



## A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments

H.J. Hendricks Franssen<sup>a,\*</sup>, A. Alcolea<sup>b</sup>, M. Riva<sup>c</sup>, M. Bakr<sup>d</sup>, N. van der Wiel<sup>e</sup>, F. Stauffer<sup>a</sup>, A. Guadagnini<sup>c</sup>

<sup>a</sup> Institute of Environmental Engineering, ETH Zurich, Wolfgang-Paulistrasse 15, CH-8093 Zurich, Switzerland

<sup>b</sup> Centre for Hydrogeology of Neuchâtel, University of Neuchâtel, Rue Emile Argand 11, CH-2009 Neuchâtel, Switzerland

<sup>c</sup> Dipartimento di Ingegneria Idraulica, Ambientale, Infrastrutture Viarie e Rilevamento (DIIAR), Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy

<sup>d</sup> Deltares (TNO), Soil and Groundwater Systems, P.O. Box 80015, 3508 TA Utrecht, The Netherlands

<sup>e</sup> Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2600 GA, Delft, The Netherlands

### ARTICLE INFO

#### Article history:

Received 4 October 2008

Received in revised form 18 February 2009

Accepted 19 February 2009

Available online 1 March 2009

#### Keywords:

Inverse modelling

Aquifer characterisation

Well catchments

Comparison study

Stochastic simulations

Conditional estimation

### ABSTRACT

Inverse modelling is a key step in groundwater-related hydrological studies. Several inversion techniques were developed during the last decades, but hardly any comparison between them was presented. We compare seven modern inverse methods for groundwater flow: the Regularised Pilot Points Method (both the estimation, *RPPM-CE*, and the Monte Carlo (*MC*) simulation variants, *RPPM-CS*), the *MC* variant of the Representer Method (*RM*), the Sequential Self-Calibration Method (*SSC*), the Moment Equations Method (*MEM*), the Zonation Method (*ZM*) and a non-iterative Semi-Analytical Method (*SAM*). These methods are applied to a two-dimensional synthetic example, depicting steady-state groundwater flow around a pumping well. Their relative performance is assessed in terms of their ability to characterise the log-transmissivity and hydraulic head fields and to predict the extent of the well catchment, both for a mildly and a strongly heterogeneous transmissivity field. The main conclusions drawn from the comparison are: (1) *MC*-based methods (*RPPM-CS*, *SSC* and *RM*) yield very similar results, regardless the degree of heterogeneity and despite they use different parameterisation schemes and objective functions; (2) statistical moments of the target quantities provided by *MEM* and *RPPM-CE* are similar to those of *MC*-based methods; (3) *ZM* and *SAM* are negatively affected by strong heterogeneity; and (4) in general, observed differences between the performances of all methods are not very large. *MC*-based inverse methods need considerably more CPU time than the other tested approaches. An advantage of *MC*-based methods is that they allow computing the posterior probability distribution of the target quantities, which can be directly fed to probabilistic risk-assessment procedures.

© 2009 Elsevier Ltd. All rights reserved.

### 1. Introduction

Inverse modelling is an important and necessary step in hydrogeological studies (e.g., [106]). In broad terms, inverse modelling (also termed history matching, scanning or tomography, amongst other synonyms) refers to the process of gathering information on the model and/or its parameters from (historical) measurements of what is being modelled [31]. Modelling groundwater flow and mass transport in permeable media is affected by different sources of uncertainty. Amongst these, we list: (i) conceptual uncertainties (including model uncertainties and incomplete knowledge of the dominant processes defining the governing equa-

tions of natural phenomena); (ii) measurement uncertainties and (iii) parameter uncertainties. The latter are linked to the spatial variability of hydraulic properties. The literature on groundwater inverse modelling mostly focuses on the estimation of parameters and its underlying uncertainty. One reason for this is the belief of some modellers that parameter uncertainty is the most relevant factor affecting mass transport predictions (e.g., [117]). A second reason is that conceptual uncertainties are difficult to be formalised in a rigorous mathematical framework [109,138]. This and the strong focus on aquifer characterisation, contamination and remediation have stimulated major developments of inverse modelling techniques during the last decades. Yet, hardly any comparisons between either available or newly developed techniques have been published. Existing comparisons by Zimmerman et al. [142] and Keidser and Rosbjerg [77] date to the nineties.

We present a comparison of seven inverse methods for aquifer characterisation. The main focus is on parameter uncertainty. However, it is emphasized that properly addressing conceptual and measurement uncertainties is of high relevance to produce

\* Corresponding author. Tel.: +41 44 6333074.

E-mail addresses: [hendricks@ifu.baug.ethz.ch](mailto:hendricks@ifu.baug.ethz.ch) (H.J. Hendricks Franssen), [andrea.alcolea@unine.ch](mailto:andrea.alcolea@unine.ch) (A. Alcolea), [monica.riva@polimi.it](mailto:monica.riva@polimi.it) (M. Riva), [mahmoud.bakr@tno.nl](mailto:mahmoud.bakr@tno.nl) (M. Bakr), [N.P.A.J.vandeWiel@CTG.TUdelft.NL](mailto:N.P.A.J.vandeWiel@CTG.TUdelft.NL) (N. van der Wiel), [stauffer@ifu.baug.ethz.ch](mailto:stauffer@ifu.baug.ethz.ch) (F. Stauffer), [alberto.guadagnini@polimi.it](mailto:alberto.guadagnini@polimi.it) (A. Guadagnini).

meaningful groundwater modelling efforts (e.g., [100,138]). The paper is organised as follows. First, an overview of modern developments in inverse modelling is presented which highlights the need for our comparison work. Section 2 establishes a general theoretical framework for inverse modelling, within which we cast the inverse methods analysed. Section 3 describes the two synthetic examples used in this comparison. Section 4 introduces the comparison criteria. Sections 5 and 6, respectively, present the results of the comparison study and a discussion about the salient features identified in this work. Finally, some conclusions are given in Section 7.

### 1.1. Literature overview

The large body of literature devoted to inverse modelling in hydrogeology has favoured the publication of numerous state-of-the-art articles [29,31,40,81,93,140]. Thus, this paper is not aimed at establishing a new one. Yet, for the sake of completeness, we present a brief (and therefore non-exhaustive) summary of the literature on inverse modelling.

Incorporating spatial variability of the unknown aquifer properties into a (predictive) forward model is usually accomplished by conditional estimation (e.g., variants of kriging) or simulation (based on available information on such properties [42]). Geostatistical direct methods typically make use of two-points statistical measures (i.e., variograms). This stems from the fact that parameters such as the (natural) logarithm of hydraulic conductivity are often considered as multi-Gaussian random functions. This is supported by statistical analysis of field data (or some suitable transform), which often display a Gaussian behaviour. In addition, multi-Gaussian models are parsimonious (i.e., defined simply by an expected value and a covariance function). Unfortunately, independent univariate Gaussian distributions do not warrant a multi-Gaussian model [54]. A Gaussian model fails to reproduce connectivity features of hydraulic properties, because the spatial continuity of their extreme values is minimal (i.e., the model maximises entropy). Thus, multi-Gaussian models do not allow characterising geological settings presenting curvilinear 'crispy' geometries (e.g., braided channels, river beds) unless their geometry is not very sophisticated and/or the information content of measurements of dependent state variables (e.g., hydraulic heads, concentrations, fluxes) is sufficiently adequate to identify connectivity patterns (e.g., [4,78,96]).

To overcome these limitations, one needs to use an approach that integrates direct stochastic methods capable of accommodating such geometries explicitly. Such methods were reviewed by de Marsily et al. [41]. Examples are the Boolean method [60], the sequential indicator simulation method [51], the truncated plurigaussian method [88], the method of transition probabilities [25] and Multiple-Point (MP) geostatistics [58,73,120]. These works were devoted to the simulation of geological structures accounting only for direct observations of lithology and correlated soft data arising, e.g., from geophysics. MP geostatistics is embedded in the Probability Perturbation Method [20,21,72], a novel technique capable of accommodating heads as conditioning data. Inversion methods that do not rely on the multi-Gaussian assumption include the conditional probabilities method [24] and the gradual deformation method, that is applied to truncated plurigaussian [71] and Boolean models [75]. The Sequential Self-Calibration Method (SSC; [53,114]) was extended to invert lithofacies distributions from state variable data within the framework of truncated Gaussian simulation [136].

Different methods were developed in the context of (multi-Gaussian) inverse modelling to further condition hydrogeological models on dependent state variables. In general, inverse modelling develops according to three main steps. First, a conceptual model

of the aquifer is proposed (conceptualisation). Second, unknown aquifer properties are defined in terms of a number of model parameters, desirably keeping the ability to reproduce spatial variability (parameterisation). Finally, most inverse methods obtain optimum values of model parameters by minimising a suitable objective function (optimisation/parameter calibration). The latter quantifies the difference between model outputs and available information. While the last two steps are clearly different, one can find little distinction in the literature between inverse methods and parameterisation techniques. We use both terms interchangeably in this section. Yet, Section 2 establishes a clear distinction between parameterisation and inversion. Amongst the methodologies devoted to inversion, the Pilot Points Method [38] is a flexible and widely-spread technique [39,85,108,133]. The Pilot Points Method assumes that the spatial distribution of the unknown aquifer property is expressed as the superposition of two fields: a deterministic initial guess, conditioned to the direct measurements of the property (traditionally log-hydraulic conductivity), and an uncertain perturbation, which further conditions the model to dependent state variables (traditionally hydraulic heads). This uncertain residual is a linear combination of model parameters (i.e., the value of the hydraulic property at a set of so-called 'pilot points'). In the original formulation by de Marsily [38], only heads were considered as conditioning data for optimisation. It is now known that this may lead to instability of the inverse problem due to over-parameterisation [35]. Adding a regularisation term to the objective function helps to alleviate this problem [43,82,125,126]. A recent modification by Alcolea et al. [4] included model plausibility in the calibration through a regularization/plausibility term. This penalises large departures of model parameters from their prior estimates. We will refer to this method as *RPPM-CE*. Another geostatistical approach to the inverse problem is the linearised co-kriging method. This was originally formulated by Kitanidis and Vomvoris [79] and further developed by Dagan [37], Yeh et al. [139], Kitanidis [80] and Li [89]. The linearised co-kriging method accounts for uncertainties on variogram parameters. Carrera and Neuman [26,27] proposed Maximum Likelihood estimation using an iterative approach, known popularly as Zonation Method (*ZM*). This approach is able to handle strongly nonlinear problems and allows considering uncertainties of the external forcing terms (e.g., recharge, sink/source terms) by including them as additional model parameters. *ZM* also accommodates conceptual model uncertainties [28], including those of variogram parameters, by considering suitable model discrimination criteria [2,3,61,76,95,110,115,138]. Carrera and Glorioso [30] showed that the linearised co-kriging is a particular case of *ZM* if the latter is stopped after the first iteration. Woodbury and Ulrych [137] proposed a full-Bayesian approach to the groundwater flow inverse problem. The method explicitly incorporates conceptual uncertainties on the variogram model and its parameters. The aforementioned conditional estimation methods [26,27,79,137] are also able to handle measurement uncertainty. These obtain a 'single best' solution conditioned to available measurements and are based on deterministic equations governing flow and transport. The most important drawbacks of a 'single best' solution are two. First, the estimated aquifer property is more smoothed than it is in reality, thus leading to biased contaminant transport predictions [7]. Second, a 'single best' solution does not allow evaluating the inherent uncertainty. The hybrid regularised inversion methodology [127] obtains a 'single best' solution (including uncertainty estimates) that also displays details of small-scale variability. Vermeulen et al. [132] solve the inverse problem in a reduced space, maintaining only a limited number of leading eigenvectors of the covariance matrix (i.e., a small set of meaningful parameters). This technique, also known as singular value decomposition, has been further developed by Christensen and Doherty [33]. A

'single best' estimate, together with an estimate of the lower bound of the posterior covariance matrix of model parameters, allows the statistical sampling of equally likely stochastic realisations [123,124]. However, in the authors' opinion, these stochastic realisations underestimate the true uncertainty (especially for strongly heterogeneous fields) and need not honour conditioning measurements.

Following a different perspective, Monte Carlo (MC) methods were devised to generate equally likely solutions to the inverse problem, termed realisations. Each realisation displays a natural variability reminiscent of that observed in the field and is conditioned to both direct measurements of aquifer properties and dependent state variables. As such, MC methods lean on a stochastic interpretation of the governing deterministic equations. The current trend is to apply MC-based inverse methods to increasingly complex subsurface flow and mass transport models. Amongst these methods, we mention the Sequential Self-Calibration (SSC) Method. SSC was originally proposed by Sahuquillo et al. [114] and Gomez-Hernandez et al. [53] and further extended to transient inversion, including the joint calibration of transmissivity and storage coefficient [62], the three-dimensional inverse modelling of fractured media [64] and the coupled inversion of flow and transport [65]. The Pilot Points Method was reformulated in a MC framework [86] and was modified to handle three-dimensional transient flow problems in the presence of fractures [87]. Recently, Alcolea et al. [5] extended the RPPM-CE to a MC framework and stressed the need to account for optimum model plausibility also in this context. We will here refer to this method as RPPM-CS. The Representer Method (RM) was proposed in the oceanographic literature [14] and later reformulated and applied to hydrogeological problems [130]. Bakr and Butler [11] extended the RM and formulated the method in a MC framework to examine the worth of different types of data for the probabilistic design of well capture zones. More recently, Janssen et al. [74] formulated a strategy to handle non-multiGaussian transmissivity distributions using RM.

The optimisation process involved in the aforementioned traditional inverse methods requires the calculation of the derivatives of the objective function with respect to model parameters (see Section 2). This process is tedious, hard to program and highly CPU consuming, unless adjoint state equations are implemented (e.g., [95]). The Ensemble Kalman Filter (EnKF) [19,46] overcomes this problem, allowing conditioning of multiple equally likely realisations to measurements of state variables. EnKF was introduced in the literature of porous media by Naevdal et al. [97] and was started to be applied to groundwater problems in 2006 [32,44,68,90]. EnKF does not condition stochastic realisations simultaneously to a batch of historic observations, but incorporates state variable measurements sequentially in the iterative process. The updating process is not based on traditional sensitivity-based optimisation (i.e., objective function derivatives), but on optimum weighting of model predictions and measurements. This makes EnKF very efficient. Yet, it is nearly optimal only for normally distributed state variables and parameters. In the same spirit, the Markov chain Monte Carlo (McMC) method is a sampling (as opposed to optimisation) technique that is proved to be a powerful tool for stochastic simulation. Convergence rate of McMC is still a limitation. Several alternatives to speed-up convergence include the use of sensitivity coefficients [104], a Shuffled Complex Evolution Metropolis algorithm [134], or parallel running of optimisation algorithms [135]. A recent development by Fu and Gomez-Hernandez [49] speeds up the convergence by means of a multi-scale blocking scheme.

Other inverse geostatistical approaches rely on a stochastic interpretation of the governing equations. The Moment Equations Method (MEM) is based on the groundwater flow equations which are satisfied by the (ensemble) moments of the state variables

(hydraulic heads and fluxes). Recursive approximations of these (otherwise non-local) moment equations of steady-state hydraulic heads are presented by Guadagnini and Neuman [56,57]. These are embedded within a sensitivity-based geostatistical inverse framework [70] to obtain the first two (ensemble) moments of the posterior probability density functions (pdfs) of transmissivity and hydraulic head. In principle, MEM is also suitable for non-multiGaussian heterogeneous systems. It allows evaluating the functional form of the log-conductivity variogram and its parameters using suitable model discrimination criteria [70].

The methods presented in this Section are mainly focused on parameter estimation and their associated uncertainty. Yet, they can handle other sources of uncertainty (e.g., on source location by Bakr [12] or on the geological scenario by Suzuki et al. [122]). Some approaches can handle uncertainty on external forcing terms by including them as additional model parameters (e.g., recharge rate in the works of Carrera and Neuman [28] and Hendricks Franssen et al. [66,67]). In the literature of surface water hydrology, other approaches consider all these uncertainties in a unified framework (see overviews and discussions in [15,59,91], amongst others). One example is the combination of sequential data assimilation with the Ensemble Kalman Filter and global optimisation [135] or the McMC approach [16,17]. These unified approaches are of great interest to hydrogeological inverse problems but are beyond the scope of this paper.

## 1.2. Motivation and scope of the comparison study

Inverse methods are often applied to synthetic examples when they are first presented in the literature. These examples are often designed to highlight the capabilities of the specific inverse method being tested. How other inverse methods would have performed on the same example remains largely unknown. In fact, one can barely find an exhaustive comparison between inverse methods in the literature, the work by Zimmerman et al. [142] being an important exception. This can be partly explained by the fact that the source codes for the inverse algorithms are, in general, not open to the public. Further to this, inversion is somewhat unpopular among practitioners [109]. Therefore, we believe that there is a need for an extensive comparison amongst available inverse methods that have undergone significant developments during the last decade and are well documented and established in the literature. We compare seven inverse methods, with special emphasis on stochastically based approaches. Three of the methods we analyse are based on a Monte Carlo stochastic interpretation of the groundwater flow equation, namely the Sequential Self-Calibration Method (SSC; [53,63,114]), the Regularised Pilot Points Method in its MC version (RPPM-CS; [5,7]), and the MC variant of the Representer Method (RM; [11,14,129]). Two methods are formulated directly in terms of (ensemble) moments of the groundwater flow equation, i.e., the Moment Equations Method (MEM; [56,57,70]), and a linearised Semi-Analytical Method (SAM; [118,119]) specifically developed for non-uniform flow towards a well. Two conditional estimation methods are also included in the comparison, i.e., the Regularised Pilot Points Method in its estimation variant (RPPM-CE; [4,7]) and the most widely spread (in the authors' opinion) classical method for the inverse modelling of groundwater flow, i.e., the Zonation Method (ZM; [26–28]).

