Published in Water Resources Reserach 46, W08511 1-18, 2010
which should be used for any reference to this work

Blocking Moving Window algorithm: Conditioning multiple-point simulations to hydrogeological data

Andres Alcolea[1,2] and Philippe Renard[1]

[1] Connectivity constraints and measurements of state variables contain valuable information on aquifer architecture. Multiple-point (MP) geostatistics allow one to simulate aquifer architectures, presenting a predefined degree of global connectivity. In this context, connectivity data are often disregarded. The conditioning to state variables is usually carried out by minimizing a suitable objective function (i.e., solving an inverse problem). However, the discontinuous nature of lithofacies distributions and of the corresponding objective function discourages the use of traditional sensitivity-based inversion techniques. This work presents the Blocking Moving Window algorithm (BMW), aimed at overcoming these limitations by conditioning MP simulations to hydrogeological data such as connectivity and heads. The BMW evolves iteratively until convergence: (1) MP simulation of lithofacies from geological/geophysical data and connectivity constraints, where only a random portion of the domain is simulated at every iteration (i.e., the blocking moving window, whose size is user-defined); (2) population of hydraulic properties at the intrafacies; (3) simulation of state variables; and (4) acceptance or rejection of the MP simulation depending on the quality of the fit of measured state variables. The outcome is a stack of MP simulations that (1) resemble a prior geological model depicted by a training image, (2) honor lithological data and connectivity constraints, (3) correlate with geophysical data, and (4) fit available measurements of state variables well. We analyze the performance of the algorithm on a 2-D synthetic example. Results show that (1) the size of the blocking moving window controls the behavior of the BMW, (2) conditioning to state variable data enhances dramatically the initial simulation (which accounts for geological/geophysical data only), and (3) connectivity constraints speed up the convergence but do not enhance the stack if the number of iterations is large.

1. Introduction

[2] Reliable hydrogeological models demand accurate characterizations of aquifer heterogeneity and a quantification of the inherent uncertainties. To that end, available data are integrated in stochastic inverse models. In broad terms, inverse modeling refers to the process of gathering information about the model and its parameters from measurements of what is being modeled [*Carrera et al.*, 2005]. These include direct measurements of aquifer properties and indirect observations of dependent state variables such as heads or concentrations. Inverse modeling has become an important focus of research, as revealed by the numerous state-of-the-art publications [*Carrera*, 1987; *Carrera et al.*, 2005; *de Marsily et al.*, 1999; *Hendricks Franssen et al.*, 2009; *Kool et al.*, 1987; *McLaughlin and Townley*, 1996; *Yeh*, 1986]. Inverse methods incorporate direct models that interpolate measurements of the property under study (e.g., measurements of hydraulic conductivity arising from prior interpretation of pumping tests) and can account also for other correlated variables such as geophysics. Traditionally, inverse methods make use of conditional estimation and simulation tools. Examples are variants of (co-)kriging and sequential Gaussian simulation. These techniques rely on two-point geostatistical measures such as variograms or covariances and assume that the joint distribution of aquifer parameters is multi-Gaussian. This is motivated by the fact that, first, histograms of field data (or some suitable transform of them) often display a Gaussian character and, second, multi-Gaussian models are parsimonious (i.e., simply defined by an expected value and a covariance function). Unfortunately, univariate Gaussian distributions of aquifer parameters do not guarantee a joint multi-Gaussian distribution [*Gomez-Hernandez and Wen*, 1998]. To make things worse, Gaussian models maximize entropy. Therefore, they fail to reproduce connectivity features of aquifer properties because the spatial continuity of their extreme values is minimum [*Journel and Deutsch*, 1993; *Zinn and Harvey*, 2003]. Thus, in general, methods based on two-point statistics are not capable of reproducing realistic geological structures presenting discrete features (e.g., channels, fractures, etc.) and/or characterized by curvilinear "crispy" geometries such as

[1] Centre for Hydrogeology and Geothermics, University of Neuchâtel, Neuchâtel, Switzerland.
[2] Now at TK Consult AG, Zurich, Switzerland.

alluvial fans or braided channels [*Falivene et al.*, 2006; *Kerrou et al.*, 2008; *Knudby and Carrera*, 2005].

[3] Discrete geological features cause abrupt changes in aquifer properties. These large "jumps" over small distances exert a major control on groundwater flow and, especially, on contaminant transport [*Alcolea et al.*, 2008; *Knudby and Carrera*, 2005; *Trinchero et al.*, 2008]. The influence of the spatial variability of aquifer properties at a smaller scale (i.e., within lithofacies) is also dominant, but to a lesser extent. Those discrete features must be modeled explicitly even if their exact geometry and position is not well-known. In some specific cases, the identification of such highly connected patterns is possible using traditional multi-Gaussian techniques in an inverse problem framework if, first, their geometry is not very sophisticated and, second, the information contained in indirect observations (e.g., heads or concentrations) is sufficiently adequate to identify connectivity patterns [*Alcolea et al.*, 2006a, 2006b; *Meier et al.*, 2001]. To overcome these limitations, some inversion methods incorporate stochastic direct models capable of handling abrupt changes in aquifer properties explicitly. Such direct models were reviewed by *de Marsily et al.* [2005]. Examples are the Boolean method [*Haldorsen and Lake*, 1984], the sequential indicator simulation method [*Gomez-Hernandez and Srivastava*, 1990], the truncated pluri-Gaussian method [*Le Loc'h and Galli*, 1997], the transition probabilities method [*Carle and Fogg*, 1997] and multiple-point (MP hereinafter) geostatistics [*Guardiano and Srivastava*, 1993; *Strebelle*, 2002]. *Falivene et al.* [2006] applied most of the aforementioned techniques to a turbidite sandstone deposit. Inversion methods not assuming a multi-Gaussian model include the conditional probabilities method [*Capilla et al.*, 1999] and the gradual deformation method, that has been applied to truncated pluri-Gaussian [*Hu et al.*, 2001] and Boolean models [*Jenni et al.*, 2007]. The sequential self-calibration method [*Gomez-Hernandez et al.*, 1997] was extended to invert lithofacies distributions from state variable data within the framework of truncated Gaussian simulation [*Wen et al.*, 2000].

