a tidally forced aquifer. *Ground Water*, 1998; 36(3): 427-433.
- Trefry MG. Periodic forcing in composite aquifers. *Adv Water Resour*, 1999; 22(6): 645-656.
- Trefry MG, Bekele E. Structural characterization of an island aquifer via tidal methods. *Water Resour Res* 2004; 40(1): W01505.
- UPC (2002). Ephebo 1.1 rev 2. Estimación de parámetros hidráulicos mediante ensayos de bombeo. Hydrogeology Group. Dpt. Of geotechnical engineering and geosciences, UPC.
- Wang J, Tsay TK. Tidal effects on groundwater motions. *Transport in Porous Media*, 2001; 43(1): 159-178.
- Weiss R, Smith L. Parameter space methods in joint parameter estimation for groundwater flow models. *Wat Resour Res*, 1998; 34(4):647-661.

APPENDIX 1. Kriging with external drift for reconstructing heads at a borehole

Time evolution of heads at boreholes affected by tidal fluctuations exhibit a marked non-stationary behaviour. These data can be modelled as the sum of a deterministic drift (heads at a reference borehole in this case, $h^{\text{ref}}(t)$) and a stochastic component $\varepsilon(t)$, which is an intrinsic random function with zero mean and known variogram $\gamma_\varepsilon(t)$. Imposing unbiasedness constraints and minimizing the error variance leads to the system of equations of kriging with external drift:

$$\sum_{j=1}^N \gamma_\varepsilon(t_i - t_j) + \mu_1 + \mu_2 h^{\text{ref}}(t_i) = \gamma_\varepsilon(t_i - t) \quad i = 1, \dots, N \quad (\text{A1.1})$$

$$\sum_{j=1}^N \lambda_j = 1 \quad (\text{A1.2})$$

$$\sum_{j=1}^N \lambda_j h^{\text{ref}}(t_j) = h^{\text{ref}}(t) \quad (\text{A1.3})$$

This system is solved for the N kriging weights λ (corresponding to N head measurements) and for μ_1, μ_2 (Lagrange multipliers of the constraints A1.2 and A1.3), at each time where $\varepsilon(t)$ is estimated. Cross validation was performed to select the variogram $\gamma_\varepsilon(t)$ and to test the statistical significance of the estimation. Optimum results were obtained with a monomic model ($\gamma_\varepsilon = kt^\theta$; $k=0.12$; $\theta=0.04$). Mean error (which should be close to zero) was 0.0954 and dimensionless mean quadratic error (which should be close to one) was 1.044. In addition, the selected variogram was validated by estimating heads at reference borehole S5, where all measurements were available. Measured heads at reference borehole S9 were used as external drift (Figure A1).

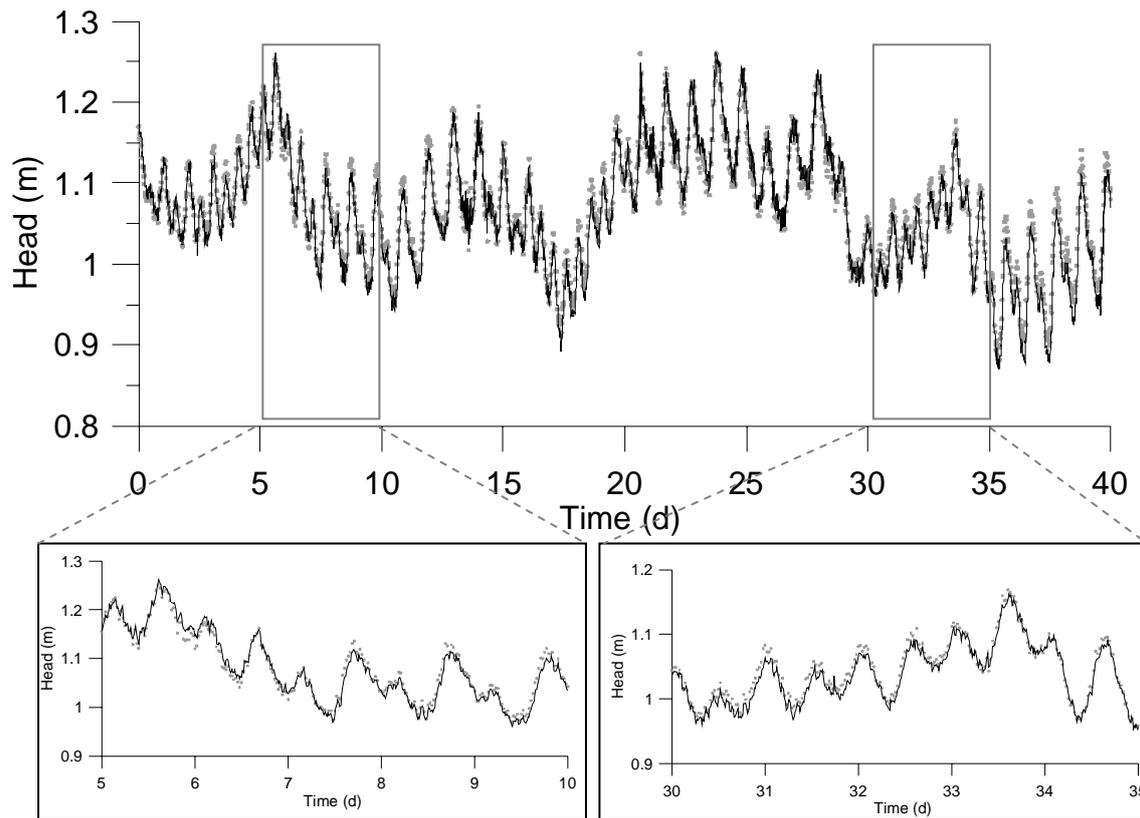


Figure A1. Measured (circles) and reconstructed (line) tidal response at borehole S5 using S9 measurements as external drift.