We focus the comparison on the performance of the inverse methods and their predictive capabilities. The influence of the type of numerical discretisation to solve the groundwater flow equation (i.e., finite elements/differences) is also investigated. In order to avoid mixing of concepts, we did not analyse the capabilities of the methods to provide an estimate of the unknown variogram and its parameters. In this spirit, we only consider parameters' uncertainty in this analysis. Thus, the good (or bad) performance

of an inverse method is heavily dependent on the associated inverse methodology. It is worthwhile mentioning that variogram estimation was part of the comparison study presented by Zimmerman et al. [142]. These authors compared the performance of seven inverse methods according to their capability to predict groundwater travel times on four synthetic data sets. These were designed to represent different conceptual models of the transmissivity distribution at the WIPP site [84]. Here, we analyse only two synthetic cases (resembling mildly and strongly heterogeneous transmissivity distributions), but we account for both predictive and characterisation capabilities. In both test cases, we consider 25 (error-free) transmissivity and hydraulic head measurements for conditioning purposes. We then characterise hydraulic head and log-transmissivity fields and predict the extent of the catchment of a pumping well. Characterisation criteria of transmissivity and hydraulic head distributions include: (a) the absolute average error between observed and predicted quantities, (b) the root mean square error, and (c) the average standard deviation of the errors. Prediction criteria include: (a) the mismatch between observed and predicted well catchments (evaluated over the whole domain and over an inner region, far away from the boundaries), and (b) the uncertainty inherent to the predictions (for the methods allowing this type of analysis).

## 2. The seven inverse methods

As pointed out by Carrera et al. [31], most inverse methods do not differ in their essence, but only in the implementation and computational details. Here, we start by outlining a general inversion framework within which the seven tested inverse methods are cast. It includes the following steps: (a) parameterisation, (b) optimisation, and (c) posterior statistical analysis of the results. Then, we outline some operational details of each method.

### 2.1. Parameterisation

A given aquifer property,  $f$ , can be expressed as the superposition of two fields, i.e.,

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

Here,  $\mathbf{x}$  denotes a point in space and  $t$  is time;  $f_D$  represents an initial guess of  $f$  and is conditioned to available information on  $f$  and, possibly, to correlated variables arising, i.e., from geophysics;  $f_p$  can be viewed as the perturbation to  $f_D$  needed for the output of the adopted model to resemble the state variable measurements (e.g., hydraulic heads). For the sake of simplicity, and without loss of generality, it is here assumed that hydraulic properties do not vary in time. The initial guess is often obtained from a (possibly unknown) geostatistical model (e.g., a variogram) and the  $n$  available measurements of the hydraulic property (in the synthetic scenarios presented in Section 3, the hydraulic property is transmissivity), organised in a vector  $\mathbf{f}^*$ . The initial guess can also be defined by criteria that are not geostatistically based, including assimilation of geological information and/or prior experience. Initial guesses based on geostatistical models can either be obtained by conditional estimation or simulation. In the former case, a ‘single best’ solution is obtained. In the latter case, the initial guess represents a random function and many equally likely realisations are generated (i.e., there are several different initial guesses). Classical Zonation Method (ZM) is an example of parameterisation which is not geostatistically based: here, the model domain is partitioned in (deterministically defined) regions, based, e.g., on geological information (aquifer properties) or soil use/meteorological information (recharge rate). In this case, constant (or varying in a prescribed manner) aquifer properties are assigned to each region. Note that

‘geological’ zonation does not preclude the use of geostatistics [34]. In some cases, the spatial variability of the aquifer property is defined within each region using a geostatistical model. The latter can vary from one zone to another. Note also that different aquifer properties can present different ‘zonations’. For the case of linear interpolation, a realization  $k$  of  $f_D$  (which is unique for the conditional estimation case) can be expressed as

$$f_D^k(\mathbf{x}) = \sum_{i=1}^n \lambda_i^k(\mathbf{x}) f_i^* \quad (2)$$

where  $\lambda_i^k$  are interpolation functions weighting the  $n$  measurements,  $f_i^*$ , of the hydraulic property. As such,  $f_D$  honours available data. Correlated information such as geophysics can be accommodated as secondary variables or external drifts. Current implementations of this parameterisation scheme allow the use of a large number of estimation/simulation methods (simple, ordinary or universal kriging, kriging with measurement errors, locally varying mean or external drift, sequential Gaussian simulation and variants of co-kriging/co-simulation when more than one property is estimated/simulated at the same time). The perturbation  $f_p$  can be expressed as a linear combination of the  $n_p$  model parameters, organised in a vector  $\mathbf{p}$ , as

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

where  $\lambda_j^k$  are interpolation (weighting) functions obtained jointly with those in Eq. (2). It is worth mentioning that the interpolation functions in Eqs. (2) and (3) can vary along the iterative process (Section 2.2). This might occur if the locations where aquifer properties are estimated (or simulated) change or if the kriging matrix is updated after each iteration. The latter happens when the posterior covariance of model parameters is accounted for.

### 2.2. Optimisation of model parameters

The formulation of the inverse problem adopted by many methods is based on maximising the likelihood of model parameters given a set of measurements. To this end, we consider a vector of parameters  $\boldsymbol{\beta} = (\mathbf{p}, \boldsymbol{\theta})$  comprising model parameters,  $\mathbf{p}$  (i.e., those defining the unknown aquifer properties), and statistical parameters,  $\boldsymbol{\theta}$  (i.e., those controlling uncertainties associated with aquifer properties). Conditioning measurements are organised in a vector  $\mathbf{z}^* = (\mathbf{u}^*, \mathbf{p}^*)$ , comprising state variable measurements,  $\mathbf{u}^*$ , and prior estimates of model parameters,  $\mathbf{p}^*$ . Within the Maximum Likelihood framework, the likelihood of the parameters  $\boldsymbol{\beta}$  given the data  $\mathbf{z}^*$  can be written as [45]

$$L(\boldsymbol{\beta}|\mathbf{z}^*) = (2\pi)^{-D/2} |\mathbf{C}_z|^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{z} - \mathbf{z}^*)^t \mathbf{C}_z^{-1} (\mathbf{z} - \mathbf{z}^*) \right] \quad (4)$$

Here, superscript  $t$  denotes transpose;  $\mathbf{z}$  is a vector including predictions of state variables and model parameters (organised accordingly to  $\mathbf{z}^*$ );  $\mathbf{C}_z$  is a block diagonal covariance matrix comprising two full matrices,  $\mathbf{C}_u$  and  $\mathbf{C}_p$ , respectively containing the covariances of the measurement errors of the state variables and prior estimates of parameters;  $D$  is the number of data (i.e., the dimension of  $\mathbf{z}$  or  $\mathbf{z}^*$ ). Maximising Eq. (4) is equivalent to minimising the support function  $S = -2 \cdot \log(L)$ , i.e.,

$$S = D \ln(2\pi) + \log |\mathbf{C}_z| + (\mathbf{z} - \mathbf{z}^*)^t \mathbf{C}_z^{-1} (\mathbf{z} - \mathbf{z}^*) \quad (5)$$

Carrera and Neuman [26] proposed an algorithm for maximising the likelihood with respect to the model parameters and some statistical parameters by scaling the covariances  $\mathbf{C}_u$  and  $\mathbf{C}_p$ , defining matrix  $\mathbf{C}_z$ . The remaining statistical parameters (which can include, e.g., variogram type and integral scale) can be estimated a posteriori

by means of suitable model selection criteria (see Section 2.3). Neglecting terms that depend only on statistical parameters, the objective function to be minimised becomes [26]

$$F = \sum_{i=1}^{nstat} \alpha_i F_{s,i} + \sum_{j=1}^{ntypar} \mu_j F_{p,j}$$

$$= \sum_{i=1}^{nstat} \alpha_i (\mathbf{u}_i - \mathbf{u}_i^*)^t \mathbf{V}_{u_i}^{-1} (\mathbf{u} - \mathbf{u}_i^*) + \sum_{j=1}^{ntypar} \mu_j (\mathbf{p}_j - \mathbf{p}_j^*)^t \mathbf{V}_{p_j}^{-1} (\mathbf{p}_j - \mathbf{p}_j^*) \quad (6)$$

where  $nstat$  denotes the number of state variables,  $\mathbf{u}_i$ , with available measurements  $\mathbf{u}_i^*$  ( $\mathbf{u}^* = [\mathbf{u}_1^*, \dots, \mathbf{u}_{nstat}^*]$ ) and covariance matrix  $\mathbf{V}_{u_i}$  (e.g.,  $i = 1$  for heads and  $i = 2$  for concentrations);  $ntypar$  is the number of types of model parameters  $\mathbf{p}_j$ , with prior information  $\mathbf{p}_j^*$  and covariance matrix  $\mathbf{V}_{p_j}$  (e.g.,  $j = 1$  for transmissivity and  $j = 2$  for storage coefficient);  $\alpha_i$  and  $\mu_j$  are weighting scalars correcting the specification of the covariance matrices ( $\mathbf{C}_{u_i} = \alpha_i^{-1} \mathbf{V}_{u_i}$ ;  $\mathbf{C}_{p_j} = \mu_j^{-1} \mathbf{V}_{p_j}$ ), which can also be viewed as factors controlling the relative importance of each data source. Minimising the first term in Eq. (6) warrants a good agreement between calculated and measured state variables. If only this term is considered, the inverse problem might become ill-posed and its solution unstable. The second term in Eq. (6) is a regularisation contribution that penalises large departures of model parameters from their prior estimates. Including this term in the objective function enhances the stability of the problem [125,126].

For the cases analysed in this study (available measurements are heads and the only uncertain hydraulic property is transmissivity  $T$  or its log transform,  $Y$ ), Eq. (6) can be written as:

$$F = \alpha_h F_h + \mu F_p$$

$$= \alpha_h (\mathbf{h} - \mathbf{h}^*)^t \mathbf{V}_h^{-1} (\mathbf{h} - \mathbf{h}^*) + \mu (\mathbf{p} - \mathbf{p}^*)^t \mathbf{V}_p^{-1} (\mathbf{p} - \mathbf{p}^*) \quad (7)$$

From a mathematical point of view, minimisation of Eq. (7) is an optimisation problem, which can be tackled by different strategies, including Broyden-like methods [18], conjugate gradient [47] or variational methods [116], amongst others (reviews can be found in [48,50]). The most widely-spread technique is possibly the Levenberg–Marquardt’s method [48,92,103]. Using Levenberg–Marquardt’s method entails the calculation of the Jacobian (containing the derivatives of hydraulic heads with respect to the model parameters) and can be very CPU intensive. A CPU-efficient alternative is provided by the adjoint state method (e.g., [95,121,128]). Several convergence criteria for stopping the minimisation process can be found in the literature [48].

### 2.3. Posterior statistical analysis

The statistical weights  $\alpha_h$  and  $\mu$  are unknown a priori and can be estimated either jointly with the model parameters or a posteriori. Posing the problem in a Maximum Likelihood framework allows estimating the optimum set of statistical parameters by minimising the support function,  $S_2$ , of the maximum expected likelihood of the parameters given the measurements. The latter is given by the following expression [95]

$$S_2 = D + \log |\hat{\mathbf{H}}| + D \log \left( \frac{\hat{F}}{D} \right) - n_h \log(\alpha_h) - n_p \log(\mu) \quad (8)$$

Here,  $n_h$  and  $n_p$ , respectively are the number of head measurements and model parameters ( $D = n_h + n_p$ );  $\hat{F}$  and  $\hat{\mathbf{H}}$ , respectively denote the objective function and the associated Hessian, evaluated for the optimum set of parameters. The optimisation process is repeated for different values of  $\alpha_h$  and  $\mu$ , whose optimum values are those leading to the minimum of Eq. (8). Note that this posterior analysis handles some of the statistical model uncertainties associated with the prior guess of the hydraulic properties (i.e., the vari-

ance), but does not accommodate the uncertainty of other unknown statistical parameters, e.g., the variogram type. Yet, the methodology is general and can be extended to include the estimation of the remaining statistical parameters. Alternatives to Eq. (8) are the minimisations of other model discrimination criteria, such as the Akaike Information Criterion [2], the Modified Akaike Information Criterion [3,110,115], the Hannan Information Criterion [61] and the Kashyap Information Criterion [76], amongst others. For simplicity, here we do not consider conceptual model and measurement uncertainties and assume unit values for  $\alpha_i$  and  $\mu$ . Therefore, analysis of model selection/information criteria was not carried out, an exception being the *RPPM*.

### 2.4. Regularised Pilot Points Method

The Regularised Pilot Points Method was first introduced by Alcolea et al. [4]. It is a modification of the well-known Pilot Points Method [38,85,86,108]. Early formulations by de Marsily and co-workers considered only the term  $F_h$  in the objective function expressed by Eq. (7). This causes instabilities due to overparameterisation [35,36]. It is well known that adding a regularisation term ( $F_p$  in Eq. (7)) overcomes this problem [125,126]. Doherty [43] and Kowalsky et al. [82] included regularisation for the first time in the context of pilot points. Unfortunately, the role of the regularisation term was not explored. The *RPPM* accounts for the plausibility of model parameters through a regularisation contribution, named plausibility term. This penalises departures of model parameters from their prior estimates. The use of plausibility permits the use of a large number of pilot points (actually, as large as computationally feasible), that allows large parameterisation schemes and, therefore, enhances the number of degrees of freedom to represent heterogeneity. In addition, optimum weighting of the plausibility term overcomes the risk of instabilities and enhances the characterisation of the unknown aquifer properties. Alcolea et al. [4,5,7] present the algorithmic details of the *RPPM* in its conditional estimation (*RPPM-CE*) and in Monte Carlo conditional simulation variants (*RPPM-CS*), as implemented in the finite element code TRANSIN [94,95]. Those works pioneer the posterior analysis of the statistical parameters ( $\alpha_h$  and  $\mu$  in Eq. (7)) in the context of the Pilot Points Method. The *RPPM* has been applied to a number of synthetic and real case studies. Alcolea et al. [7] used both variants to analyse the effect of the small-scale variability of hydraulic conductivity on non-Fickian transport (e.g., tailing in breakthrough curves). The *RPPM* was applied successfully to the management of two coastal aquifers in Spain [6] and Oman [8].

The *RPPM* follows the steps defining the general inversion framework. Model parameters are the values of the unknown aquifer properties at the pilot points’ locations. These can vary along the iterative process, together with the interpolation functions defining the parameterisation (Eqs. (2) and (3)). Interpolation methods include several variants of kriging/co-kriging or sequential Gaussian simulation for the *RPPM-CS* case. The objective function to be minimised is reported in Eqs. (6) and (7). Head measurement errors can be uncorrelated ( $\mathbf{V}_h$  is diagonal) or not ( $\mathbf{V}_h$  is full). For the *RPPM-CE* case, the parameter covariance matrix  $\mathbf{V}_p$  is the kriging error covariance matrix ( $\mathbf{V}_k$ ). The latter is corrected for the *RPPM-CS* case ( $\mathbf{V}_p = 2\mathbf{V}_k$ ). As such,  $\mathbf{V}_p$  is a full matrix and correlation between model parameters is included in the calibration process. By definition, the estimation variance at pilot points located close to measurements is small and values of model parameters are similar when pilot point locations are close (i.e., there is a large correlation). In this work (and only for *RPPM-CE*), the only unknown statistical parameter is  $\mu$ . It is assessed in a Maximum (expected) Likelihood framework by minimising Eq. (8). We found (not shown) that the optimum value of  $\mu$  was close to unity, thus reflecting the absence of uncertainties on the

adopted geostatistical model. We set  $\alpha_n = 1$  (i.e., measurement uncertainties are not taken into account). Additional algorithmic details are summarized in Table 1.

### 2.5. Sequential Self-Calibration

The Sequential Self-Calibration Method [53,114] was the first inverse method that used multiple equally likely stochastic realisations for  $f_p$ . It is therefore able to provide unbiased inverse estimates of heads and concentrations. The equally probable initial guesses are generated by multi-Gaussian sequential co-simulation [52]. A second important innovation was the proposed parameterisation of  $f_p$ . Sahuquillo et al. [114] and Gomez-Hernandez et al. [53] used the pilot points concept. Instead of locating a pilot point at the location with the largest sensitivity,  $\delta F/\delta p$ , they laid out a large number of master blocks (analogous to the pilot points in the RPPM literature), regularly distributed throughout the simulation domain. Model parameters are the values of the unknown aquifer properties at the master blocks. These always include the locations where measurements of the aquifer properties are taken. Hendricks Franssen et al. [62] further extended this concept and varied the position of the master blocks during the optimisation process, avoiding artifacts in the inverse conditioned fields. The interpolation functions in Eqs. (2) and (3) vary accordingly. Capilla et al. [22,23] tested the methodology in a synthetic example and in a real-world case study. The SSC was then extended to handle the stochastic inverse modelling of transient groundwater flow and the joint calibration of transmissivity and storativity fields [62], three-dimensional inverse modelling of fractured media [55,64], coupled inverse modelling of groundwater flow and mass transport [65] and the inclusion of exhaustive soft information from geophysical

surveys or remote sensing [67]. The methodology was implemented in the finite differences code INVERTO [63].

The objective function to be minimised in the sequential self-calibration method follows the formalism of Eq. (7). The plausibility term  $F_p$  is not often accounted for in the context of SSC (as in this work). Nevertheless, Hendricks Franssen et al. [66] found that for the joint inverse simulation of transmissivities and recharge rates it was essential to include a plausibility term  $F_p$  accounting for prior estimates of recharge rates. As an alternative to the use of plausibility, imposing boundary restrictions upon model parameters helps to alleviate inverse problem instabilities. Table 1 gives some further details on the Sequential Self-Calibration Method.

### 2.6. Moment Equations Method

The methodology was originally presented by Hernandez et al. [69,70]. It relies on a nonlinear geostatistical inverse algorithm for steady-state groundwater flow that allows estimating jointly the spatial variability of log-transmissivity, the underlying variogram and its parameters, and the variance-covariance of the estimates. Exact mean flow equations [98,99] are rendered workable by means of a suitable second-order approximation [56,57] (in terms of a small parameter, representing the standard deviation of the underlying random log-transmissivity).