[4] Among the aforementioned techniques, MP geostatistics presents the ability of simulating a wide variety of geological structures and an accurate conditioning to local data because of its pixel-based nature [*Xu*, 1996]. This technique became important after the work by *Strebelle* [2002], who developed the first efficient MP algorithm, based on the single normal equation (*snesim*). Recently, *Hu and Chugunova* [2008] published the first detailed state-of-the-art on MP geostatistics. Lithofacies maps are simulated using multipoint probability distributions conditioned to the heterogeneity patterns displayed in the surrounding data. These probability distributions are borrowed from a training image which depicts a (prior) conceptual geological model defining the expected patterns of geological heterogeneity. As such, training images need not to honor any data. Several training images can be used at a time, depicting different patterns of heterogeneity, possibly at different observation scales. Training images are often borrowed from aerial photos, outcrop observation or from the (sometimes subjective) decision of a modeler. *Caers and Zhang* [2004] address different ways for generating stationary training images. Recent works by *Chugunova and Hu* [2008] and *de Vries et al.* [2009] address the problem of nonstationarity. MP geostatistical techniques have been successfully applied to a number of case studies including petroleum engineering and hydrogeology [*Caers et al.*, 2003; *Feyen and Caers*, 2006; *Liu et al.*, 2004; *Strebelle*, 2006]. Those works were devoted to the direct simulation of geological structures accounting only for direct observations of lithology at available wells and correlated soft data arising, for example, from geophysics. However, dependent state variables such as heads and geometrical connectivity constraints contain valuable information on aquifer architecture and must be accounted for in meaningful hydrological models. Geometrical connectivity data (i.e., point "*i*" connected with point "*j*") can be accommodated in MP simulations [*Allard*, 1994; *Renard and Caers*, 2008]. Usually, the connectivity information is not geometric in a strict sense, but extracted from state variable measurements, such as heads (flow connectivity) or concentrations (transport connectivity from tracer breakthrough curves). *Knudby and Carrera* [2005] and *Trinchero et al.* [2008] present comparisons between different measures of connectivity. Conditioning to state variable data is often carried in the framework of inverse problem.

[5] In the context of inverse problem, MP geostatistics has been embedded in the Probability Perturbation Method, PPM hereinafter [*Caers*, 2003; *Caers and Hoffman*, 2006; *Hu*, 2008]. The PPM is an iterative technique. Each iteration consists of perturbing the current simulation (or the initial) so that the new solution fits state variable data better than the previous one. The degree of perturbation is controlled by a single parameter r_D ranging from 0 (no perturbation) to 1 (a completely new MP simulation). r_D is used to modify a local probability map which is then combined with prior probabilities of the initial MP simulation using the tau model [*Journel*, 2002]. This parameter is optimized at each iteration. *Ronayne et al.* [2008] presented the first hydrogeological application of this method. That work is aimed at identifying the position and geometry of a discrete channel deposit in the alluvial fan system that underlies the Lawrence Livermore National Laboratory (U.S.) from dynamic data arising from pumping tests. Their results suggest that the objective function being minimized is not, in general, a continuous function of r_D. Discontinuities may appear when, at a given iteration, some high or low hydraulic conductivity structures (e.g., sand channels or shale lenses) are suddenly connected or disconnected. This feature is intrinsic to inverse problems with a discrete distribution of lithofacies and can hinder the identification of the optimum value of r_D. A second difficulty is inherent to the use of the tau model, because the posterior distribution (conditioned to state variables) depends on the value of the coefficient used in the tau model [*Caers*, 2007]. Current techniques are not able to estimate a priori the optimum value of that coefficient. An alternative, also embedding MP geostatistics, consists of considering the inverse problem as a search problem [*Suzuki and Caers*, 2006; *Suzuki et al.*, 2008], as opposed to the traditional sensitivity-based optimization. These authors consider a large ensemble of MP simulations of the reservoir conditioned to geological data only and use search methods (e.g., Nearest Neighborhood) to select those that match state variable data. To speed up the search, they consider a modified similarity Hausdorff distance [*Dubuisson and Jain*, 1994] between the initial simulations. This approach is appealing because it considers several prior geological models defining different structures of the reservoir (i.e., generated with different training images). As such, prior

conceptual uncertainties are alleviated. Yet, this search method requires the prior generation of all possible structures. As such, new patterns of heterogeneity cannot be included as the search process evolves.

[6] Markov chain Monte Carlo methods (McMC hereinafter) are an alternative to overcome the difficulties of the sensitivity-based minimization of discontinuous objective functions [*Balakrishnan et al.*, 2003; *Fu and Gomez-Hernandez*, 2008, 2009; *Liu and Oliver*, 2003; *Oliver et al.*, 1997; *Robert and Gasella*, 1999; *Tarantola*, 2005; *Zanini and Kitanidis*, 2009]. These techniques sample the posterior distribution of parameters (in a Bayesian sense) by, first, drawing samples of the prior distribution and, second, accepting or rejecting them on the basis of a Metropolis-Hastings type criterion. The latter accounts for the fit to measurements not accounted for in the sampling of the prior distribution. These "derivative-free" approaches are also appealing because they do not assume any particular distribution of parameters or any linearization. In addition, they guarantee (by construction) that the samples of the posterior pdf are also samples of the prior. However, these methods are generally CPU time consuming owing to the large number of samples of the prior that have to be tested. This limits their applicability to problems with a small number of unknown parameters [*Balakrishnan et al.*, 2003; *Hassan et al.*, 2009; *Liu and Oliver*, 2003]. Fortunately, efficient algorithms that allow the inference of highly parameterized fields have been proposed recently [*Fu and Gomez-Hernandez*, 2009; *Holloman et al.*, 2006; *Zanini and Kitanidis*, 2009] or are developed nowadays.

[7] In an attempt to overcome some of the aforementioned difficulties, we present the Blocking Moving Window algorithm (BMW hereinafter), which combines MP geostatistics and forward simulation (as opposed to traditional sensitivity-based inversion) of state variables. The derivative-free character of the BMW alleviates the problems caused by the discontinuous nature of the objective function. The algorithm is similar to the one devised by *Fu and Gomez-Hernandez* [2009] for multi-Gaussian fields, the main novelties being (1) the inclusion of a MP simulator as direct model instead of the sequential Gaussian simulator in the work by *Fu and Gomez-Hernandez* [2009] and (2) the inclusion of a cooling schedule in the acceptation/rejection process. The MP simulator allows one to reproduce non-Gaussian fields presenting connectivity features and "crispy" geometries regardless of the information contained in the state variables data set. The cooling schedule makes our method to deviate from the standard McMC approach (actually, the BMW contains the basic "ingredients" of Simulated Annealing, SA hereinafter [*Kirkpatrick et al.*, 1983]). Yet, it makes the iterative process less random and, therefore, more efficient. The BMW is described in detail and applied to a synthetic example mimicking the groundwater flow to a well in a channelized geological scenario. We investigate the role of the different conditioning data sets (lithology and head at selected observation wells and connectivity constraints). Results show that conditioning to heads enhances dramatically the initial MP simulation (i.e., accounting for geological/geophysical information only). The use of connectivity constraints does not improve the reliability of the characterization but enhances the convergence of the algorithm. This paper is organized as follows. First, the workflow of the BMW is presented in detail. Second, the synthetic example is described. The paper ends with a discussion of the results and some conclusions about the use of hydrogeological data to condition MP geostatistical simulations.