CONCLUDING COMMENTS

CONCLUDING COMMENTS

This dissertation presents a modification of the pilot points method (PPM) including a plausibility term in the optimization process. This modification allows extending the range of applications of the PPM. The suggested approach has been tested in a number of synthetic examples and in a real case. Conclusions are summarized below.

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

In the first paper, the methodology in its conditional estimation variant has been tested on a synthetic example. Three items are explored: (1) the role of the plausibility term, (2) the sensitivity to the number of pilot points and (3) the effect of coercing the variation of model parameters during the optimization process. Results show that:

- Neglecting the plausibility term, which is the standard approach in the context of pilot points, leads to the best fit of state variable data, but to an unstable identification of 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, giving too much importance to plausibility biases the solution towards prior information. If the geostatistical model contains little information about actual variability patterns (which is often the case), the estimated field yields also a poor identification of heterogeneity. In fact, in most cases, conditioning to state variable data worsens the results if the plausibility term is not properly weighted. The use of a maximum likelihood statistical framework allows the estimation of the optimum weight of the plausibility term.

- Good fits to measured state variables were obtained when neglecting (assigning very low weights to) prior information. Still, nearly as good fits were obtained with stable estimations when moderate (optimum) weights were assigned to prior information.
- A large number of pilot points should be used for obtaining a precise identification of heterogeneity. The use of the plausibility term, which reduces the risk of overparameterization / instability, allows the use of a large number of pilot points (in fact, as large as computationally feasible) and leads to enhanced resolution.
- The inclusion of a coercing factor in the variation of model parameters does not offer any improvement to the identification of heterogeneity. Coercing variations of model parameters only adds computational effort, while the solution remains unaltered.

The main conclusion of this work is that 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.

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

In this paper, we explore the possibility of using a plausibility term in the case of seeking stochastic simulations of the unknown properties conditioned to direct measurements of these properties and dependent variables. Results show that optimum weighting of the plausibility term is necessary. This weight must be calculated for each conditional simulation. Often, a large number of simulations is calculated for evaluating uncertainty. Thus, to search the optimum weight for each simulation can be tedious. However, for each simulation, the optimum weight is the same as the one obtained using conditional estimation. This frees the modeller of the burden of having to seek the optimum weight at each simulation.

PAPER III. Regularized pilot points method for accommodating small scale variability of hydraulic conductivity. Application to simulations of contaminant transport.

The RPPM is framed in the context of the universal scaling theory. The objective of this paper is to test the ability of the RPPM for reproducing the effect of small scale variability of hydraulic conductivity. Heterogeneity of $\log_{10}K$ is simulated using two nested variograms of short and long ranges, representing small scale variability and large scale variability patterns, respectively. Accepting that the high frequency fluctuations cannot be characterized, we aim at evaluating whether including their presence impedes the characterization of the large connectivity patterns defining large scale heterogeneity. In parallel, we explore whether including small scale variability allows reproducing tailing in breakthrough curves. Results show that:

- The simulated fields reproduce the statistics of the “true” field (assumed known).
- Adding a component of small scale variability does not impede the characterization of the large scale connectivity patterns. In addition, it leads to increased tailing in the simulated breakthrough curves. The tail slope is much larger than that observed in practice. Yet, results suggest that the slope may be decreased by adding more nested variograms. This lends support to the universal scaling theory of Neuman.
- Simulated fields reproduce the main features of the “true” breakthrough curves (arrival time, peak concentration and slope of the tail). This confirms the main conjecture that motivated the work, namely that one does not need to identify small scale variability, but to simulate its presence.
- When small scale variability is ignored, simulated BTCs reproduce arrival time and peak concentration, but not the tail. This is not critical when the small scale variability is negligible. In such cases, optimal (smooth) estimation of hydraulic conductivity yields good results. However, in view of the ubiquity of tailing, this will be rarely the case in practice.

PAPER IV. Geostatistical inverse modelling of a coastal aquifer using tidal response and hydraulic tests as calibration data.

The aim of this work is to characterize a contaminated site near the coast. Remediation requires a reliable characterization of preferential flow paths, which are best measured by hydraulic diffusivity. This can be derived from the time lag and/or the amplitude of tidal fluctuations using the tidal response method (TRM). Unfortunately, this method assumed homogeneity. The objective of this work is to overcome this limitation and use tidal response to identify preferential flow paths. The RPPM is used to characterize spatial variability. Hydraulic test data are added, which allows resolving diffusivity into transmissivity and storage coefficient. However, practical application is complicated by the need of filtering tidal effects from the response to injection and by the need to deal with different types of data. An integrated methodology, which ranges from data gathering and filtering to model calibration using the RPPM, is applied. The contaminated site was occupied by a factory. Therefore, valuable information of the construction is available, what allows comparing the identified flow paths with “as-built” maps of the construction project. To this end, two model structures are applied. They differ on whether or not this information is explicitly used for zonation. Results show that:

- The flow paths identified by the hydraulic information-based model (i.e., geological information not accounted for) are consistent with those revealed by “as-built” maps. This lends support to the robustness of the methodology.
- Excellent fits of measured tidal response and injection test data were obtained with both model structures.
- Calculated concentrations of a transport prediction compare well with the observed ones.