The Moment Equations Method (MEM) can also be cast in the general inversion framework, albeit with some peculiarities. Model parameters are the estimates of the conditional unbiased (ensemble) mean,  $(Y(\mathbf{x}))_c$ , of (natural) log-transmissivity at discrete measurement points and selected pilot point locations. This deterministic parameterisation roughly follows Eq. (1). It can incorporate (optionally) uncertainty of transmissivity values at

**Table 1**  
Computational details and selected references for the inverse methods used in the comparison. Note that SAM is excluded.

	RPPM-CE	RPPM-CS	SSC	MEM	RM	ZM
Methodological reference	[4]	[5,7]	[114]	[69,70]	[11]	[26]
Software	TRANSIN [94,95]	TRANSIN [94,95]	INVERTO [63]	–	InvFeTrans	TRANSIN [94,95]
Problem dimensionality/ governing Equations <sup>1</sup>	1D, 2D, 3D <sup>+</sup> , SS/T, GW/CT-MD/HT	1D, 2D, 3D <sup>+</sup> , SS/T, GW/CT-MD/HT	1D, 2D, 3D, SS/T, GW/CT	1D, 2D, SS, GW	1D, 2D, SS, GW/CT	1D, 2D, 3D <sup>+</sup> , SS/T, GW/CT-MD/HT
Available parameters for calibration <sup>2</sup>	$\mathbf{K}_{3D}(\mathbf{T}_{3D}), S_s(S), q_r,$ $\alpha_L, \alpha_T, \phi, D_m, R, \lambda$	$\mathbf{K}_{3D}(\mathbf{T}_{3D}), S_s(S), q_r,$ $\alpha_L, \alpha_T, \phi, D_m, R, \lambda$	$\mathbf{K}_{3D}(\mathbf{T}_{3D}), S_s(S),$ $q_r, H, \phi, R$	$\mathbf{T}_{2D}$	$\mathbf{K}_{2D}(\mathbf{T}_{2D}), \alpha_L, \alpha_T, R$	$\mathbf{K}_{3D}(\mathbf{T}_{3D}), S_s(S), q_r, H, Q,$ $l, \alpha_L, \alpha_T, \phi, D_m, R, \lambda, c', M$
Discretisation <sup>3</sup>	FE	FE	FD	FE	FE	FE
Coupled to zonation	YES	YES	YES	YES	NO	YES
Parameterisation details <sup>4</sup>	C/VP	C/VP	C/VP	VC	VC	C
Objective function	Eqs. (6) and (7)	Eq. (7)	Eq. (7); $\alpha_n = 1; \mu_V = 0$	Eq. (7); $\alpha_n = 1$	Eq. (7) plus additional term	Eq. (7)
Calculation of derivatives <sup>4</sup>	Analytical, no adjoint state	Analytical, no adjoint state	Analytical, adjoint state	Finite differences	Analytical, adjoint state	Analytical, adjoint state
Optimisation method	Marquardt	Marquardt	Various; alternating	Marquardt	Variational method	Marquardt
Posterior statistical analysis	$\alpha_i, \mu_j$	$\alpha_i, \mu_j$	Rarely	$\mu$	No	$\alpha_i, \mu_j$
Out of general framework	Combination of deterministical/ geostatistical inverse problem [6]		Conditioning to remote sensing images or other exhaustive soft information [67]		Multi-phase flow [129]	

<sup>1</sup> 3D<sup>+</sup>: 1D and 2D elements can be embedded in a 3D model. In that case, 1D and 2D elements can be used to model wells and planar fractures, respectively. List of abbreviations: SS (Steady-State), T (Transient), GW (Groundwater Flow), CT (Conservative Transport), RT (Reactive Transport); MD (analytical formulation of Matrix Diffusion process), HT (Heat Transport).

<sup>2</sup> List of available parameters for calibration:  $\mathbf{K}_{3D}$  (anisotropic 3D hydraulic conductivity or equivalent transmissivity tensor  $\mathbf{T}_{3D}$ ),  $S_s$  (specific storage or equivalent storage coefficient  $S$ ),  $q_r$  (areal recharge),  $H$  (prescribed head),  $Q$  (prescribed flow rate),  $l$  (leakage coefficient) for groundwater flow;  $\alpha_L$  (longitudinal dispersivity),  $\alpha_T$  (transverse dispersivity),  $\phi$  (porosity),  $D_m$  (molecular diffusion coefficient),  $R$  (retardation factor),  $\lambda$  (decay coefficient) and  $M$  (contaminant mass flux).

<sup>3</sup> FE: Finite elements; FD: Finite differences.

<sup>4</sup> C denotes constant parameterisation scheme along the iterative process. VP denotes variable parameterisation due to variable location of the pilot points/master blocks along the iterative process. VC denotes variable parameterisation scheme through the use of posterior kriging.

measurement locations. The location of the pilot points in the current implementation of *MEM* is kept constant during the optimisation procedure. Projection of the  $\langle Y(\mathbf{x}) \rangle_c$  field and the second conditional moment of the associated estimation errors,  $C_{Yc}(\mathbf{x}, \mathbf{y})$ , onto a computational grid is performed by universal kriging, upon considering the variance of measurement errors at actual data points (assumed to be uncorrelated) and the covariance matrix of estimation errors of model parameters. Hence, the interpolation functions defining the parameterisation in Eqs. (2) and (3) vary after each iteration of the optimisation process.

Inversion entails minimising the objective function Eq. (7). The covariance matrix of head errors is assumed to be diagonal (i.e., spatially uncorrelated measurement errors). In Eq. (7),  $\mathbf{V}_p \equiv \mathbf{V}_k$  is the prior kriging error covariance matrix. Posterior statistical analysis can be performed in order to select the optimum value of the weight  $\mu$ . For the sake of simplicity, in the current application we set  $\alpha_h = 1$  (thus implying that we have at our disposal the correct  $\mathbf{C}_h$  matrix) and assume that the optimum  $\mu$  is known ( $\mu = 1$ ).

A unique feature of the method is its capability of providing estimates of the prediction errors of hydraulic heads and fluxes, which are calculated a posteriori upon solving the corresponding moment equations. In this framework, estimates of hydraulic conductivity and head are used together with the posterior kriging covariance matrix of hydraulic conductivity to solve the second-order conditional moment equations for the (posterior) covariances of head and flux. The latter provide measures of predictive uncertainty. Table 1 provides further details on the algorithms adopted in the *MEM*-based procedure. The estimate of the mean well catchment and its associated variance is performed a posteriori according to the work of Riva et al. [111].

### 2.7. Representer Method

The Representer Method (*RM*; [14]) is widely used in meteorological and oceanographic sciences (e.g., [83,101,102]). In the hydrogeological inversion context, *RM* is used to estimate large-scale patterns of heterogeneity of transmissivity and/or dispersivities from scattered measurements of head, concentrations and transmissivity (e.g., [9,105,107,129]) and can also be applied to the characterisation of other aquifer properties. The method either uses a first-order Taylor approximation or a *MC* scheme [11] to provide posterior covariances of estimated parameters and predicted state variables. A recent study by Bakr [13] compares these two uncertainty propagation techniques.

Despite the notation found in the existing literature is far different from the one adopted in this work, the methodology can also be cast in the general framework presented. Model parameters are the transmissivity values at the hydraulic head and transmissivity measurement locations. Parameterisation of transmissivity follows Eqs. (1)–(3). Using the matrix notation of oceanographic literature, Eqs. (1)–(3) can be formulated in terms of log-transmissivity as

$$Y(\mathbf{x}) = \bar{Y}(\mathbf{x}) + \psi^t \mathbf{p} \quad (9)$$

Here,  $\bar{Y}(\mathbf{x})$  is a vector of prior means of log-transmissivity, which is analogous to  $f_D$  in Eq. (1). The Sequential Gaussian Simulation algorithm as implemented by Pebesma and Wesseling [105] is used to generate the initial guesses conditioned to direct measurements of  $Y$ . The second term of Eq. (9) includes  $\psi$ , the so-called measurements parameter representer matrix (it is equivalent to the cross-covariance between parameters and predicted measurements) and the vector  $\mathbf{p}$  containing unknown weighting coefficients of size equal to the number of measurements. This reduces the number of unknowns from the number of finite element nodes to the total number of measurements (see for instance Bakr and Butler [11] for details). The second term of Eq. (9) corresponds to  $f_p$  in Eqs.

(2) and (3) and encompasses the simultaneous conditioning to transmissivity and hydraulic head measurements [10,130]. For linear problems Eq. (9) provides an exact, closed form solution to the least squares problem, which becomes similar to co-kriging. Eq. (9) can also be posed such that the first term is given by multiple unconditional stochastic realisations, and the second term is a perturbation that conditions those realisations both to transmissivity and hydraulic head measurements.

An iterative solution of Eq. (9) is calculated for nonlinear problems. The optimisation procedure is based on minimising an objective function similar to that in Eq. (7), with the only difference that an extra term is included, penalising departures of log-transmissivities from their prior means throughout the entire simulation domain (i.e., at all elements). The other methods analysed (*SSC*, *MEM*, *RPPM-CE*, *RPPM-CS* and *ZM* (see Section 2.8)) do not include this extra term in the objective function. It is worth noting that prior statistics of  $\mathbf{Y}$ ,  $\mathbf{h}$  and  $\mathbf{p}$  are assumed to be known (i.e.,  $\mathbf{V}_Y$  is known and both  $\alpha_h$  and  $\mu$  are set to 1). Thus, posterior statistical analysis is not performed. Table 1 provides further details on the methodology.

### 2.8. Zonation Method

The Maximum Likelihood framework incorporating prior information on model parameters was developed by Carrera and Neuman [26,27] and is popularly known as the Zonation Method (*ZM*). *ZM* is extensively adopted by practitioners (e.g., [1,131]) mainly because of its simplicity and ease of use.

The model domain is partitioned into a set of subdomains and the parameterisation scheme defined in Eq. (1) is used within each of them. Model parameters are the values of the properties within each zone. Point or cell estimation methods are a particular case of zonation (i.e., a zone is defined for each node or element/cell of the discretisation [96]). Hydraulic properties within zones can be constant or vary in a predefined (i.e., deterministic) manner, thus allowing to accommodate heterogeneity within zones. This does not preclude the use of geostatistics in the context of *ZM*. This is clearly demonstrated in the work by Clifton and Neuman [34], who used block kriging to estimate the initial guess of zonal log-hydraulic conductivities. The value of the interpolation functions in Eqs. (2) and (3) is zero if the estimation point  $\mathbf{x}$  falls outside the zone being considered. This parameterisation is a particular case of the general formulation presented in Eqs. (2) and (3). The objective function to be minimised appears in Eq. (6) or its particularization in Eq. (7). Posterior statistical analysis is carried out by minimising the maximum expected likelihood of the parameters given the data.

### 2.9. The Semi-Analytical Method of Stauffer et al. [118,119]

The Semi-Analytical Method (*SAM*) we analyse here displays a strong deviation from the general methodology outlined in Sections 2.1–2.3. Although it allows conditioning to transmissivity and hydraulic head data, its main difference from the other methods is that the solution is calculated non-iteratively (i.e., the method does not include calculating and minimising an objective function). For this reason we provide a stand-alone description of the methodology without referring to the general framework and *SAM* is not included in Table 1.

The prerequisite for the uncertainty analysis using the Semi-Analytical Method as applied in this study is to establish a deterministic steady-state flow model for equivalent hydraulic conductivity in a rectangular two-dimensional flow field. Under such conditions, a pathline approximately represents the ensemble mean trajectory. To estimate the uncertainty of a particle location along the mean pathline, the longitudinal and transverse particle displacement var-

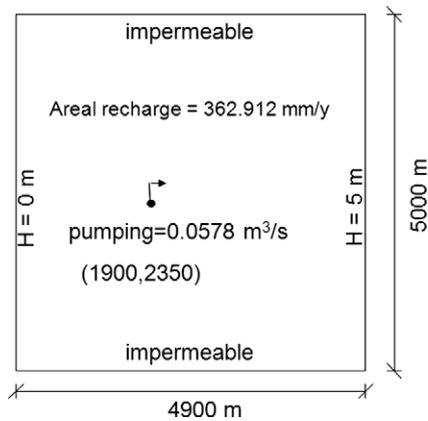


Fig. 1. General set-up for the two synthetic cases studies used for comparing the different inverse methods.

iance can be expressed as a combination of the particle increments over time [118]. To that end, the two-point velocity covariance has to be evaluated along the mean particle trajectory.

The unconditional longitudinal and lateral two-point velocity covariance can be analytically approximated and scaled according to Stauffer et al. [118] using the solution for uniform flow by Rubin [112]. The approximation turned out to be successful except close to stagnation points and boundaries of the flow domain [118,119]. An alternative, although computationally more demanding, is to numerically evaluate the velocity covariance. As the locations at which such a covariance needs to be calculated are not known in advance, the covariances are evaluated for a regular finite differences grid over the flow domain. The head covariance matrix can be obtained from the covariance matrix of  $Y$  and the sensitivity matrix [141]. The velocity covariances are here obtained by numerical differentiation. The conditional expected location of the well catchment is determined as follows. The transverse second moment of the particle displacements along the mean pathline can be obtained by conditioning the expected mean and covariance of velocity to hydraulic conductivity and/or head measurements. The expected mean and covariance of velocity can be determined using the method of conditional probabilities [113]. Determining the covariance of velocities requires the calculation of the cross-covariances between velocity and heads and between velocities and hydraulic conductivity. The latter are evaluated by finite differences. The uncertainty bandwidth along the mean trajectory is finally calculated as twice the square root of the transverse second moment of the particle location. The latter is assumed to follow a Gaussian distribution.

### 3. Synthetic fields

The performance of the seven methods is compared on the basis of a synthetic set-up depicting groundwater flow around a pumping well. Two scenarios including mild and strong reference transmissivity fields (referred to as test cases 1 and 2, respectively) are analysed. We consider transmissivity as an under-sampled parameter and assume its spatial distribution to be the only unknown of the problem. All the remaining aquifer parameters are assumed to be deterministically known. The log-transmissivity variogram is also assumed to be known. Hence, the methods relying on a plausibility term (*MEM* and *RPPM*) employ a unit plausibility weight. Both synthetic cases share the common flow configuration described below.

The domain has a length of 4900 and 5000 m in  $x$  and  $y$  directions, respectively. Prescribed heads of 0 and 5 m are imposed at the western and eastern boundaries, respectively. The northern and southern boundaries are impermeable. A uniform recharge

rate of 362.912 mm/y is imposed all over the domain. A well pumps steadily 0.0578 m<sup>3</sup>/s at location (1900 m, 2350 m). Fig. 1 displays the common flow set-up for both synthetic problems.

Methods using a centred finite differences (*FD*) scheme (*SSC*, *RM*) adopt a regular grid with 49 × 50 cells of 100 × 100 m. *RPPM-CS*, *RPPM-CE* and *MEM* use a second-order Galerkin finite element (*FE*) approach, the pumping well being located at a node. In order to obtain consistent comparisons, *RPPM-CS* and *RPPM-CE* use a more refined *FE* discretisation comprising a regular grid with 98 × 100 square elements of size 50 × 50 m (i.e., four elements defining a single *FD* grid cell). This negatively affects the CPU time needed by *RPPM-CS* and *RPPM-CE*. We tested the effect of the numerical scheme (i.e., *FD* or *FE*) by comparing the reference hydraulic head fields obtained by *FD* and *FE* forward simulations of the groundwater flow set-up and using the reference transmissivity fields. We find that the numerical scheme has very little impact on the final outcomes. Differences are observable only in the vicinity of the pumping well. Therefore, for the sake of robustness of the comparison, we choose to compare all results with the same reference hydraulic head ( $h$ ) field, obtained by finite differences.

Two reference heterogeneous transmissivity ( $T$ ) fields are generated. The first field is mildly heterogeneous, with mean  $\log_{10}T$  of  $-2.932$  and  $\log_{10}T$  variance equal to  $0.189 \log_{10}(\text{m}^2/\text{s})$  (i.e.,  $\sigma_{\ln T}^2 = 1.0$ ). Patterns of heterogeneity are defined by imposing an isotropic exponential variogram without nugget effect and effective range of 500 m (1/10th of the domain size). The second field has also a mean  $\log_{10}T$  of  $-2.932$  with a  $\log_{10}T$  variance equal to  $1.0 \log_{10}(\text{m}^2/\text{s})$  (i.e.,  $\sigma_{\ln T}^2 = 5.3$ ) and is characterised by an isotropic spherical variogram without nugget effect and an effective range of 500 m.

The reference well catchment is determined upon releasing 100 particles, regularly distributed at each grid cell. These are moved by particle tracking until they reach the eastern or western boundaries or are captured by the well. In the latter case, the grid cell from which the particle starts is considered to belong to the well catchment. Particle tracking on the reference field is carried out with 3DTRANSP [63] with linear interpolation of the velocity field. For the sake of simplicity, we consider advection as the only transport mechanism.

The  $Y$  and  $h$  reference fields are sampled at 25 locations, randomly and regularly distributed, respectively for test cases 1 and 2. These data are assumed to be error-free and used for (inverse) conditioning. The comparison focuses on methods for parameter estimation. Thus, we avoid the additional noise arising from corrupted measurements. Nonetheless, measurement uncertainty is an important issue and should be accounted for in meaningful real-world models.