2. Methodology

[8] The BMW algorithm consists of four main steps, repeated iteratively until convergence. First, a MP conditional simulation of the lithofacies distribution is drawn by perturbing a previous state. This simulation is consistent with a prior geological conceptualization depicted by a suitable training image (i.e., high-order statistics are borrowed from it). As such, it belongs to the prior space of plausible geological models. It is also anchored to hard data of lithofacies at observation wells and correlates with soft data. Connectivity constraints (i.e., point "*i*" connected with point "*j*") can also be used as conditioning data for the MP simulation. Only a portion of the domain, selected randomly (i.e., the blocking moving window), is simulated. This makes the iterative process less random and therefore, more efficient [*Fu and Gomez-Hernandez*, 2009]. Prior conditioning to connectivity data [*Renard and Caers*, 2008] also enhances the convergence rate. Second, hydraulic properties at the intrafacies are populated. Third, a forward simulation of state variables is carried out and a suitable objective function is calculated. This function measures the misfit between calculated and measured state variables. If the objective function is smaller than a user-supplied threshold value, the MP simulation becomes a stack member regardless of the stage of the iterative process (i.e., this can occur at very early iterations). Finally, the simulation is accepted or rejected according to a "survival" probability sampled from a Gibbs's distribution. The final outcome is a stack of simulations that (1) belong to the prior space of plausible lithofacies distributions (i.e., they resemble the training image), (2) are anchored to lithofacies and connectivity constraints, (3) correlate with soft data, and (4) fit measurements of state variables well. Note that at a given iteration, the MP simulation drew at step 1 is not at all perturbed. Thus, MP simulations populating the stack belong to the prior space of plausible geological models (i.e., unconditional, resembling the prior geological model only) and to the posterior subspace, conditioned to available data. Details on these basic steps at a generic iteration "*k*" are provided below.

2.1. Step 1: Multiple-Point Conditional Simulation of Lithofacies

[9] Step 1.1 is the random generation of a squared blocking moving window. The location of the center is randomly drawn (i.e., with uniform probability). In this work, the window size is supplied by the user. At first iteration, the window encompasses the whole simulation domain. The window size is the parameter that exerts a major control on the algorithm behavior. Early approaches in the McMC literature modified a single element per iteration [*Oliver et al.*, 1997]. This procedure was proved effective but inefficient owing to the large number of iterations required.

[10] Step 1.2 is the generation of the conditioning data set. We only simulate elements encompassed by the blocking moving window. The remaining elements (simulated at the last accepted iteration; see step 4) are considered as conditioning data. Available lithofacies data within the window

and all the connectivity constraints are also included in the conditioning data set.

[11] Step 1.3 is the multiple-point conditional simulation of a test lithofacies distribution, L^k_{test}. This resembles the variability patterns depicted by the training image (or images). As such, L^k_{test} belongs to the prior space of plausible geological models. It is already conditioned to the aforementioned hard data sets and correlates with soft data arising, for example, from geophysics. In this work, we have used *Impala*, our own MP kernel (J. Straubhaar et al., Multi-point reservoir modeling, submitted to *Mathematical Geosciences*, 2010). Yet, it is worth to mention that our implementation of the BMW is general and other codes for stochastic simulation (MP or not) can be accommodated at this step.

[12] Step 1.4 is the prior rejection of L^k_{test}. The forward simulation of state variables is often CPU-intensive. Thus, prior to solving the forward problem, a lithofacies distribution can be rejected if it does not resemble other sources of geological or geometrical information that cannot be included as conditioning data at step 1.3. If L^k_{test} is rejected at this stage, algorithm recalls step 1.1 (i.e., a new lithofacies distribution is drawn). Prior rejection is decided by the so-called "proxies," that is, criteria based on a fast evaluation of L^k_{test}. Among others, available proxies implemented in the BMW are (1) proportions of the training image (L^k_{test} is rejected at this stage if the proportions of the different lithofacies do not fit the corresponding proportions at the training image or user defined ranges) and (2) global and directional connectivity features of the training image (L^k_{test} is rejected if it does not resemble the global and directional connectivity depicted by the training image).

2.2. Step 2: Population of Hydraulic Properties

[13] Hydraulic properties are populated at the intrafacies defined by L^k_{test}. Options implemented in the BMW include constant values and variants of conditional estimation (i.e., of kriging/cokriging) and sequential Gaussian (co-) simulation. In the latter cases, a geostatistical model (e.g., a suitable variogram) defining the heterogeneity within each lithofacies is required.

2.3. Step 3: Simulation of Dependent State Variables

[14] Dependent state variables are obtained by solving numerically the equations of the forward problem. In this work, we use TRANSIN [*Medina and Carrera*, 2003], a finite element code for the simulation of groundwater flow and contaminant transport. Still, simulators of other state variables can be accommodated easily. This confers flexibility to the implementation of the BMW.

2.4. Step 4: Posterior Rejection

[15] The goal is to obtain a stack of MP simulations that fit measurements of state variables well. To that end, we evaluate an objective function F^k_{test} measuring the mismatch between calculated and measured state variables:

$$F^k_{test}\left(L^k_{test}\right) = \sum_{i=1}^{N_{type}} \mu_i \sum_{j=1}^{n_i} \omega_j \left(c_{i,j}\left(L^k_{test}\right) - o_{i,j}\right)^2, \qquad (1)$$

where N_{type} is the number of types of observations (e.g., $i = 1$ for heads, $i = 2$ for concentrations, etc.); n_i is the number of observations of type i; $c_{i,j}$ and $o_{i,j}$ are the calculated and measured values, respectively; and ω_j and μ_i are balance weights associated with measurement "j" of type "i," respectively. The posterior acceptance (or rejection) of the MP simulation drew at step 1, L^k_{test}, is decided upon by a Metropolis type criterion [*Metropolis et al.*, 1953]. If F^k_{test} is smaller than the previous accepted one (termed F_{prev}, with corresponding lithofacies distribution L_{prev}), L^k_{test} is termed superior and is accepted directly. Otherwise, the simulation is termed inferior and can still be accepted according to a survival probability sampled from the Gibbs distribution:

$$p = \min\left\{1,\ \exp\left(-\frac{F^k_{test} - F_{prev}}{T^k}\right)\right\}, \qquad (2)$$

where p is the probability of accepting an inferior MP simulation and T^k is the so-called annealing temperature. If L^k_{test} is accepted, the system is updated (i.e., $L_{prev} = L^k_{test}$ and $F_{prev} = F^k_{test}$). L^k_{test} is added to the stack if F^k_{test} is smaller than a user-defined threshold value. This measures the required extent to which model outputs are in agreement with their measured counterparts. Thus, it defines the greed of the final stack (i.e., a greedy tolerance will lead to a reduced stack). The annealing temperature T is controlled by the cooling schedule. A key issue in the definition of the cooling schedule is the cooling rate. A fast cooling may freeze the algorithm in a local minimum of the objective function. A slow cooling can make the computational effort unaffordable. A number of cooling schemes can be applied [*Van Laarhoven and Aarts*, 1987], the most widely spread being the geometric progression:

$$T^{k+1} = \alpha T^k \ ;\ \text{if mod}\ (k, n_{cycle}) = 0, \qquad (3)$$

where α is a scalar in the range (0,1) and n_{cycle} is the number of iterations per temperature step. In this work, we adopted $\alpha = 0.9$, which behaves correctly in most cases [*Alcolea et al.*, 2000; *Van Laarhoven and Aarts*, 1987]. At first iteration, the annealing temperature must be very high (and, correspondingly, the probability of accepting inferior distributions). This allows a broad scanning of the prior space of geologically coherent lithofacies distributions. As the iterative process evolves, T diminishes and we only accept inferior distributions whose objective function does not depart too much from the previous accepted one. The annealing temperature T is lowered after a number of iterations n_{cycle}. At the end of the iterative process (T small), the probability defined in equation (2) is very small. As such, MP simulations presenting large departures from the best one (in theory, the global minimum) are not accepted.

2.5. Step 5: Convergence Check

[16] The aforementioned steps are repeated iteratively until convergence is achieved. The three traditional convergence criteria have been encoded in our implementation. These are (1) a small value of the objective function, (2) target stack size, and (3) maximum number of iterations. An important issue to guarantee the convergence of the BMW is the number of iterations. This must be large enough to ensure the stationarity of the stack. Under such conditions, the addition of new members does not change the statistical moments of the stack. Two additional convergence

checks analyze the stationarity of the process, as measured by the evolution of the average and the variance of the ensemble mean (denoted by $\langle L_{stack} \rangle$ and σ^2_{stack}, respectively):

$$\Delta\langle L_{stack}\rangle = \frac{\left|\langle L_{stack}\rangle^{j+1} - \langle L_{stack}\rangle^{j}\right|}{\left|\langle L_{stack}\rangle^{j+1}\right|} \qquad \Delta\sigma^2_{stack} = \frac{\left|\sigma^{2\,j+1}_{stack} - \sigma^{2\,j}_{stack}\right|}{\left|\sigma^{2\,j+1}_{stack}\right|} \tag{4}$$

Here, j denotes the number of stack members. The convergence criteria in equation (4) are calculated whenever a new member is added to the stack. The BMW stops when either $\Delta\langle L_{stack}\rangle$ or $\Delta\sigma^2_{stack}$ reach threshold values during a number of stack updates (i.e., when $\langle L_{stack}\rangle$ or σ^2_{stack} reach a plateau).

[17] Note that the use of a cooling schedule makes the BMW to deviate slightly from McMC methods. These have been proved to sample the posterior pdf properly. The BMW resembles more a mixture of SA, aimed at obtaining the maximum likelihood point of the pdf [*Tarantola*, 2005], and a genetic algorithm. As such, the goal of the BMW is not to sample accurately the posterior pdf but to obtain a stack of MP simulations that (1) belong to the prior space of plausible geological models and (2) honor or fit well available data. The typical behavior of the BMW is as follows. At early iterations, when the temperature is sufficiently high, the BMW explores broadly the prior space of plausible geological models (i.e., those resembling the prior geological model) and adds to the stack MP simulations that correspond to potential local minima of the posterior pdf. On the contrary, the BMW is greedy at late stages. Then, the stack is updated with MP simulations that do not deviate much from the optimum one (theoretically, the global minimum). As such, the stack is made of a cluster of members that are similar to the optimum MP simulations plus several "isolated" members that are different and that were attained at early to intermediate stages of the iterative process. In any case, it is worth to mention that the BMW is a McMC exhaustive sampler of the posterior pdf if the cooling schedule is removed (i.e., $\alpha = 1$ in equation (3)).

3. Application

[18] The objective of this application is threefold. First, we illustrate the performance of the BMW. Second, we explore the role of the different conditioning data sets (lithofacies and head data at selected observation wells and connectivity constraints). Third, we analyze the sensitivity of the algorithm to its dominant parameter, that is, the size of the blocking moving window. Results are explored on the basis of a synthetic example depicting groundwater flow toward a well in a channelized (binary) geological scenario.

[19] Figure 1 summarizes the geological setup. A domain of 100 × 100 m^2 is discretized in 100 × 100 square finite elements of 1 × 1 m^2. A training image [*Strebelle*, 2002] depicts the expected patterns of heterogeneity (Figure 1a). These are highly connected sand channels presenting pseudo-horizontal orientation, embedded in a matrix made of shale. Information contained in the training image includes (1) proportions of sand and shale, (2) channel width and orientation, and (3) channel tortuosity. Note that the size of the training image (250 × 250 m^2, also discretized in 1 × 1 m^2 elements) is larger than the simulation domain. Simulations are conditioned to nine lithofacies measurements only (six at sand and three at shale; Figure 1c). Seismic data (Figure 1d) are included as correlated, but not conditioning, exhaustive soft information. The training image for this secondary variable is depicted in Figure 1b. Figure 1e displays the ensemble mean of 100 plausible MP lithofacies simulations conditioned to lithofacies data and correlated with soft data. Working only with two lithofacies allows one to interpret Figure 1e as the probability to find sand (p = 1) or shale (p = 0) at a given element. The stack variance represents the related uncertainty (Figure 1f). The small amount of lithofacies data is not sufficient to delineate a well-defined sequence of channels. Small clusters of elements with probability of 1 are found only at the vicinity of the measurement locations at the sand formations (Figure 1e). Therefore, the related uncertainties are large, as depicted in Figure 1f. The uncertainty at the vicinity of observation wells O7 and O8 is very small. This can be explained by a strong correlation with soft data in these regions. Connectivity constraints are not used to generate these lithofacies distributions.