### 4. Comparison measures

The performance of the seven methods is assessed in terms of both reliability of the inverse solution and their predictive capability. The former is evaluated using the following statistics:

- (1) Mean absolute error: it measures the mismatch between calculated and reference values

$$AAE(X) = \frac{1}{N} \sum_{i=1}^N |\bar{X}_i - X_{ref,i}| \quad (10)$$

where  $X$  is either  $Y = \log_{10}T$  or  $h$ ,  $N$  is the number of elements (when  $X = Y$ ) or grid nodes (when  $X = h$ ), the overbar refers to an ensemble average (i.e.,  $\bar{X}_i$  denotes the average of  $X_i$  over all the simulations for *SSC*, *RM* and *RPPM-CS* or the 'single best' value at element ' $i$ ' for the non-MC-based methods) and the subscript *ref* denotes reference value.

- (2) Root mean square error: this measure is similar to AAE but evidences the largest mismatches. As such, it is more sensitive to sampling fluctuations. It is defined as:

$$RMSE(X) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\bar{X}_i - X_{ref,i})^2} \tag{11}$$

$RMSE(Y)$  should be smaller than the a priori standard deviation of  $Y$  (square root of the variogram sill, i.e., 0.43 and  $1.0 \log_{10}(\text{m}^2/\text{s})$  for test cases 1 and 2, respectively), if conditioning on  $Y$  and  $h$  improves the characterisation of the  $Y$  field.

- (3) Average ensemble standard deviation:

$$AESD(X) = \frac{1}{N} \sum_{i=1}^N \sigma_{X_i} \tag{12}$$

where  $\sigma_{X_i}$  is the ensemble standard deviation of  $X$  at node (or element)  $i$ . Although a low value of this measure does not warrant a good characterisation, it is a meaningful comparison criterion, as it indicates which method provides the largest uncertainty reduction.

The above measures (Eqs. (10)–(12)) are calculated both for (1) the whole simulation domain and for (2) the whole simulation domain excluding a box centred at the pumping well with dimensions  $5 \times 5$  elements (or nodes). This is further discussed in Section 6.

The comparison of the predictive capabilities of the tested methods in terms of calculated well catchments is not straightforward because *MEM* and *SAM* calculate continuous solutions of the median well catchment (50% capture probability) and its 95% probability interval (2.5% and 97.5% capture probabilities) assuming that the well catchment is Normally distributed. Instead, *MC* methods calculate a probability value at each grid cell. Thus, outcomes of *MC* and non-*MC* methods cannot be directly compared. In order to do a meaningful comparison we categorise calculated probabilities and assess whether a point belongs to areas with capture probability larger than (a) 97.5%, (b) 50% or (c) 2.5%. We then compute the following quantities:

- (4) Mismatch well catchment: it measures the misfit between the calculated and reference well catchments (in  $\text{km}^2$ ):

$$MWC = A \sum_{i=1}^N |\overline{CP}_{0.50,i} - CP_{ref,i}| \tag{13}$$

where  $CP_{0.50,i}$  is a binary value that indicates whether a certain grid cell  $i$  belongs to the median well catchment ( $CP_{0.50,i} = 1$ ) or not ( $CP_{0.50,i} = 0$ ),  $CP_{ref,i}$  is the reference well catchment (1 or 0 as well) and  $A$  is the surface area of a grid cell ( $A = 0.01 \text{ km}^2$ ).

- (5) Uncertainty of well catchment (*UWC*): This quantity measures the extent (in  $\text{km}^2$ ) of the area with capture probabilities ranging between 2.5% and 97.5%. As the results provided by *SAM* are expected to deteriorate close to the domain boundaries, *MWC* and *UWC* are also calculated over inner domains of different sizes.

- (6) Average absolute error of well catchment. Finally, a comparison of *MC*-based inversion techniques was carried out by calculating an additional statistic, accounting for the full probability distribution of capture probabilities:

$$AAE(WC) = \frac{1}{N} \sum_{i=1}^N |\overline{CP}_i - CP_{ref,i}| \tag{14}$$

where  $CP$  is the capture probability associated with location  $i$ .  $CP_{ref,i}$  is deterministic, with value 0 or 1. All the aforementioned statistics (Eqs. (10)–(14)) are also calculated for a stack of 500 unconditional (*UNC*) simulations of the  $Y$  fields and the corresponding groundwater flow and transport solutions. These will be considered as a base case and the comparison of performances will be presented in terms of reduction with respect to the *UNC* scores. This allows us to assess how conditioning improves characterisation and predictions.

## 5. Results

### 5.1. Test case 1. Mildly heterogeneous transmissivity field ( $\sigma_{\ln T}^2 = 1.0$ )

All methods but *SAM* provide estimates of the  $Y$  and  $h$  fields. *SAM* only renders an estimate of the well catchment. Conditional estimation methods (*RPPM-CE* and *ZM*) do not calculate the uncertainty associated with the  $Y$  and  $h$  fields and the well catchments. The quality of the characterisation results is assessed according to the comparison criteria presented in Section 4. Some computational details of the seven methods, such as the software used and the computational burden are summarised in Table 2. Note that *CPU* times are only indicative and cannot be compared directly because different methods employ different discretisations and number of model parameters.

#### 5.1.1. Reproduction of the reference $Y$ and $h$ fields

The scores for the comparison criteria listed in Section 4 are displayed in the upper part of Table 3. Fig. 2 displays the mean  $Y$  fields obtained by the different methods, together with the  $Y$  reference distribution. Fig. 3 juxtaposes the reference  $h$  field and the  $h$  distribution estimated by the classical *ZM* (the  $h$  fields estimated by the other tested methods are very similar and are not reported here). Fig. 4 displays the ensemble standard deviations associated with both  $Y$  and  $h$  fields.

**Table 2**

Computational burden and algorithmic details of the seven inverse methods.  $N_{iter}$  is the number of iterations, CPU-1 and CPU-2 are the amount of CPU time needed for case 1 and case 2,  $\|g\|$  is the estimated gradient norm of the objective function at the current iteration and  $\|g^1\|$  is the estimated gradient norm of the objective function at the last iteration.

	<i>RPPM-CE</i>	<i>RPPM-CS</i>	<i>SSC</i>	<i>MEM</i>	<i>RM</i>	<i>ZM</i>	<i>SAM</i>
Author	A. Alcolea	A. Alcolea	H.J. Hendricks Franssen	M. Riva & A. Guadagnini	N. van de Wiel & M. Bakr	A. Alcolea	F. Stauffer
Software	TRANSIN	TRANSIN	INVERTO	–	InvFeTrans	TRANSIN	W_CATCH
$n_p$	900	900	400	100	50	–	–
Location	Prescribed	Prescribed	Variable	Prescribed	Prescribed	–	–
Criterion	$\frac{\ g\ }{\ g^1\ } \leq 10^{-6}$	$\frac{\ g\ }{\ g^1\ } \leq 10^{-6}$	$F \leq 0.01$ or $N_{iter} > 3000$	$(F_i - F_{min})/F_i \leq 10^{-3}$ for 5 consecutive times	$\text{Max}e_h \leq 10^{-3}$	$\frac{\ g\ }{\ g^1\ } \leq 10^{-6}$	–
Machine	PC, 2 GHz, 2 GB	PC, 2 GHz, 2 GB	Linux PC, 2.4 GHz, 1 GB	PC, 2 GHz, 2 GB	Single processor Linux Beowulf cluster	PC, 2 GHz, 2 GB	PC, 2 GHz, 1 GB
Grid	$98 \times 100$	$98 \times 100$	$49 \times 50$	$49 \times 50$	$49 \times 50$	$98 \times 100$	$49 \times 50$
CPU-1	125 s	118 h	40 h	8 h	10 h	162 s	~ min
CPU-2	296 s	423 h	40 h	8 h	50 h	171 s	~ min

**Table 3**  
Comparison measures for log-transmissivity and hydraulic head for case 1 (mildly heterogeneous log-transmissivity field) and case 2 (strongly heterogeneous log-transmissivity field). For the deterministic methods, some of the measures could not be evaluated (N.A.). Comparison measures related to hydraulic heads are calculated twice, considering (1) the whole simulation domain, but excluding a box of  $5 \times 5$  elements centred at the pumping well (number to the left of the slash) and (2) the whole simulation domain (right of the slash). Characterisation criteria obtained using a stack of 500 unconditional simulations (actually, conditional to the variogram only) are also listed as *UNC*.

Method	AAE(Y) ( $\log_{10}$ (m <sup>2</sup> /s))	AESD(Y) ( $\log_{10}$ (m <sup>2</sup> /s))	RMSE(Y) ( $\log_{10}$ (m <sup>2</sup> /s))	AAE(h) (m)	AESD(h) (m)	RMSE(h) (m)
<i>Test case 1. Mildly heterogeneous field (<math>\sigma_{\ln T}^2 = 1.0</math>)</i>						
SSC	0.241	0.325	0.306	0.62/0.62	0.99/0.99	0.89/0.90
RPPM-CS	0.243	0.332	0.308	0.65/0.66	0.98/0.98	0.92/0.96
RM	0.241	0.325	0.307	0.62/0.62	0.94/0.94	0.89/0.90
RPPM-CE	0.242	N.A.	0.308	0.67/0.68	N.A.	0.95/0.98
MEM	0.255	0.304	0.321	0.75/0.75	1.42/1.43	1.02/1.02
ZM	0.264	N.A.	0.332	0.76/0.77	N.A.	1.06/1.09
UNC	0.300	0.430	0.379	2.26/2.30	4.89/4.95	2.75/2.87
<i>Test case 2. Strongly heterogeneous field (<math>\sigma_{\ln T}^2 = 5.2</math>)</i>						
SSC	0.712	0.911	0.894	1.45/1.49	2.20/2.48	2.11/2.47
RPPM-CS	0.713	0.754	0.893	1.59/1.73	2.64/3.11	2.50/3.95
RM	0.713	0.939	0.893	1.51/1.63	3.22/3.67	2.43/3.65
RPPM-CE	0.719	N.A.	0.910	1.49/1.51	N.A.	2.11/2.22
MEM	0.728	0.900	0.917	2.25/2.29	5.27/5.35	2.93/2.99
ZM	0.798	N.A.	0.999	2.78/2.78	N.A.	3.56/3.56
UNC	0.765	0.996	0.953	3.44/3.58	5.73/5.96	4.37/5.15

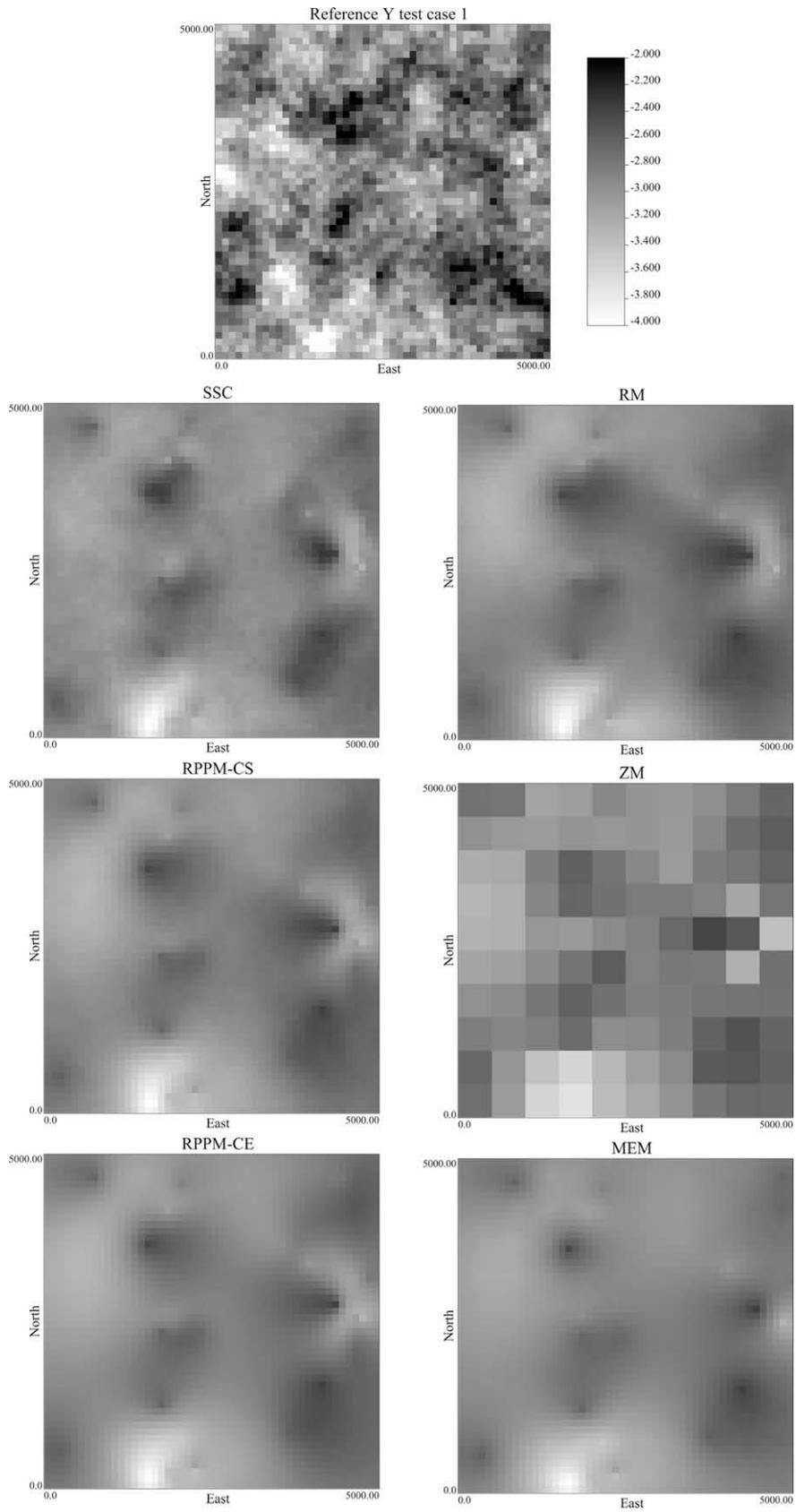
For this case, the differences between the methods are (in general) very mild. Two main features become apparent. First, one can observe the striking similarity between the statistics of the MC-based inversion methods (Table 3). Second, all inverse methods provide (ensemble) mean  $Y$  and  $h$  fields that (i) resemble the reference ones (i.e., they are capable of identifying the regions of high and low transmissivity/hydraulic head) and (ii) are very similar amongst them although, naturally, the  $Y$  field estimated by ZM displays a less grainy texture when compared to other solutions. The benefit of conditioning to available information is apparent when comparing the unconditional and conditional values of AAE( $Y$ ). The three MC-based inversion methods and RPPM-CE reduce AAE( $Y$ ) by 19–20%. MEM and ZM have somewhat less improvement in the characterisation of the  $Y$  field, with reductions of 15% and 12%, respectively. Still, these measures do not deviate significantly from those obtained by MC methods. Very similar results (in terms of reduction of characterisation criteria) are found by observing RMSE( $Y$ ). The uncertainty of the  $Y$  field, as measured by AESD( $Y$ ), does not differ significantly between the methods. It is worth mentioning that AESD( $Y$ ) is more strongly affected than AAE( $Y$ ) by statistical sampling fluctuations. Conditioning on data is more beneficial to the characterisation of the  $h$  field than to the characterisation of the  $Y$  field. This is not surprising, as it is a common observation in inverse modelling studies. For all methods, conditional AAE( $h$ ) is reduced of at least 66%, when compared to the unconditional case. ZM and MEM are associated with slightly larger errors than the other methods, even though the reproduction of the reference  $h$  field is good for all cases. The uncertainty associated with the  $h$  field, as measured by AESD( $h$ ), is again very similar for the three MC-based inverse methods and increases slightly for MEM. Results for RMSE( $h$ ) are systematically consistent with those of AAE( $h$ ). All the aforementioned results hold for the two domains where AAE( $h$ ), AESD( $h$ ) and RMSE( $h$ ) were calculated (i.e., the whole simulation domain and the sub-region obtained upon excluding a box centred at the pumping well).

Fig. 4 shows that MEM yields more smoothed distributions of the ensemble standard deviations than those of MC-based inverse methods. This is consistent with the conclusions by Guadagnini and Neuman [56], albeit in the context of forward simulations. On the other hand, MC-based methods render very similar distributions of ensemble standard deviations of  $Y$ . Observing the distribution of  $Y$  ensemble standard deviations of MEM, one can clearly distinguish the location of pilot points (i.e., the standard deviation at pilot points is substantially reduced). This can be partly ex-

plained by the fact that the interpolation functions defining the parameterisation of MEM vary along the iterative process [70]. Other methods avoid this artifact by moving the location of the pilot points along the iterative process (e.g., SSC) and/or using an increased pilot point density and plausibility of model parameters (e.g., RPPM). By doing so, “lumpy bull’s eyes” of the ensemble  $Y$  fields are reduced and the characterisation of the reference fields is improved. Ensemble standard deviations of  $h$  obtained by MEM are larger than those obtained by the other methods (Table 3). Further work is needed to assess if this observed behaviour of MEM is general or if it depends on the particular configuration chosen for the analysed set-up.

### 5.1.2. Reproduction of the well catchment

The comparison between the demarcation of the reference well catchment and the calculated median well catchments (MWCs) shows that the differences amongst methods are very limited (upper part of Table 4). SAM and, to a lesser extent ZM, display larger misfits (1.17 and 0.93 km<sup>2</sup>, respectively) than the other methods (from 0.75 to 0.86 km<sup>2</sup>). Nevertheless, the MWC value obtained by unconditional simulations (1.58 km<sup>2</sup>) is reduced substantially in all cases (MWC reductions range from 26% for SAM to 53% for SSC). With the exception of SAM, the differences between methods are not significant and are associated with statistical sampling fluctuations and with the type of comparison measure. For instance, evaluation of AAE(WC) leads to a radically different ranking. The smallest values of AAE(WC) are found for two conditional estimation methods (RPPM-CE and ZM). As transport predictions are rather sensitive to small-scale variability of the  $Y$  field, we believe that the differences we found amongst the various performance measures are not significant. The somehow large errors obtained by SAM can be partly explained by the influence of the boundaries. Yet, when the comparison is limited to the inner domain (i.e., at least two integral scales separated from the boundaries, MWC-inner), SAM still displays the largest errors. This issue will be further explored in Section 6. The uncertainty of the mean well catchment, as measured by UWC, is also very similar for the three MC-based inverse methods and slightly larger for MEM and SAM. Note that the reduction in uncertainties with respect to the unconditional simulations is very large (31–44%). Fig. 5 displays the estimated mean well catchments and their uncertainty bounds. Once again, the differences amongst all methods are very small also from a qualitative point of view, with the exception of SAM, at locations close to the boundaries.