[20] The reference lithofacies distribution (Figure 2d) is randomly selected among the aforementioned 100 MP plausible simulations. Groundwater flow is expected to be governed by the spatial distribution of lithofacies and, to a lesser importance, by the spatial variability of hydraulic conductivity at the intrafacies. Thus, hydraulic conductivity is assumed to be constant (10 and 10^{-3} m/d for sand and shale, respectively). These values cause a large contrast in the corresponding reference head field. For simplicity, we assume steady state groundwater flow. Yet, the methodology is general and also applies to transient problems. A well PW, tapping a sandy channel, pumps 21.6 m^3/d at coordinates (45, 45). Additionally, a regional steady state groundwater flow is simulated by prescribing heads at left and right boundaries (1 and 0 m, respectively). Upper and lower boundaries are impervious. Figure 2d displays the reference heads corresponding to this setup. As one can see, the effects of the pumping and of the channels are noticeable. Reference heads are collected at the nine available wells depicted in Figure 1c. Connectivity constraints are also extracted from the reference lithofacies distribution. These are the connections of pumping well PW to observation wells O1 to O4. Thus, we do not forbid a possible connection of observation well O5 with PW and O1 to O4 (a light gray channel connects the lower and central sets of channels; Figure 1e). Obviously, observation wells at different lithofacies cannot be connected.

4. Results

[21] The performance of the method is evaluated both qualitatively and quantitatively. From the qualitative point of view, we compare the lithofacies and head distributions with the corresponding reference counterparts. In addition, we analyze the mean and the variance of the final stack. In this work, the stack is made of lithofacies distributions leading to values of the objective function smaller than 0.1. From equation (1), this is equivalent to an average head residual of 0.105 m, a small value compared to the range of reference heads [−5.5, 1]. Weights in equation (1) were all set to 1.0. A total of 24 cases are solved, varying the size of

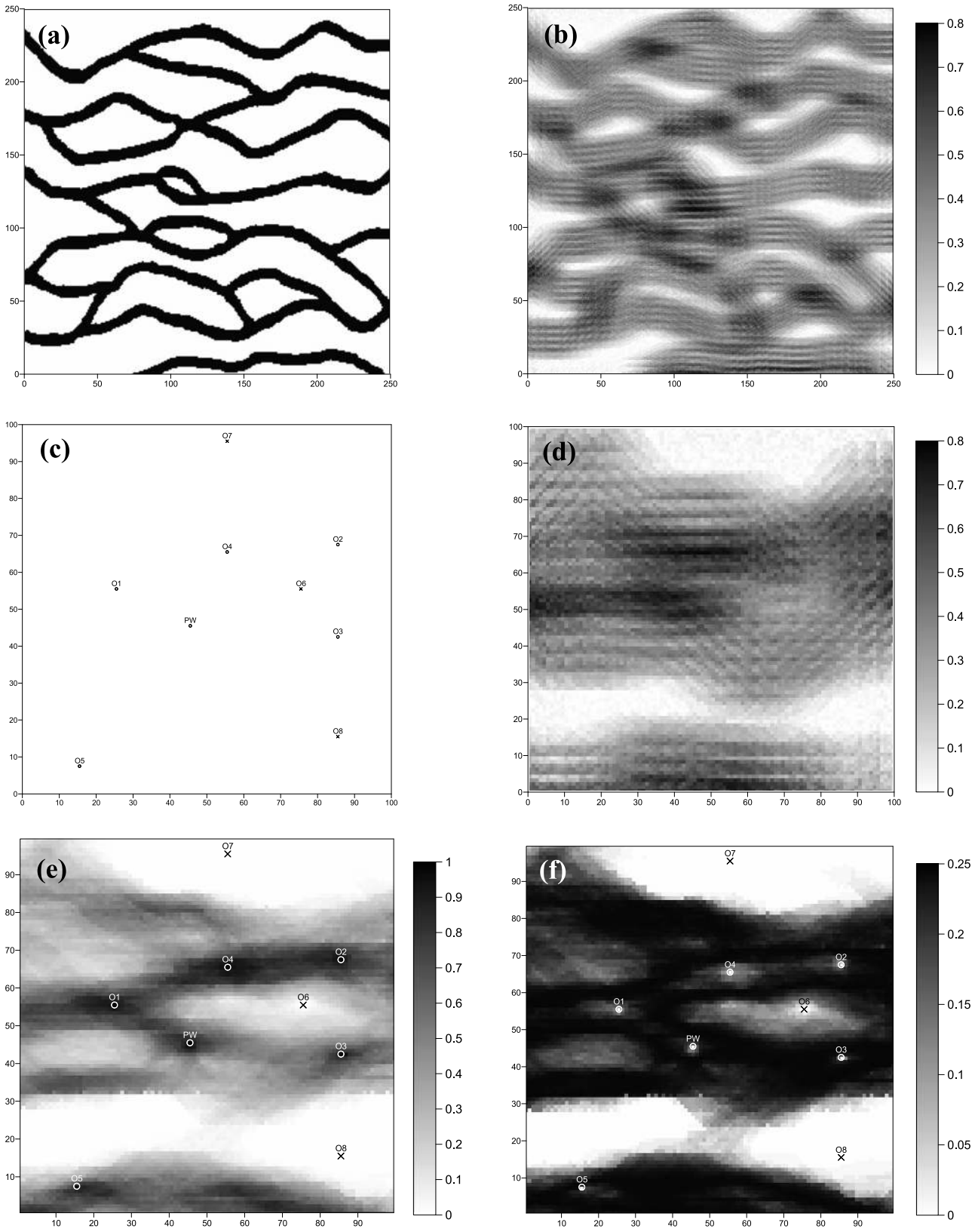

Figure 1. Geological setup. (a) Training image of the primary variable [*Strebelle*, 2002]. (b) Training image of the secondary variable (seismic data). (c) Measurement locations at sand and shale lithofacies (circles and crosses, respectively). (d) Exhaustive seismic data. (e) Ensemble mean of 100 plausible simulations. (f) Corresponding ensemble variance.