**Fig. 2.** The reference Y field and ensemble mean Y fields ( $\log_{10}(\text{m}^2/\text{s})$ ) estimated by six inverse methods for the mildly heterogeneous case.

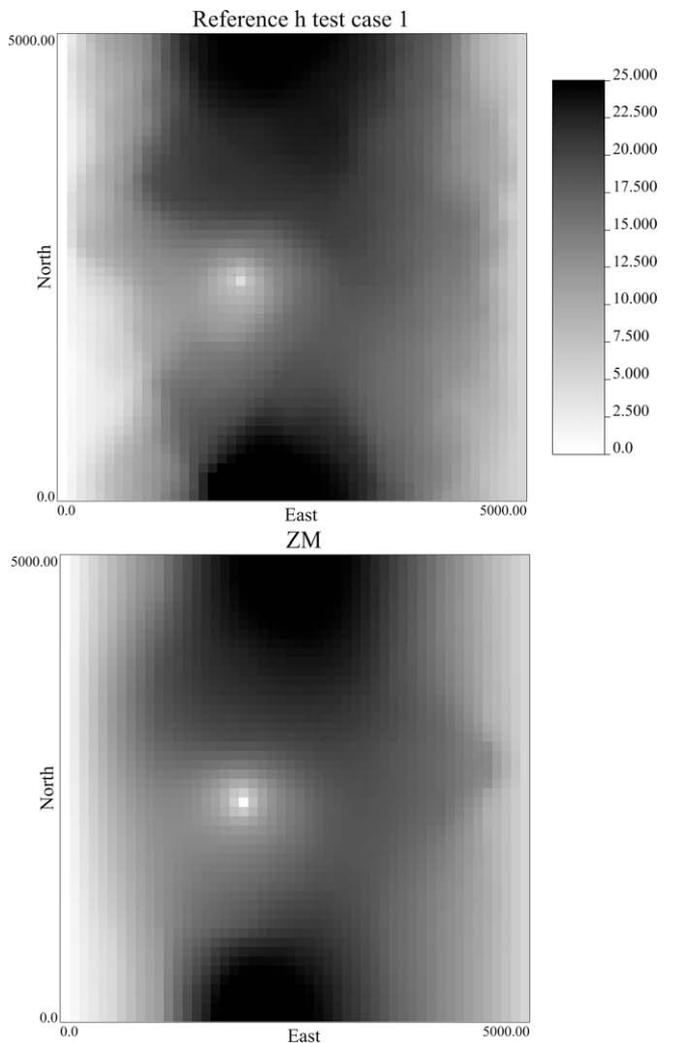


Fig. 3. The reference  $h$  field and the ensemble mean  $h$  field (m) estimated by the Zonation Method for the mildly heterogeneous case.

## 5.2. Test case 2. Strongly heterogeneous transmissivity field ( $\sigma_{\ln T}^2 = 5.2$ )

### 5.2.1. Reproduction of the reference $Y$ and $h$ fields

The scores for the comparison criteria listed in Section 4 are displayed in the lower part of Table 3. Fig. 6 displays the mean  $Y$  fields obtained by the different methods, whereas Fig. 7 depicts the calculated mean  $h$  fields. Fig. 8 shows the corresponding ensemble standard deviations. The differences amongst the methods are slightly larger than those of test case 1. Yet, they are not compelling because the results differ significantly only at a few locations in the vicinity of the well. As expected, reductions in  $AAE(Y)$  are smaller than for the mildly heterogeneous case, ranging from 7% ( $MC$ -based methods) to 5% ( $MEM$ ). This reflects that large heterogeneity renders the calibration process more difficult and less effective. We note that  $ZM$  worsens the unconditional simulation results by 4%. This is so because the rough texture of the  $Y$  field estimated by  $ZM$  does not allow capturing the strong heterogeneity depicted in this test case. Analysis of  $AESD(Y)$  and  $RMSE(Y)$  leads to similar conclusions as in test case 1.

The differences associated with the characterisation of the reference  $h$  field, as measured by  $AAE(h)$  and  $AESD(h)$ , are larger than those observed in test case 1. These become dramatic if  $RMSE(h)$  is considered. Yet, individual inspection of some realisations calcu-

lated by  $MC$ -based methods reveals very low  $Y$  values (much lower than the reference  $Y$  values and those provided by  $ZM$ ,  $MEM$  or  $RPPM-CE$ ) in the vicinity of the pumping well. This causes extremely low-hydraulic head values in the area surrounding the well, thus strongly affecting the resulting ensemble statistics. Significantly lower characterisation criteria are obtained when the vicinity of the pumping well is excluded. This is a consequence of the absence of a conditioning measurement at the well. As compared with test case 1, reductions in characterisation criteria of  $h$  are smaller (never larger than 58%, value obtained by  $MC$ -based methods) than those of  $Y$ . Again, it can be noted that  $MEM$  displays larger variances than the  $MC$ -based inverse methods and  $ZM$  shows a strong departure from the  $Y$  and  $h$  reference fields. One should also note that  $RPPM-CS$  and  $RPPM-CE$  yield very similar results, both for the mildly and strongly heterogeneous cases. A visual inspection of Figs. 6 and 7 reveals small differences between the outcomes of all inverse methods, the exception being the spatial distributions of  $Y$  and  $h$  obtained by  $ZM$ . These differences are now observable also from a qualitative point of view. Standard deviations of  $Y$  and  $h$  (Fig. 8) are similar once again. The large hydraulic head variances of  $MEM$  are also obvious from the figure.

### 5.2.2. Reproduction of the well catchment

Well catchments obtained by the three  $MC$ -based inverse methods do not display large differences (Fig. 9). This is consistent with the results of the characterisation exercise. As opposed to test case 1, now conditioning improves unconditional  $MWC$  results only by a maximum of 37% ( $MWC$  of  $SSC$ ), whereas improvements up to 53% were achieved for test case 1 (upper part of Table 4). The median well catchments obtained by  $SAM$ ,  $ZM$  and  $MEM$  deviate more from the reference than those of  $MC$ -based methods. Scores by  $SAM$  and  $ZM$  are almost identical to their unconditional counterparts ( $UNC$ ).  $RPPM-CE$  and  $MEM$  also yield large prediction errors for all the evaluation criteria because the large  $Y$  variance leads to an over-smoothed  $Y$  distribution that negatively affects transport predictions [7]. Similar ranking results are obtained when the analysis is limited to the inner domain. Again, we find that different prediction criteria lead to different rankings of the methods. Similar conclusions as in test case 1 can be obtained by observing uncertainty bounds, as measured by  $UWC$ . Details are provided in the lower part of Table 4. Fig. 9 depicts the calculated well catchments and their uncertainty bounds. Although the numerical differences between the methods are larger for test case 2, the calculated well catchments do not look that different.

## 6. Discussion

### 6.1. Similarity of $MC$ methods

The three  $MC$ -based inverse methods ( $SSC$ ,  $RPPM-CS$  and  $RM$ ) display very limited differences in their scores of the performance criteria. Similarities are striking with reference to the characterisation of the  $Y$  field. The latter is the heart of inverse estimation and the basis for further predictions in the same aquifer. Differences between reconstructed (ensemble) mean hydraulic head fields are barely visible. With regard to transport prediction, these differences are also very small and the ranking of methods depends on the selected performance criterion. Therefore, this is not indicative of the fact that one method outperforms the other. The question is then: to what extent do these three  $MC$ -based inverse methods, which use different parameterisations and objective functions, characterise and predict the same, or are the same? The similarity of the global performance criteria might be due to locally good/bad results of the methods, which tend to compensate. We address this aspect by analysing the mean residuals of the  $Y$  field (differences

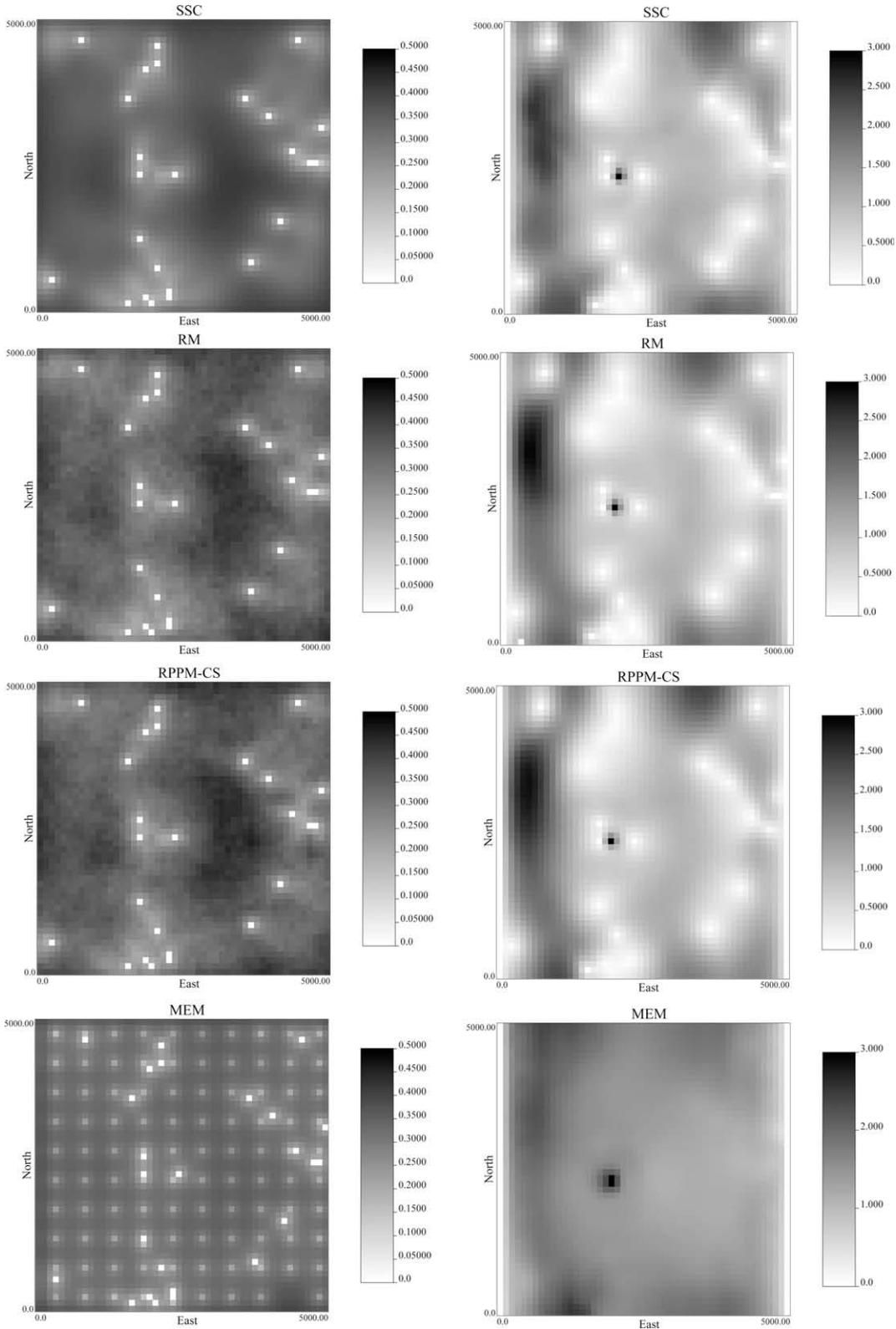


Fig. 4. Ensemble  $Y$  standard deviations ( $\log_{10}(\text{m}^2/\text{s})$ ) (left) and standard deviations for  $h$  (m) (right) for the mildly heterogeneous case.

between ensemble  $Y$  average and reference fields). The linear correlation coefficients between the mean residuals obtained by the three inverse  $MC$  methods are then calculated. As an example, Fig. 10 shows the linear regressions between the mean  $Y$  residuals obtained by the three  $MC$ -based methods for the ensemble of 500 realisations. We argue that the similarities between  $MC$  methods

increase with the number of realisations. To investigate this point, this evaluation was carried out using five groups of 100 realisations and for the whole ensemble of 500 realisations (only for the large variance case). Table 5 displays the results of this analysis. It can be seen that the similarity between methods (i.e., the correlation coefficients between residuals) is large and increases with the number

**Table 4**

Comparison measures for the evaluation of the well catchment for case 1 (mildly heterogeneous log-transmissivity field) and case 2 (strongly heterogeneous log-transmissivity field). For the deterministic methods, some of the measures could not be evaluated (N.A.). For the two methods that estimated a median well catchment with confidence bounds,  $AAE(WC)$  was not calculated. The median well catchment was calculated twice, considering (1) the whole simulation domain and (2) an inner domain of size  $3 \times 3 \text{ km}^2$ . Reference unconditional simulation values (UNC) are also listed.

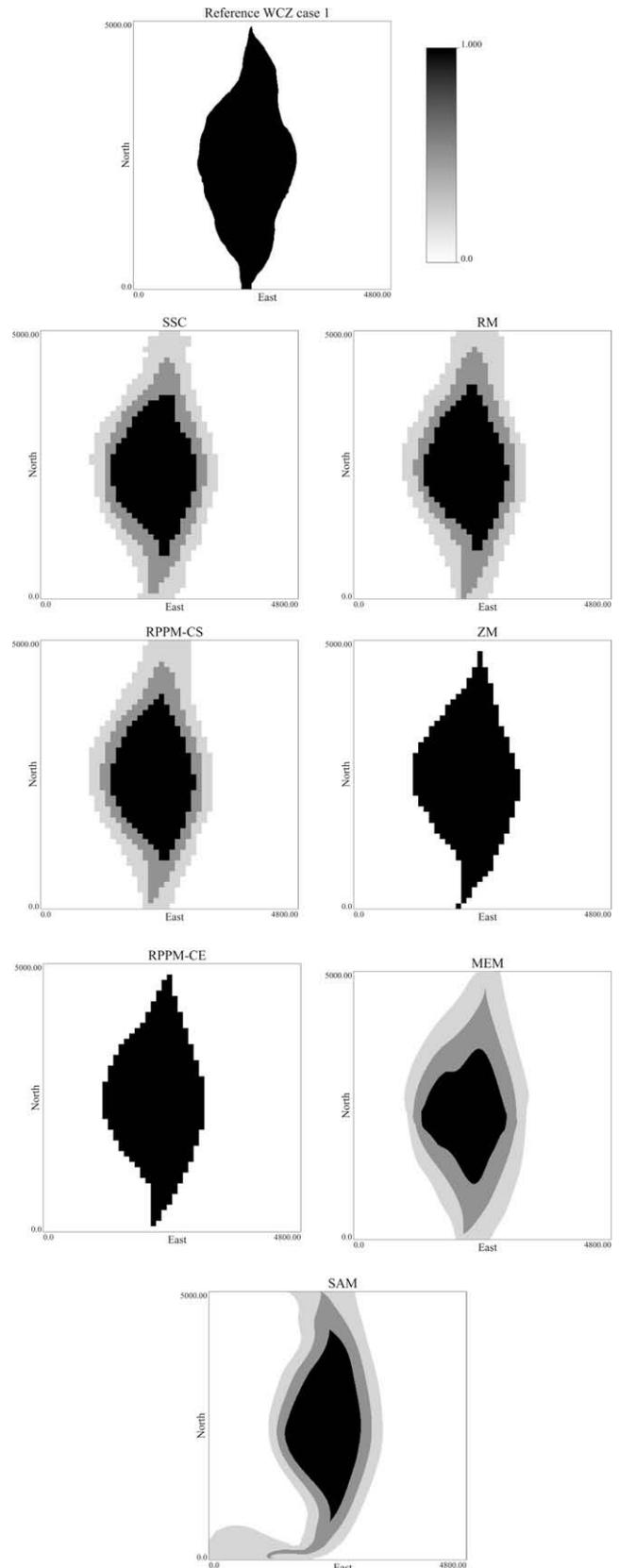
Method	MWC (km <sup>2</sup> )	UWC (km <sup>2</sup> )	AAE(WC)	MWC Inner domain (km <sup>2</sup> )
<i>Test case 1. Mildly heterogeneous field (<math>\sigma_{\ln T}^2 = 1.0</math>)</i>				
SSC	0.75	4.54	0.046	0.405
RPPM-CS	0.86	4.57	0.045	0.490
RM	0.80	4.69	0.042	0.479
RPPM-CE	0.85	N.A.	0.036	0.484
MEM	0.80	5.60	N.A.	0.406
SAM	1.17	5.20	N.A.	0.680
ZM	0.93	N.A.	0.040	0.556
UNC	1.58	8.11	0.092	0.800
<i>Test case 2. Strongly heterogeneous field (<math>\sigma_{\ln T}^2 = 5.2</math>)</i>				
SSC	1.07	9.37	0.031	0.842
RPPM-CS	1.28	7.92	0.037	1.016
RM	1.21	9.04	0.018	0.934
RPPM-CE	1.40	N.A.	0.058	1.192
MEM	1.52	11.38	N.A.	1.068
SAM	1.66	Not solved	N.A.	1.391
ZM	2.28	N.A.	0.058	1.393
UNC	1.69	11.54	0.092	1.400

of realisations. This suggests that part of the differences between the inverse simulations of  $Y$  by MC-based methods is likely due to sampling fluctuations. This is particularly true for the hydraulic head values close to the well. Thus, an almost perfect correlation is expected for very large stacks of realisations.

Our results differ somehow from those by Zimmerman et al. [142], who found large differences between their tested MC-based methods. The comparison by Zimmerman et al. [142] mainly focused on the blind prediction of four tracer tests. These predictions are more affected by small-scale variability of hydraulic conductivity than groundwater flow predictions (e.g., [7]). In addition, particle transport in those tracer experiments took place only in a limited part of the aquifer. Therefore, heterogeneity of hydraulic conductivity was under-sampled. This made those predictions more prone to (random) variations that are not necessarily related to the inverse methods and could be a reason for the larger differences between the MC-based inverse methods in that study. Another reason is that variogram estimation was also a part of the prediction exercise by Zimmerman et al. [142].

## 6.2. Influence of the boundaries

For some methods like SAM, one can expect that the characterisations and transport predictions are worse close to the boundaries. In order to investigate the boundary influence, MWC was also evaluated over inner domains that exclude grid cells/elements within 1.0, 1.4, 1.8, 2.0, 2.6 and 3.0 integral scales far away from the boundary. Results of this detailed comparison are shown in Fig. 11. For both test cases (especially the mildly heterogeneous one), SAM displays larger errors close to (within one integral scale from) the boundaries than the other methods. However, this only partly explains its worse performance as compared to other methods. SAM also shows large errors in areas close to the median well catchment. This is likely related to the limited accuracy of the numerical evaluation of the covariances and the first-order approach. As expected, MEM performs relatively better than the other methods in the vicinity of the well than further away from it (Fig. 11). The MWC for distance classes larger than two integral scales from the boundary corresponds implicitly to that of the area



**Fig. 5.** Ensemble median well catchments and their 95% uncertainty intervals, estimated by seven inverse methods, for the mildly heterogeneous case. The darkest colour indicates estimated capture probabilities larger than 97.5%, the dark grey capture probabilities between 50% and 97.5%, the light grey capture probabilities between 2.5% and 50% and white capture probabilities below 2.5%. The reference well catchment is also displayed.

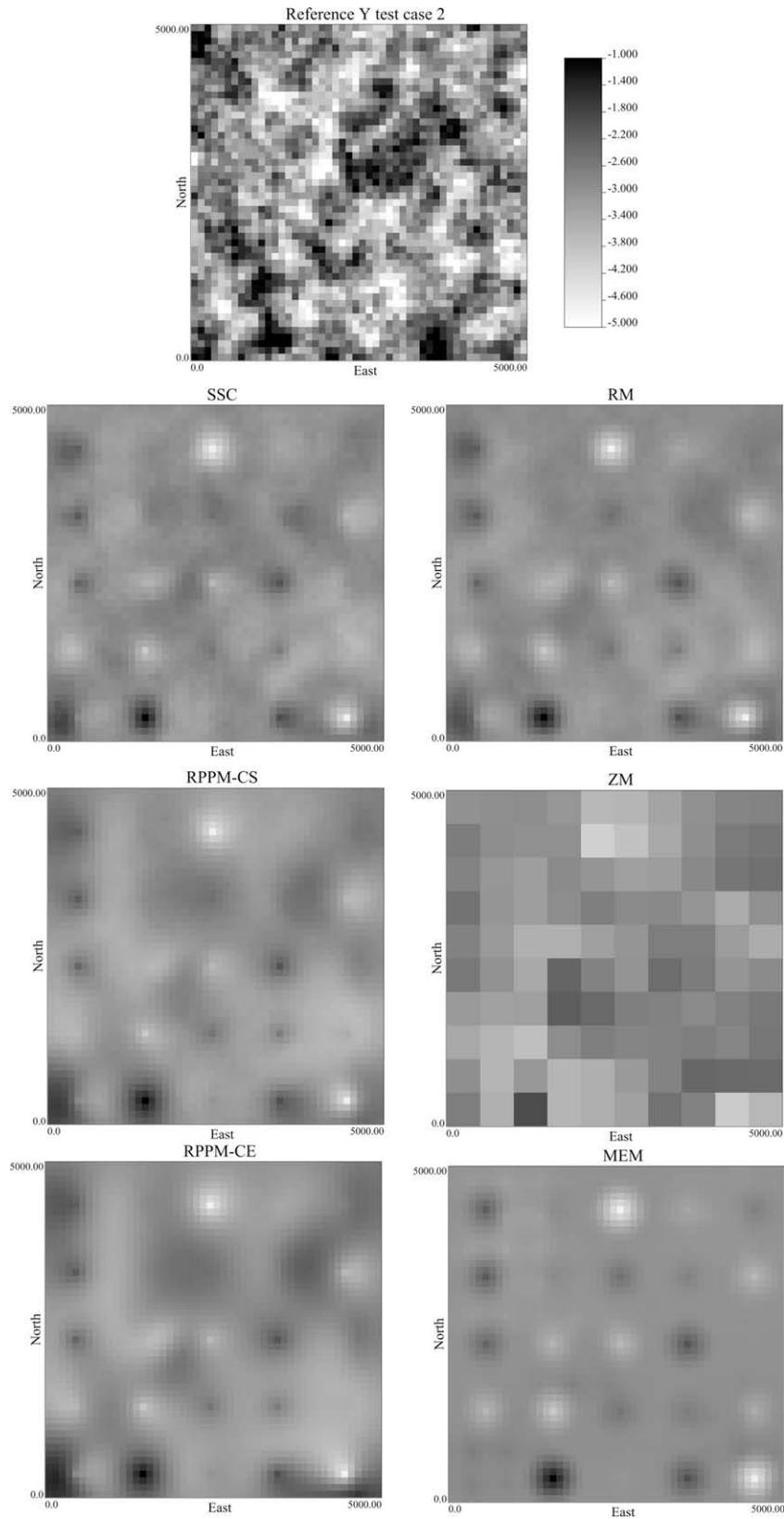


Fig. 6. The reference Y field and the ensemble mean Y fields ( $\log_{10}(\text{m}^2/\text{s})$ ) estimated by six inverse methods for the strongly heterogeneous case.

close to the well. It is for these distance classes that MEM performs good for the mildly and strongly heterogeneous cases. The other methods did not show a consistent pattern with regard to the error distribution as a function of the distance class.

### 6.3. Which method is best?

This paper is by no means aimed at establishing a ‘ranking’ of inversion methods. In fact, it is difficult, even hardly possible, to se-

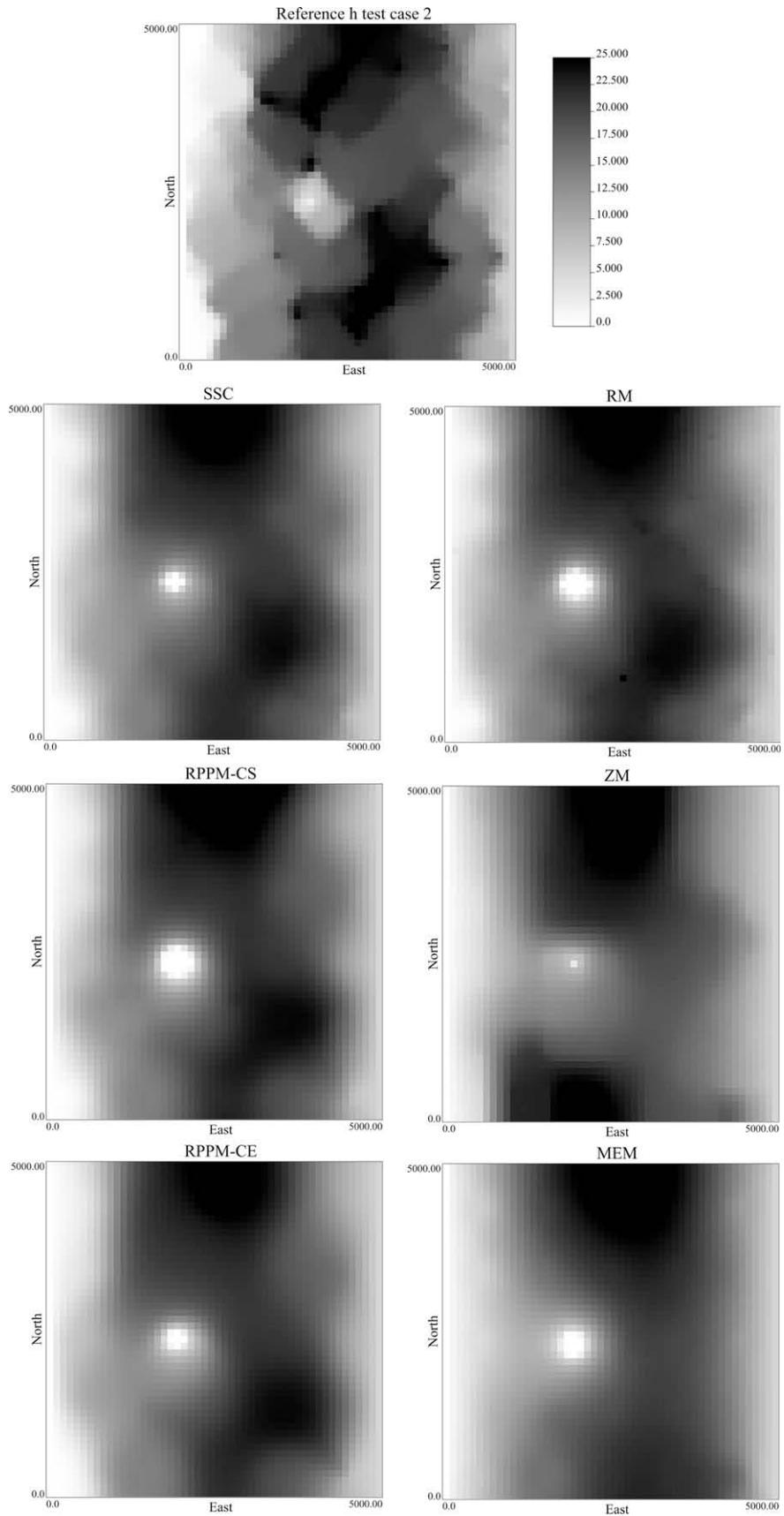


Fig. 7. The reference  $h$  field and the ensemble mean  $h$  fields (m) estimated by six inverse methods for the strongly heterogeneous case.

lect unambiguously which method (amongst the ones tested) is best. Although the  $MC$ -based inverse methods ( $RPPM-CS$ ,  $RM$ ,  $SSC$ )

yield the best overall scores, the differences with the other methods are generally small for the mildly heterogeneous formation.

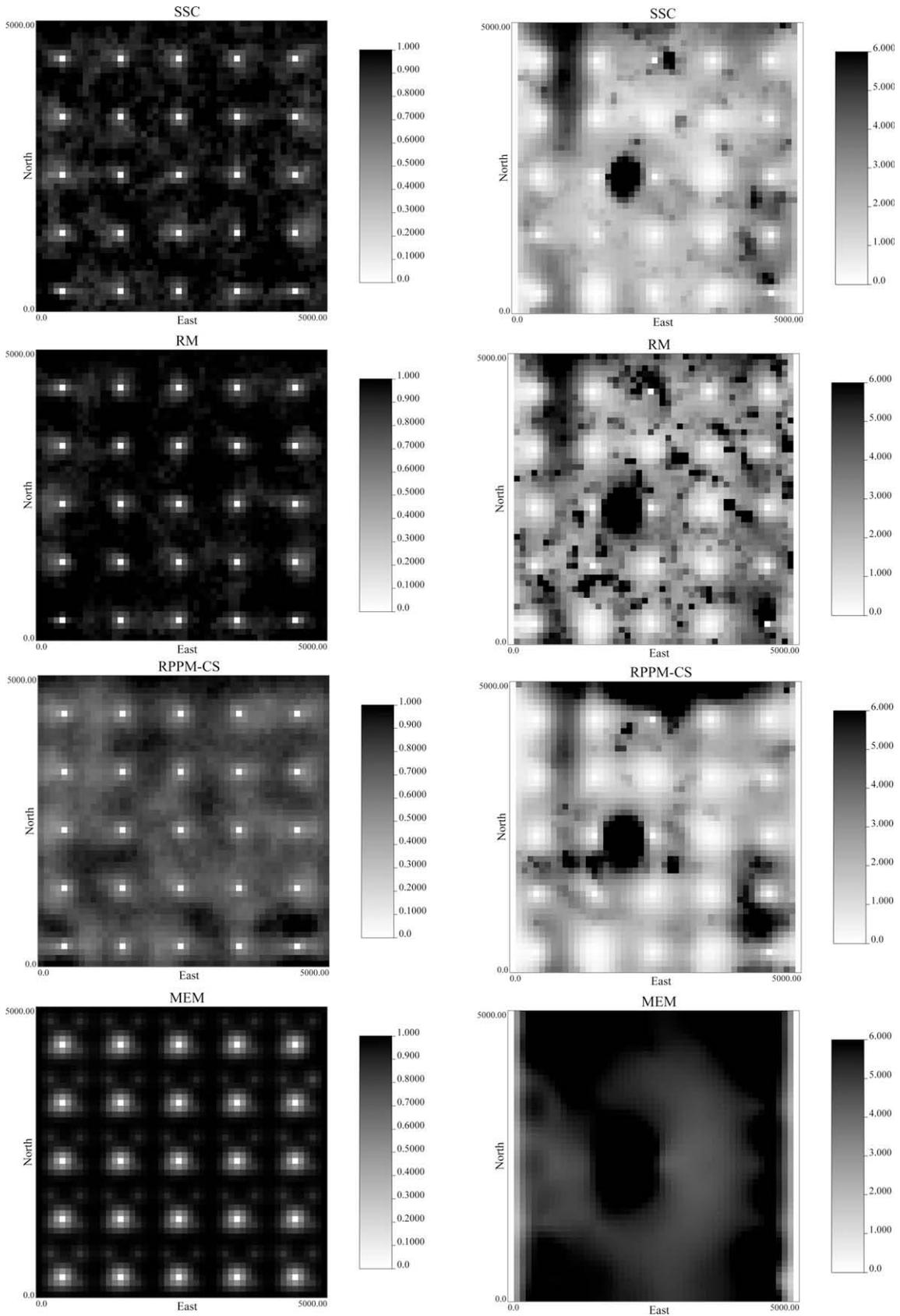
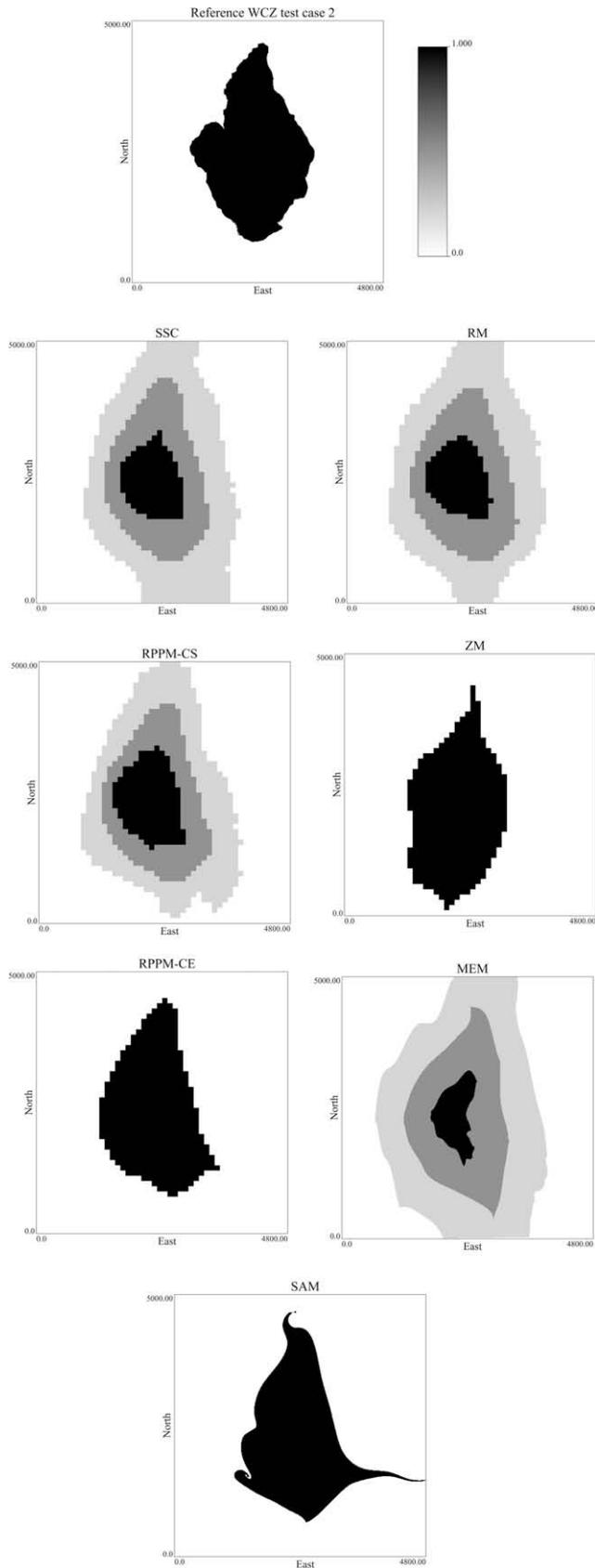


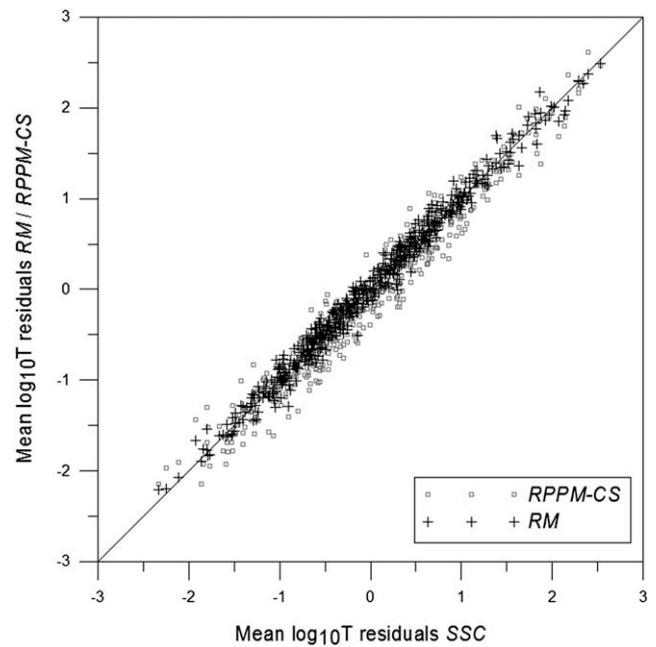
Fig. 8. Ensemble  $Y$  standard deviations ( $\log_{10}(\text{m}^2/\text{s})$ ) (left) and standard deviations for  $h$  (m) (right) for the strongly heterogeneous case.