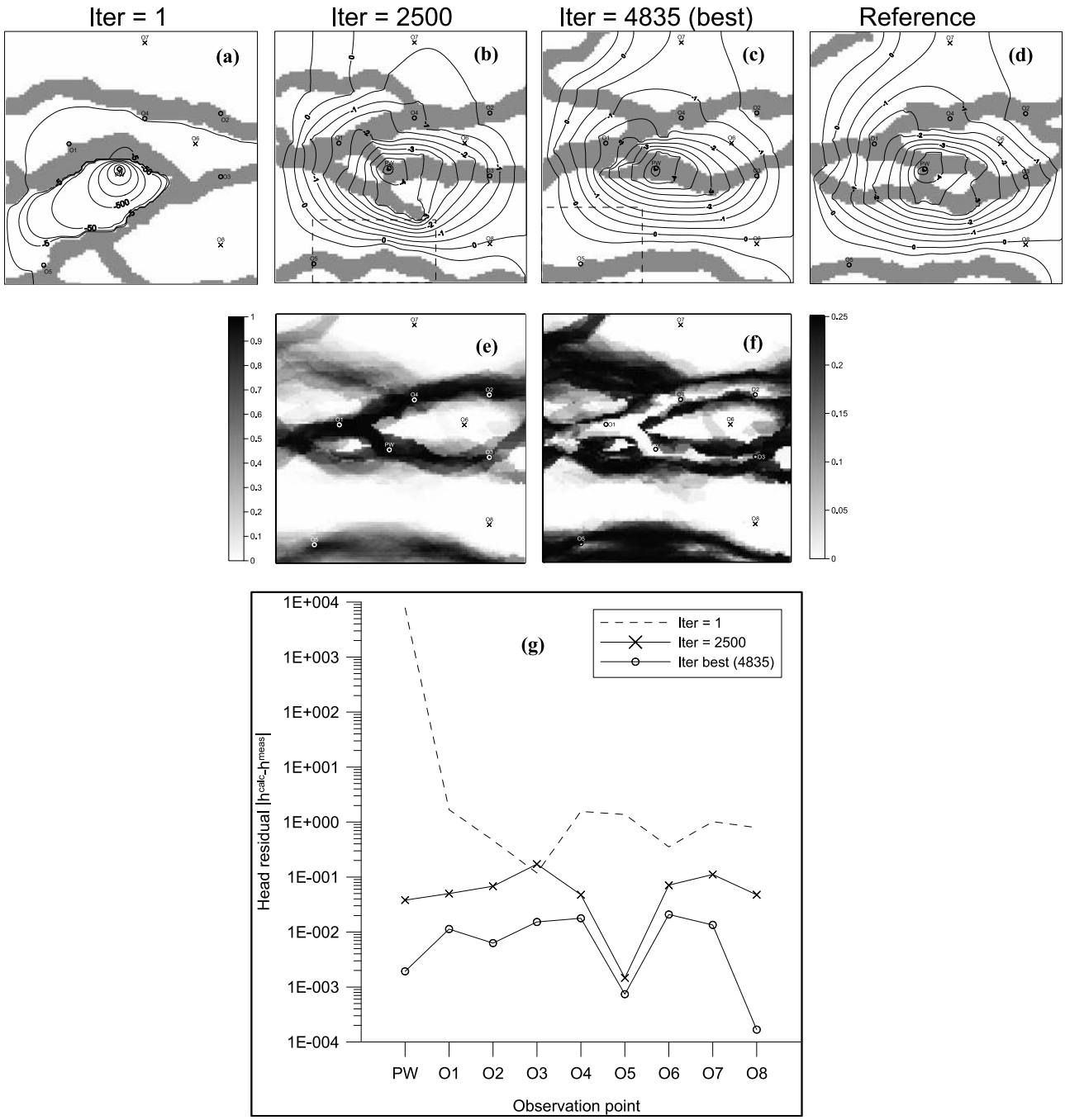

Figure 2. Algorithm evolution in the absence of connectivity constraints. (a–c) Lithofacies distributions and corresponding heads attained at early, intermediate, and final stages of the algorithm (iterations 1, 2500, and 4835, respectively; the latter leads to the minimum objective function). The blocking moving window is depicted by a dashed line. At first iteration it encompasses the whole simulation domain. (d) Reference lithofacies distribution and reference head field. (e) Mean of the final stack. This can be interpreted as the probability to find sand (p = 1) or shale (p = 0) at a given element. (f) Corresponding variance (i.e., related uncertainty). (g) Evolution of head residuals (absolute value of calculated minus measured values) at measurement locations.

the blocking moving window (12 sizes uniformly distributed, ranging from 8 to 96 elements) and the conditioning data (heads only or heads and connectivity constraints). Secondary variables are used as correlated soft data and lithofacies at measurement locations are conditioning data in all cases. Prior rejection criteria (see step 1.4 in section 2) are not used in this work. Thus, we will always refer to the posterior acceptance or rejection of lithofacies distributions described in the step 4 of the methodology. For a given run, the size of the blocking moving window is constant throughout the iterative process. From the quantitative point of view, we analyze the following statistics.

[22] 1. We analyze the objective function F_{test}. A small value of this measure guarantees a good average fit of available head measurements.

[23] 2. We analyze the minimum objective function at iteration "k":

$$F^k_{min} = \min_{i=1,k} F^i_{test}. \quad (5)$$

[24] 3. We analyze the cumulative sums of accepted and rejected lithofacies distributions at a given iteration and the size of the final stack. These criteria report about the quality of the convergence.

[25] 4. We analyze the head residuals at measurement locations (i.e., absolute value of the difference between calculated and measured heads) and the root mean square error ($RMSE_h$) of heads. The latter does not consider the errors at measurement locations only but measures the global reproduction of the reference head field:

$$RMSE_h = \left(\frac{1}{N_n} \sum_{i=1}^{N_n} \left| h_i^{calc} - h_i^{ref} \right|^2 \right)^{1/2}, \quad (6)$$

where N_n denotes the number of nodes (10201 in this case); h^{calc} and h^{ref} are the calculated and reference heads at mesh nodes, respectively.

[26] 5. We analyze the proportion of elements matching the reference lithofacies distribution.

[27] First, we show the performance of the algorithm in the absence of connectivity constraints. To this end, we select a constant, arbitrary, intermediate window size of 48 elements. A large number of iterations (5000) is considered as the only convergence check. This allows us to gain insights into the late behavior of the algorithm. Nonetheless, the stationarity of the process is also analyzed. Figure 2 displays three snapshots of the algorithm at initial, intermediate, and late stages. The first observation that becomes apparent is the strong effect of the conditioning to head data. The initial lithofacies distribution (Figure 2a) was selected purposefully among the plausible simulations in such a way that the pumping well was isolated from the rest of observation wells. Thus, the initial lithofacies distribution is far different from the reference one (Figure 2d). In addition, observation wells O2 and O4 are disconnected from O1 and O3 and observation well O5 is connected with O1 and O3. As expected, simulated heads do not resemble the reference head field (Figure 2d) and residual heads are very large (Figure 2g, dashed line). Consequently, the initial objective function is very large (~$6 \cdot 10^7$; Figure 3) mainly owing to the large residual at the pumping well. Such a "bad" initial distribution was selected because we wanted to test the ability of the algorithm independently of the goodness of its initial state.

[28] At some early stage of the algorithm (iteration 10, not displayed here) the blocking moving window encompasses the pumping well and its connection with observation wells O1 to O4 is attained. This causes a dramatic drop in the objective function (from $6 \cdot 10^7$ to $2 \cdot 10^0$; Figure 3). Yet, this connection might be broken again at any iteration. Consequently, very large values of F_{test} can be attained throughout the process. Corresponding lithofacies distributions should be rejected, especially at late stages of the algorithm. The probability of acceptance or rejection is controlled by the Gibbs distribution (equation (2)). This probability depends largely on the temperature of the system. In this work, we start with an initial temperature of 15 and cool it every 50 iterations, using a geometric progression with mean 0.9. At early stages of the process, such a cooling scheme allows accepting inferior lithofacies distributions leading to a 10 unit increment in the objective function, with probability 0.5. In other words, initially we accept almost all plausible lithofacies distributions and reject mainly those at which PW is disconnected. The large number of acceptances of inferior distributions (or equivalently, the small number of rejections) can be observed by inspecting the slopes of the thin lines in Figure 3.

[29] At intermediate stages of the iterative process (e.g., iteration 2500), the algorithm accepts inferior lithofacies distributions leading to increments in the objective function of 0.05 only (also with probability 0.5). Thus, the search is already quite greedy and most of the inferior lithofacies distributions are rejected. Consequently, only those similar to the optimum distribution are accepted from now on. At this stage (in fact, much before), the reproduction of the reference lithofacies field is already acceptable and the undesired connections/disconnections at the initial iteration have been corrected. For instance, PW is connected with observation wells O1 to O4 and the channel containing O5 is disconnected from the channels in the central part of the domain. Consequently, the reproduction of the reference head field is already good, but not good enough. In fact, the stack contains only 10 members at iteration 2500, possibly owing to the scarcity of data (these members are potential local minima of the objective function; Figure 4). Head residuals (Figure 2g, line with crosses) have diminished significantly as compared to those at iteration 1, especially at the pumping well PW. Therefore, the value of the objective function is small (~$5 \cdot 10^{-2}$; Figure 3).

[30] The lithofacies distribution leading to the smallest objective function (theoretically, the global minimum; Figure 2c) is attained at iteration 4835, by the very end of the iterative process. It resembles the connectivity features of the reference distribution slightly better than that at iteration 2500. The differences between these two lithofacies distributions are not large. This can be explained by the fact that the algorithm already extracted the connectivity information from head data at iteration 2500. Late iterations of the algorithm are devoted to better profiling the channels depicted at intermediate stages and most lithofacies distributions are rejected (increasing slope of the solid thin line in Figure 3). However, some new channels (not informed by data) may still appear, as the one in the upper left corner of the domain. This effect is ubiquitous during the iterative process. Still, some of these MP simulations are added to the stack and are far different from the one that will lead to the global minimum. Thus, as expected, the amount of similarity and the consequent uncertainty are controlled by the amount and quality of available information. The reproduction of the reference head field is much better. Consequently, the head residuals diminish (Figure 2g, line with circles) and the minimum value of the objective function is smaller (~10^{-3}; Figure 3).

[31] The final stack is, by construction, made of members of the prior space of models that resemble the prior geological model as depicted by the training image and

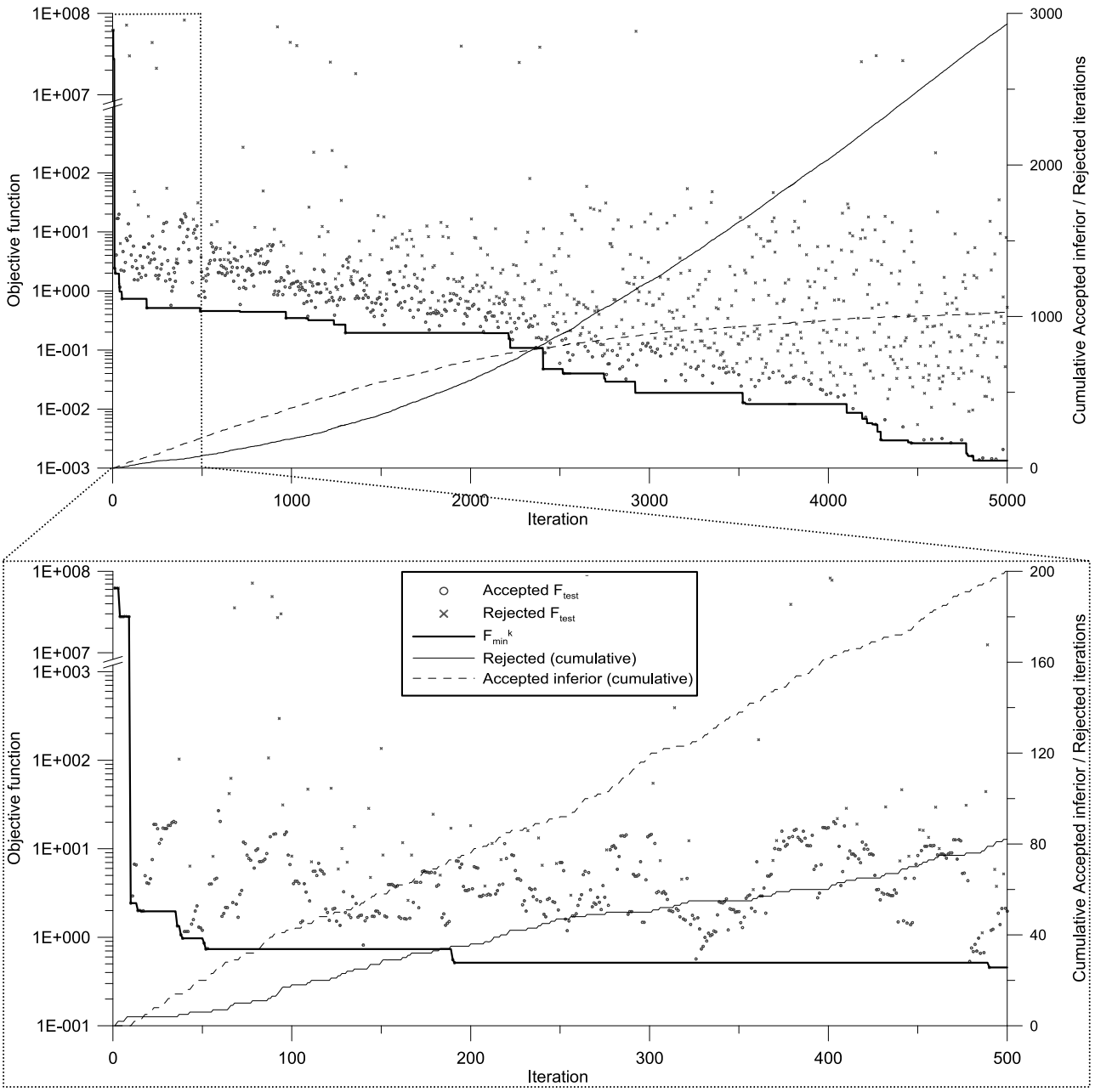

Figure 3. Evolution of the objective function in the absence of connectivity constraints. Note that the left vertical axis is broken in the range (10^3, 10^7). The bottom plot shows a selected zoom at early stages of the algorithm.