In case of *MEM* and *RPPM-CE*, the observed differences are also small for the strongly heterogeneous case. On the other hand, the

non-MC methods need considerably less CPU time to calculate the 'single best' solution. Therefore, the choice of an inverse meth-



**Fig. 9.** Ensemble median well catchments and their 95% uncertainty interval, estimated by seven inverse methods, for the strongly heterogeneous case. The darkest colour indicates estimated capture probabilities larger than 97.5%, the dark grey capture probabilities between 50% and 97.5%, the light grey capture probabilities between 2.5% and 50% and white capture probabilities below 2.5%. The reference well catchment is also displayed.



**Fig. 10.** Scatter plot of the mean Y residuals (ensemble mean minus reference) for SSC versus RM and RPPM-CS, for the strongly heterogeneous case.

**Table 5**

Linear correlation coefficients between the mean residuals (calculated ensemble log-transmissivity – reference log-transmissivity) obtained with the three MC-based inverse methods (SSC, RPPM-CS and RM), for the strongly heterogeneous case. These are calculated both for the ensemble of 500 realisations (right column) and for five groups of 100 realisations each (average presented in the left column).

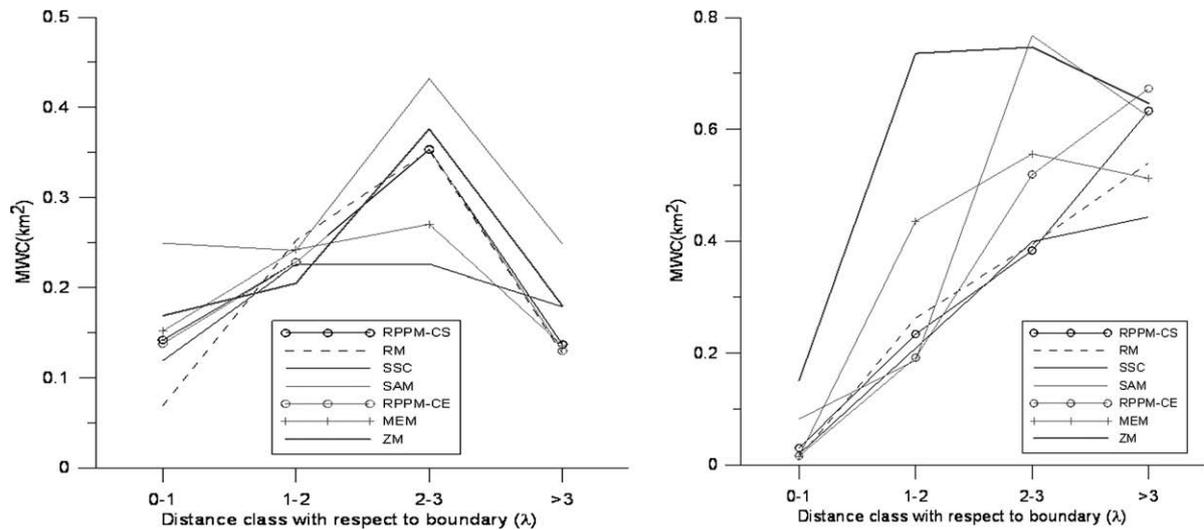
Methods compared	100 realisations	500 realisations
SSC vs. RPPM-CS	0.970	0.975
SSC vs. RM	0.983	0.990
RM vs. RPPM-CS	0.975	0.981

od should depend on the problem under study (e.g., on the degree of heterogeneity), the detail of the required solution and the amount of available time to obtain it. By looking at the results of this comparison study, it is suggested that practitioners should consider the use of other available inverse methods, instead of focusing on the very popular Zonation Method (ZM). In this work, ZM was applied in such a way that each zone includes one integral scale of Y in the x and y directions. If the aquifer would have been divided in a larger number of smaller zones, the performance would have been better.

The authors feel that, given the inherent stochastic nature of the subsurface, and groundwater modelling in general, it is very important to quantify the uncertainty of model predictions. The characterisation of the uncertainty of model predictions should also become much more common in practical applications [109]. Unfortunately, the assessment of the uncertainty of groundwater flow and mass transport predictions is still uncommon outside academia.

#### 6.4. What comes next?

This comparison study helps to gain some insights on the strong and weak points of different inverse modelling techniques. The detailed definitions of the two test cases analysed here are available for the readers, and could be used as a benchmark to test other inverse conditioning algorithms. Actually, we hope that this study will trigger more comparisons in a nearby future.



**Fig. 11.** Comparison measure  $MWC$  ( $\text{km}^2$ ) for the different inverse methods and its values calculated for different distance classes.  $0-1\lambda$  means that only the grid cells are included that are within one integral scale of the boundary,  $1\lambda - 2\lambda$  means that only the grid cells are included that are between one and two integral scales from the boundary, etc. (left) mildly heterogeneous case, (right) strongly heterogeneous case.

Given the fact that the studied inverse methods often yield very similar results (some of them obtain nearly identical results), we feel that future research should be oriented towards the:

- (1) Inverse characterisation of more complex geological formations exhibiting complex channelling structures. These can be modelled, e.g., by multi-point geostatistical techniques. Some on-going research is referenced in Section 1.
- (2) Simultaneous identification of multiple parameters (of different types) for different sources of uncertainty. An example is the prediction of regional groundwater flow, whose uncertainty is jointly governed by parameters such as transmissivity, recharge rate and quantities associated with interactions between rivers and aquifer. In principle, the uncertainty of additional parameters could be handled by including different parameter types in the second term of Eq. (6), and by optimising  $\alpha_i$  and  $\mu_j$  on the basis of suitable statistical model selection criteria. Nevertheless, there are very few studies that take into account the joint calibration of complex regional models with many different unknown parameters.
- (3) Incorporating additional data sources for inverse conditioning. Examples include: natural tracer data, geophysical information, two-point connectivity data, remote sensing images and other new data sources to condition a groundwater flow and mass transport model.
- (4) Inverse modelling of flow and transport problems with more complicated physico-chemical processes. An example is the inverse conditioning to concentration data in the case of reactive transport or multi-phase flow.

## 7. Conclusions

For two different synthetic cases, corresponding to mildly and strongly heterogeneous transmissivity fields, seven inverse methods were used to estimate the mean log-transmissivity and hydraulic head fields and their corresponding variances. The solutions are then used to predict the catchment of a pumping well and its associated uncertainty. In both cases, the inverse conditioning is done on the basis of 25 transmissivity and hydraulic head data, either irregularly (first test case) or regularly (second test case) dis-

tributed over the flow domain. Our comparison study leads to the following key conclusions:

- (1) In both cases, the MC-based inversion methods (Sequential Self-Calibration, Regularised Pilot Points Method and the Representer Method) yield very similar results. Residuals (difference between predicted ensemble mean and reference values) obtained by the three inverse MC methods are strongly correlated, displaying a linear correlation that increases with the number of stochastic realisations. This suggests that the small differences between the methods are most probably related to the stack size, and that the methods would give almost identical results for a very large ensemble.
- (2) The Regularised Pilot Points Method in its estimation variant (RPPM-CE) yields similar results to those of the MC-based inverse methods for most comparison criteria, an exception being the predictions of transport in the high variance scenario (test case 2). This suggests that, below a certain heterogeneity threshold, a zero-order approximation is sufficient for obtaining good characterisations of the  $Y$  and  $h$  fields, as well as meaningful transport predictions. In addition, the CPU cost associated with zero-order methods is considerably lower than that of MC-based or second-order methods (e.g., MEM).
- (3) The Moment Equations Method yields (both for the mildly and strongly heterogeneous case) very similar results to those of MC-based inversion methods, but somewhat worse results for the characterisation of the hydraulic head field.
- (4) The Semi-Analytical Method (which only estimates the well catchment) does not match the well catchments as good as the MC-based inverse methods. Possibly, this problem is more related to the determination of the location and the second moment at the stagnation points, than to the large variance of the  $Y$  field of test case 2.
- (5) It is found that the popular Zonation Method gives worse results than the other methods, especially for the large variance case. This can be explained by the rough texture of the estimated  $Y$  field, which does not allow accommodating strong heterogeneities in a proper way. The use of a large number of smaller zones helps to alleviate this problem.

- (6) In general, the differences amongst all tested methods are limited. Then, the fact that the non-MC inverse approaches use much less CPU time than the MC-based approaches makes it necessary to balance, on one side, the cost of applying a certain method and, on the other side, the level of detail of the required solution.

The comparison presented here is limited to some extent, mainly because: (1) only one groundwater flow and transport scenario is considered; (2) only parameter uncertainty is accounted for (measurement and conceptual uncertainties (especially the latter) are of great importance in real-world applications); (3) for the sake of simplicity and to facilitate the comparison, the geological scenarios considered here are simple (real-life heterogeneity is, indeed, more complex); (4) uncertainty of other parameters is also important for practical applications. In short, much remains to be done. In this work, we have identified some salient features of seven inverse methods. We hope that this comparison will trigger further research on comparisons between other methods, perhaps in a more realistic context.

### Acknowledgements

This study was partly performed within the European Research Project “Stochastic Analysis of Well Head Protection and Risk-Assessment” W-SAHARA. For H.J. Hendricks Franssen and F. Stauffer, this European research project was supported by the Swiss Federal Office for Education and Science (BBT), Project BBW 99.0543. Andres Alcolea wishes to thank the Swiss National Science Foundation for providing funding for his part of the work (Contract PP002-106557). This work was supported in part by funding from MIUR (Italian Ministry of Education, Universities and Research-PRIN2006; project: “Statistical estimation of heterogeneity in complex randomly heterogeneous geologic media). We would like to thank Shlomo Neuman and Philippe Renard for reviewing earlier versions of this paper. Their comments helped to improve this paper.

### References

- [1] Abarca E, Vazquez-Sune E, Carrera J, Capino B, Gamez D, Batlle F. Optimal design of measures to correct seawater intrusion. *Water Resour Res* 2006;42(9):1–14. doi:10.1029/2005wr004524. W09415.
- [2] Akaike H. A new look at statistical model identification. *IEEE Trans Autom Control* 1974;AC19(6):716–23.
- [3] Akaike H. On entropy maximization principle. In: *Applications of statistics*. Amsterdam: The Netherlands; 1977.
- [4] Alcolea A, Carrera J, Medina A. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Adv Water Res* 2006;29(11):1678–89. doi:10.1016/j.advwatres.2005.12.009.
- [5] Alcolea A, Carrera J, Medina A. Inversion of heterogeneous parabolic-type equations using the pilot points method. *Int J Numer Meth Fluid* 2006;51(9–10):963–80.
- [6] Alcolea A, Castro E, Barbieri M, Carrera J, Bea S. Inverse modeling of coastal aquifers using tidal response and hydraulic tests. *Ground Water* 2007;45(6):711–22. doi:10.1111/j.1745-6584.2007.00356.x.
- [7] Alcolea A, Carrera J, Medina A. Regularized pilot points method for reproducing the effect of small-scale variability: application to simulations of contaminant transport. *J Hydrol* 2008;355(1–4):76–90. doi:10.1016/j.jhydrol.2008.03.004.
- [8] Alcolea A, Renard P, Mariethoz G, Bertone F. Reducing the impact of a desalination plant using stochastic modeling and optimization techniques. *J Hydrol* 2009;365(3–4):275–88.
- [9] Bakr MI. A stochastic inverse-management approach to groundwater quality problems. PhD thesis. Technical University of Delft, The Netherlands; 2000.
- [10] Bakr MI, te Stroet CBM, Meijerink A. Stochastic groundwater quality management: role of spatial variability and conditioning. *Water Resour Res* 2003;39(4). doi:10.1029/2002wr001745. 1078.
- [11] Bakr MI, Butler AP. Worth of head data in well-capture zone design: deterministic and stochastic analysis. *J Hydrol* 2004;290(3–4):202–16. doi:10.1016/j.jhydrol.2003.12.004.
- [12] Bakr MI. A groundwater quality inverse model to estimate spatially variable hydraulic parameters and to filter system noise. In: *IV International symposium on environmental hydrology and the IV regional conference on civil engineering technology*, Cairo, Egypt; 2005.
- [13] Bakr MI. A Monte Carlo representer-based inverse method for groundwater flow problems: a comparison with single realization approach. In: *IV International symposium on environmental hydrology and the IV regional conference on civil engineering technology*, Istanbul, Turkey; 2008.
- [14] Bennet AF. *Inverse methods in physical oceanography*. New York: Cambridge University Press; 1992.
- [15] Beven K. Prophecy, reality and uncertainty in distributed hydrological modeling. *Adv Water Res* 1993;16(1):41–51.
- [16] Blasone RS, Madsen H, Rosbjerg D. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *J Hydrol* 2008;353(1–2):18–32. doi:10.1016/j.jhydrol.2007.12.026.
- [17] Blasone RS, Vrugt JA, Madsen H, Rosbjerg D, Robinson BA, Zylolowski GA. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling. *Adv Water Res* 2008;31(4):630–48. doi:10.1016/j.advwatres.2007.12.003.
- [18] Broyden CG. Quasi-Newton methods and their application to function minimisation. *Math Comput* 1967;21(99):368–81.
- [19] Burgers G, van Leeuwen PJ, Evensen G. Analysis scheme in the ensemble Kalman filter. *Mon Weather Rev* 1998;126(6):1719–24.
- [20] Caers J. Geostatistical history matching under training-image based geological model constraints. *SPE J* 2003;8(3):218–26.
- [21] Caers J, Hoffman T. The probability perturbation method: a new look at Bayesian inverse modeling. *Math Geol* 2006;38(1):81–100. doi:10.1007/s11004-005-9005-9.
- [22] Capilla JE, Rodrigo J, Gomez-Hernandez JJ. Simulation of non-Gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information. In: *Third annual conference of the international-association-for-mathematical-geology (IAMG 97)*, Barcelona, Spain Scopus; 1997.
- [23] Capilla JE, Gomez-Hernandez JJ, Sahuquillo A. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric head data – 3. Application to the Culebra formation at the Waste Isolation Pilot Plan (WIPP), New Mexico, USA. *J Hydrol* 1998;207(3–4):254–69.
- [24] Capilla JE, Rodrigo J, Gomez-Hernandez JJ. Simulation of non-Gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information. *Math Geol* 1999;31(7):907–27.
- [25] Carle SF, Fogg GE. Modeling spatial variability with one multidimensional continuous-lag Markov chain. *Math Geol* 1997;29(7):891–918.
- [26] 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.
- [27] Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions. 2. Uniqueness, stability and solution algorithms. *Water Resour Res* 1986;22(2):211–27.
- [28] Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions. 3. Application to synthetic and field data. *Water Resour Res* 1986;22(2):228–42.
- [29] Carrera J. State of the art of the inverse problem applied to the flow and solute transport problems. In: *ASI. Groundwater flow and quality modeling*. NATO; 1987. p. 549–85.
- [30] Carrera J, Glorioso L. On geostatistical formulations of the groundwater flow inverse problem. *Adv Water Res* 1991;14(5):273–83.
- [31] Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ. Inverse problem in hydrogeology. *Hydrogeol J* 2005;13(1):206–22. doi:10.1007/s10040-004-0404-7.
- [32] Chen Y, Zhang DX. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Adv Water Res* 2006;29(8):1107–22. doi:10.1016/j.advwatres.2005.09.007.
- [33] Christensen S, Doherty J. Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration. *Adv Water Res* 2008;31(4):674–700. doi:10.1016/j.advwatres.2008.01.003.
- [34] Clifton PM, Neuman SP. Effects of kriging and inverse modeling on conditional simulation of the Avra valley aquifer in Southern Arizona. *Water Resour Res* 1982;18(4):1215–34.
- [35] 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.
- [36] Cooley RL, Hill MC. Comment on RamaRao et al. [1995] and LaVenue et al. [1995]. *Water Resour Res* 2000;36(9):2795–7.
- [37] Dagan G. Stochastic modeling of groundwater flow by unconditional and conditional probabilities: the inverse problem. *Water Resour Res* 1985;21(1):65–72.
- [38] de Marsily G. *De l'identification des systemes hydrogeologiques*. PhD thesis, Paris VI, France; 1978.
- [39] de Marsily G, 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 characterisation. Part 2*. D. Reidel Pub Co; 1984. p. 831–49.
- [40] de Marsily G, Delhomme JP, Delay F, Buoro A. 40 years of inverse problems in hydrogeology. *Comptes Rendus De L Academie Des Sciences Serie Ii Fascicule a-Sciences De La Terre Et Des Planetes* 1999;329(2):73–87.
- [41] de Marsily G, Delay F, Goncalves J, Renard P, Teles V, Violette S. Dealing with spatial heterogeneity. *Hydrogeol J* 2005;13(1):161–83. doi:10.1007/s10040-004-0432-3.

- [42] Deutsch CV, Journel AG. *GSLIB, Geostatistical software library user's guide*. New York: Oxford University Press; 1998.
- [43] Doherty J. Ground water model calibration using pilot points and regularization. In: MODFLOW 2001 and other modeling Odysseys conference, Golden, Colorado; 2001.
- [44] Drecourt JP, Madsen H, Rosbjerg D. Calibration framework for a Kalman filter applied to a groundwater model. *Adv Water Res* 2006;29(5):719–34. doi:10.1016/j.advwatres.2005.07.007.
- [45] Edwards AWF. *Likelihood*. Cambridge: Cambridge University Press; 1972.
- [46] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res – Ocean* 1994;99(C5):10143–62.
- [47] Fletcher R, Reeves CM. Function minimization by conjugate gradients. *Comput J* 1964;7(2):149–54.
- [48] Fletcher R. *Practical methods of optimization*. 1. Unconstrained optimization. UK: John Wiley & Sons, Ltd; 1980.
- [49] Fu J, Gomez-Hernandez J. Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov chain Monte Carlo method. *J Hydrol* 2009;364:328–41. doi:10.1016/j.jhydrol.2008.11.014.
- [50] Gill PE, Murray W, Wright MH. *Practical optimization*. London: Academic Press; 1981.
- [51] Gomez-Hernandez JJ, Srivastava RM. ISIM3D: an ANSI-C three-dimensional multiple indicator conditional simulation model. *Comput Geosci* 1990;16(4):395–440.
- [52] Gomez-Hernandez JJ, Journel AG. Joint sequential simulation of multi-Gaussian fields. In: Fourth international geostatistics congress: Troia 92, Troy, Portugal; 1992.
- [53] Gomez-Hernandez JJ, Sahuquillo A, Capilla JE. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data – I Theory. *J Hydrol* 1997;203(1–4):162–74.
- [54] Gomez-Hernandez JJ, Wen XH. To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Adv Water Res* 1998;21(1):47–61.
- [55] Gomez-Hernandez JJ, Franssen HJH, Cassiraga EF. Stochastic analysis of flow response in a three-dimensional fractured rock mass block. *Int J Rock Mech Min Sci* 2001;38(1):31–44.
- [56] Guadagnini A, Neuman SP. Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains. 1. Theory and computational approach. *Water Resour Res* 1999;35(10):2999–3018.
- [57] Guadagnini A, Neuman SP. Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains. 2. Computational examples. *Water Resour Res* 1999;35(10):3019–39.
- [58] Guardiano F, Srivastava RM. Multivariate geostatistics: beyond bivariate moments. In: A Soares Geostatistics Troia 1992. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1993. p. 133–44.
- [59] Gupta HV, Bastidas LA, Sorooshian S, Shuttleworth WJ, Yang ZL. Parameter estimation of a land surface scheme using multicriteria methods. *J Geophys Res – Atmos* 1999;104(D16):19491–503.
- [60] Haldorsen HH, Lake LW. A new approach to shale management in field scale simulation models. *Soc Petrol Eng J* 1984;24(4):447–57.
- [61] Hannan EJ. The estimation of the order of an ARMA process. *Ann Stat* 1980;8:1071–81.
- [62] Hendricks Franssen HJ, Gomez-Hernandez JJ, Capilla JE, Sahuquillo A. Joint simulation of transmissivity and storativity fields conditional to steady-state and transient hydraulic head data. *Adv Water Res* 1999;23(1):1–13.
- [63] Hendricks Franssen HJ. *Inverse stochastic modeling of groundwater flow and mass transport*. PhD thesis, Technical University of Valencia, Spain; 2001.
- [64] Hendricks Franssen HJ, Gomez-Hernandez JJ. 3D inverse modelling of groundwater flow at a fractured site using a stochastic continuum model with multiple statistical populations. *Stoch Environ Res Risk Assess* 2002;16(2):155–74. doi:10.1007/s00477-002-0091-7.
- [65] Hendricks Franssen HJ, Gomez-Hernandez J, Sahuquillo A. Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data. *J Hydrol* 2003;281(4):281–95. doi:10.1016/S0022-1694(03)00191-4.
- [66] Hendricks Franssen HJ, Stauffer F, Kinzelbach W. Joint estimation of transmissivities and recharges – application: stochastic characterisation of well capture zones. *J Hydrol* 2004;294(1–3):87–102. doi:10.1016/j.jhydrol.2003.10.021.
- [67] Hendricks Franssen HJ, Brunner P, Makobo P, Kinzelbach W. Equally likely inverse solutions to a groundwater flow problem including pattern information from remote sensing images. *Water Resour Res* 2008;44(1). doi:10.1029/2007wr006097. W01419.
- [68] Hendricks Franssen HJ, Kinzelbach W. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resour Res* 2008;44(9). doi:10.1029/2007wr006505. W09408.
- [69] Hernandez AF, Neuman SP, Guadagnini A, Carrera J. Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion. In: ModelCARE 2002 conference, Prague, Czech Republic; 2002.
- [70] 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(5). doi:10.1029/2005wr004449. W05425.
- [71] Hu LY, Le Ravalec M, Blanc G. Gradual deformation and iterative calibration of truncated Gaussian simulations. *Petrol Geosci* 2001;7:S25–30.
- [72] Hu LY. Extended probability perturbation method for calibrating stochastic reservoir models. *Math Geosci* 2008. doi:10.1007/s11004-008-9158-4.
- [73] Hu LY, Chugunova T. Multiple-point geostatistics for modeling subsurface heterogeneity: a comprehensive review. *Water Resour Res* 2008;44(W11413). doi:10.1029/2008WR006993.
- [74] Janssen G, Valstar JR, van der Zee S. Inverse modeling of multimodal conductivity distributions. *Water Resour Res* 2006;42(3). doi:10.1029/2005wr004356. W03410.
- [75] Jenni S, Hu LY, Basquet R, De Marsily G, Bourbiaux B. History matching of a stochastic model of field-scale fractures: methodology and case study. *Oil Gas Sci Technol – Revue De L'Institut Francais Du Petrole* 2007;62(2):265–76.
- [76] Kashyap RL. Optimal choice of AR and MA parts in autoregressive moving average models. *IEEE Trans Pat Anal Mach Intell* 1982;4(2):99–104.
- [77] Keidser A, Rosbjerg D. A comparison of four inverse approaches to groundwater flow and transport parameter identification. *Water Resour Res* 1991;27(9):2219–32.
- [78] Kerrou J, Renard P, Hendricks Franssen HJ, Lunati I. Issues in characterizing heterogeneity and connectivity in non-multi-Gaussian media. *Adv Water Res* 2008;31(1):147–59. doi:10.1016/j.advwatres.2007.07.002.
- [79] Kitanidis PK, Vomvoris EG. A geostatistical approach to the inverse problem in groundwater modeling (steady-state) and one-dimensional simulations. *Water Resour Res* 1983;19(3):677–90.
- [80] Kitanidis PK. Quasi-linear geostatistical theory for inversing. *Water Resour Res* 1995;31(10):2411–9.
- [81] Kool JB, Parker JC, Van Genuchten MT. Parameter estimation for unsaturated flow and transport models. A review. *J Hydrol* 1987;91:255–93.
- [82] 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 Res* 2004;27(6):583–99. doi:10.1016/j.advwatres.2004.03.003.
- [83] Kurapov AL, Egbert GD, Allen JS, Miller RN. Representer-based variational data assimilation in a nonlinear model of nearshore circulation. *J Geophys Res–Ocean* 2007;112(C11). doi:10.1029/2007jc004117. C11019.
- [84] Lappin AR. Summary of site-characterization studies conducted from 1983 through 1987 at the Waste Isolation Pilot Plant (WIPP), Southeastern New Mexico. SAND88-0517. Albuquerque, USA: Sandia National Laboratories; 1988.
- [85] 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.
- [86] Lavenue AM, Ramarao BS, Demarsily G, 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.
- [87] 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.
- [88] Le Loc'h G, Galli A. Truncated plurigaussian method: theoretical and practical points of view. In: Baafi EY, Schofield NA, editors. *Geostatistics wollongong 1996*, vol. 1. Kluwer Academic Publishers; 1997. p. 211–22.
- [89] Li W. *Geostatistical inference of hydrogeological parameters on large scales*. PhD thesis. Technical University of Zurich, Switzerland; 2008.
- [90] Liu GS, Chen Y, Zhang DX. Investigation of flow and transport processes at the MADE site using ensemble Kalman filter. *Adv Water Res* 2008;31(7):975–86. doi:10.1016/j.advwatres.2008.03.006.
- [91] Liu YQ, Gupta HV. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resour Res* 2007;43(7). doi:10.1029/2006wr005756. W07401.
- [92] Marquardt DW. An algorithm for least-squares estimation of non-linear parameters. *J Soc Ind Appl Math* 1963;11:431–41.
- [93] McLaughlin D, Townley LR. A reassessment of the groundwater inverse problem. *Water Resour Res* 1996;32(5):1131–61.
- [94] Medina A, Alcolea A, Carrera J, Castro LF. Flow and transport modelling in the geosphere: the code TRANSIN IV. In: IV Jornadas de investigacion y desarrollo tecnologico de gestion de residuos radiactivos de ENRESA, Barcelona: ENRESA; 2000.
- [95] Medina A, Carrera J. Geostatistical inversion of coupled problems: dealing with computational burden and different types of data. *J Hydrol* 2003;281:251–64.
- [96] Meier PM, Medina A, Carrera J. Geostatistical inversion of cross-hole pumping tests for identifying preferential flow channels within a shear zone. *Ground Water* 2001;39(1):10–7.
- [97] Naevdal G, Johnsen LV, Aanonsen SI, Vefring EH. Reservoir monitoring and continuous model updating using ensemble Kalman filter. *Society of Petroleum Engineers*; 2003. p. 84372.
- [98] Neuman SP, Orr S. Prediction of steady state flow in nonuniform geological media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation. *Water Resour Res* 1993;29(2):341–64.
- [99] Neuman SP, Tartakovsky D, Wallstrom TC, Winter CL. Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation (vol. 29, p. 341, 1993). *Water Resour Res* 1996;32(5):1479–80.
- [100] Neuman SP. Maximum likelihood Bayesian averaging of uncertain model predictions. In: ModelCARE 2002 conference, Prague, Czech Republic; 2002.
- [101] Ngodock HE, Jacobs GA, Chen MS. The representer method, the ensemble Kalman filter and the ensemble Kalman smoother: a comparison study using

- a nonlinear reduced gravity ocean model. *Ocean Modell* 2006;12(3–4):378–400. doi:[10.1016/j.ocemod.2005.08.001](https://doi.org/10.1016/j.ocemod.2005.08.001).
- [102] Ngodock HE, Smith SR, Jacobs GA. Cycling the representer algorithm for variational data assimilation with a nonlinear reduced gravity ocean model. *Ocean Modell* 2007;19(3–4):101–11. doi:[10.1016/j.ocemod.2007.06.004](https://doi.org/10.1016/j.ocemod.2007.06.004).
- [103] Nowak W, Cirpka OA. A modified Levenberg–Marquardt algorithm for quasi-linear geostatistical inverting. *Adv Water Res* 2004;27(7):737–50. doi:[10.1016/j.advwatres.2004.03.004](https://doi.org/10.1016/j.advwatres.2004.03.004).
- [104] Oliver DS, Cunha LB, Reynolds AC. Markov chain Monte Carlo methods for conditioning a log-permeability field to pressure data. *Math Geol* 1997;29(1):61–91.
- [105] Pebesma EJ, Wesseling CG. Gstat: a program for geostatistical modelling, prediction and simulation. *Comput Geosci* 1998;24(1):17–31.
- [106] Poeter EP, Hill MC. Inverse models: a necessary next step in ground-water modeling. *Ground Water* 1997;35(2):250–60.
- [107] Przybysz-Jarnut JK, Hanea RG, Jansen JD, Heemink AW. Application of the representer method for parameter estimation in numerical reservoir models. *Comput Geosci* 2007;11(1):73–85. doi:[10.1007/s10596-006-9035-5](https://doi.org/10.1007/s10596-006-9035-5).
- [108] Ramarao BS, Lavenue AM, Demarsily G, 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.
- [109] Renard P. Stochastic hydrogeology: what professionals really need? *Ground Water* 2007;45(5):531–41. doi:[10.1111/j.1745-6584.2007.00340.x](https://doi.org/10.1111/j.1745-6584.2007.00340.x).
- [110] Rissanen J. Modelling by shortest data description. *Automatica* 1978;14:465–71.
- [111] Riva M, Guadagnini A, De Simoni M. Assessment of uncertainty associated with the estimation of well catchments by moment equations. *Adv Water Res* 2006;29(5):676–91. doi:[10.1016/j.advwatres.2005.07.005](https://doi.org/10.1016/j.advwatres.2005.07.005).
- [112] Rubin Y. Stochastic modeling of macrodispersion in heterogeneous porous media. *Water Resour Res* 1990;26(1):133–41.
- [113] Rubin Y. Prediction of tracer plume migration in disordered porous media by the method of conditional probabilities. *Water Resour Res* 1991;27(6):1291–308.
- [114] Sahuquillo A, Capilla JE, Gomez-Hernandez J, Andreu J. Conditional simulation of transmissivity fields honouring piezometric data. In: *Hydraulic Engineering Software IV, Fluid Flow Modeling*. Kluwer Academic Publishers; 1992.
- [115] Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978;6(2):461–4.
- [116] Smith DR. *Variational methods in optimization*. New Jersey: Prentice-Hall; 1974.
- [117] Smith L, Schwartz FW. Mass transport, 2. Analysis of uncertainty in prediction. *Water Resour Res* 1981;17(2):351–69.
- [118] Stauffer F, Attinger S, Zimmermann S, Kinzelbach W. Uncertainty estimation of well catchments in heterogeneous aquifers. *Water Resour Res* 2002;38(11):1238.
- [119] Stauffer F, Hendricks Franssen HJ, Kinzelbach W. Semianalytical uncertainty estimation of well catchments: conditioning by head and transmissivity data. *Water Resour Res* 2004;40(8). doi:[10.1029/2004wr003320](https://doi.org/10.1029/2004wr003320). W08305.
- [120] Strebelle S. Conditional simulation of complex geological structures using multiple-point statistics. *Math Geol* 2002;34(1):1–21.
- [121] Sun NZ, Yeh WWG. A stochastic inverse solution for transient groundwater flow: parameter identification and reliability analysis. *Water Resour Res* 1992;28(12):1851–60.
- [122] Suzuki S, Caumon G, Caers J. Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. *Comput Geosci* 2008;12(1):105–19. doi:[10.1007/s10596-007-9063-9](https://doi.org/10.1007/s10596-007-9063-9).
- [123] Tarantola A. *Inverse problem theory and methods for parameter estimation*. Philadelphia, USA: SIAM; 2005.
- [124] Tarantola A. Popper, Bayes and the inverse problem. *Nature Phys* 2006;2:492–4.
- [125] Tikhonov AN. Regularization of incorrectly posed problems. *Sov Math Dokl* 1963;4:1624–7.
- [126] Tikhonov AN. Solution of incorrectly formulated problems and the regularization method. *Sov Math Dokl* 1963;4:1035–8.
- [127] Tonkin MJ, Doherty J. A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resour Res* 2005;41(10). doi:[10.1029/2005wr003995](https://doi.org/10.1029/2005wr003995).
- [128] Townley LR, Wilson JL. Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models. *Water Resour Res* 1985;21(12):1851–60.
- [129] Valstar JR. *Inverse problem theory of coupled groundwater flow and mass transport*. PhD thesis. Technical University of Delft, The Netherlands; 2001.
- [130] Valstar JR, McLaughlin DB, Stroet C, van Geer FC. A representer-based inverse method for groundwater flow and transport applications. *Water Resour Res* 2004;40(5). doi:[10.1029/2003wr002922](https://doi.org/10.1029/2003wr002922).
- [131] Vazquez-Sune E, Capino B, Abarca E, Carrera J. Estimation of recharge from floods in disconnected stream-aquifer systems. *Ground Water* 2007;45(5):579–89. doi:[10.1111/j.1745-6584.2007.00326.x](https://doi.org/10.1111/j.1745-6584.2007.00326.x).
- [132] Vermeulen PTM, Heemink AW, Valstar JR. Inverse modeling of groundwater flow using model reduction. *Water Resour Res* 2005;41(6). doi:[10.1029/2004wr003698](https://doi.org/10.1029/2004wr003698).
- [133] 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.
- [134] Vrugi JA, Gupta HV, Bouten W, Sorooshian S. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour Res* 2003;39(8). doi:[10.1029/2002wr001642](https://doi.org/10.1029/2002wr001642).
- [135] Vrugi JA, Diks CGH, Gupta HV, Bouten W, Verstraten JM. Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. *Water Resour Res* 2005;41(1). doi:[10.1029/2004wr003059](https://doi.org/10.1029/2004wr003059).
- [136] Wen XH, Tran TT, Behrens RA, Gomez-Hernandez JJ. Production data integration in sand/shale reservoirs using sequential self-calibration and geomorphing: a comparison. In: *2000 SPE annual technical conference and exhibition*, Dallas, Texas: Soc Petroleum Eng; 2000.
- [137] Woodbury AD, Ulrych TJ. A full-Bayesian approach to the groundwater inverse problem for steady state flow. *Water Resour Res* 2000;36(8):2081–93.
- [138] Ye M, Meyer PD, Neuman SP. On model selection criteria in multimodel analysis. *Water Resour Res* 2008;44(3). doi:[10.1029/2008wr006803](https://doi.org/10.1029/2008wr006803).
- [139] Yeh TCJ, Jin MH, Hanna S. An iterative stochastic inverse method: conditional effective transmissivity and hydraulic head fields. *Water Resour Res* 1996;32(1):85–92.
- [140] Yeh WWG. Review of parameter identification procedures in groundwater hydrology. *Water Resour Res* 1986;22(2):95–108.
- [141] Zhang D. *Stochastic methods for flow in porous media: coping with uncertainties*. San Diego, USA: Academic Press; 2002.
- [142] Zimmerman DA, de Marsily G, Gotway CA, Marietta MG, Axness CL, Beauheim RL, et al. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. *Water Resour Res* 1998;34(6):1373–413.