(1) honor local lithofacies data, (2) correlate with exhaustive seismic data, and (3) fit available head data well (i.e., stack members lead to objective function values smaller than 0.1). The mean and the variance of the aforementioned stack are depicted in Figures 2e and 2f, respectively. The stack mean is quite similar to the best lithofacies distribution (Figure 2c) because the annealing schedule was very greedy. Consequently, the number of rejected distributions was large and the stack size (390) small compared with the total number of iterations. The largest uncertainties, as measured by the stack variance, are found mainly at the contours of the channels and at areas not informed by measurements (e.g., the two channels at the upper left corner). We address the issue of stationarity of the iterative process by analyzing the convergence criteria in equation (4) (Figure 4). The stack starts to be populated after a large number of iterations (2405). The initial behavior of the stack is quite erratic, what is translated into large fluctuations of its statistical moments and their corresponding variations in equation (4). This is explained by the large differences between the stack members at this stage. The annealing temperature is still high, what allows an exhaustive scanning of the prior space of models (i.e., the search is completely random). Consequently, the stack members look really different at this stage and are far different from the optimum MP simulation. This occurs during ~4000 iterations, when the annealing temperature is low and new stack members look very similar to the optimum and do not display features that are not already

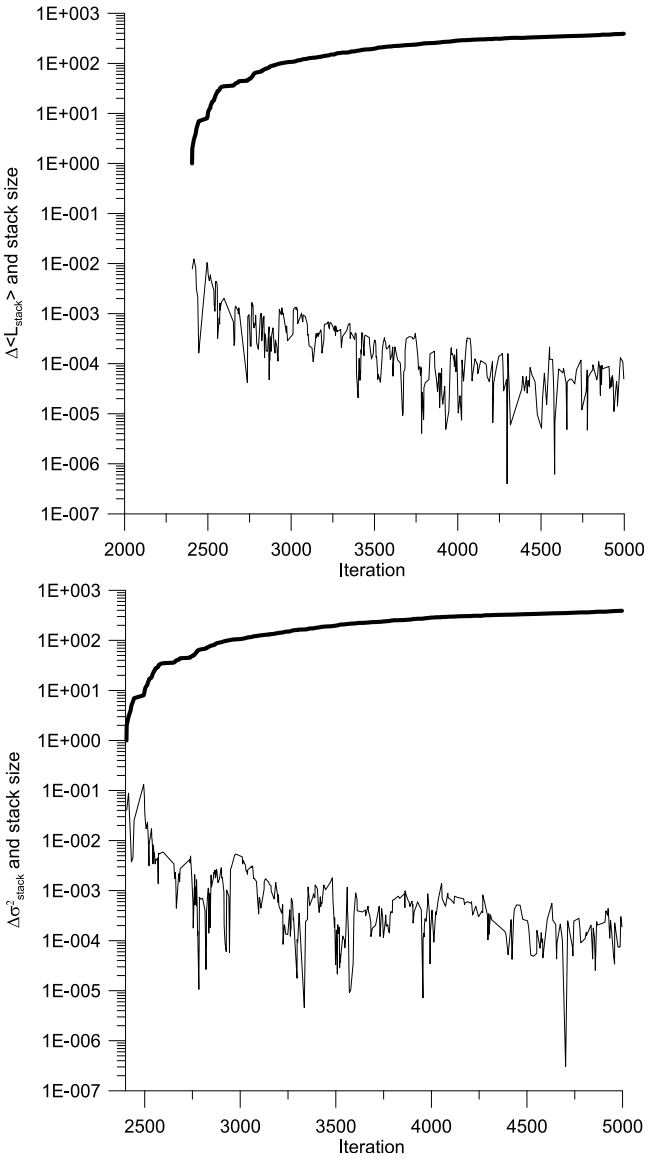

Figure 4. Stationarity of the process: evolution of (top) the average of the stack mean and (bottom) the variance of the stack mean. The stack size is depicted with a thick line in both cases.

contained in the stack. This is confirmed by the late-time pseudohorizontal slopes of $\Delta\langle L_{stack}\rangle$ and $\Delta\sigma^2_{stack}$ in Figure 4. The addition of a stack member containing new geometrical features (i.e., deviating from those already contained in the existing stack) is manifested with peaks in the variation of the statistical moments. This effect is ubiquitous throughout the iterative process (Figure 4).

[32] The role of connectivity constraints is best observed by comparing Figures 2 and 5. In this experiment, the only difference (from the algorithmical point of view) is that connectivity constraints are imposed as conditioning data for the simulation (step 1 in section 2). These restrictions do not inform about "how" the wells are connected. This information could be enriched with other sources of data, that is, travel times, but this issue is out of the scope of this paper. Still, this makes the algorithm to be far more efficient and enhances the convergence. In this case, the algorithm starts with a simulation conditioned to lithofacies and connectivity constraints. For instance, at iteration 1 (in fact, at any iteration; Figure 5a) the disconnection of PW is not allowed. Therefore, the objective function is already quite small ($\sim 2 \cdot 10^0$ versus $\sim 6 \cdot 10^7$ in the absence of connectivity constraints; Figure 6). However, the connectivity features of the reference lithofacies distribution are not yet reproduced because we still did not condition to heads. As the algorithm evolves (Figure 5b), reproductions of reference fields are enhanced and the head residuals become smaller (Figure 5g). The best lithofacies distribution (Figure 5c) is now attained at earlier stages (iteration 4421 versus 4835 in the previous case). The strong effect of conditioning to heads becomes apparent one more time. From a qualitative point of view, final reproductions of the reference fields are not very different (Figures 2c and 5c). Consequently, the head residuals are similar and very small in both cases (Figures 2g and 5g). The final stack (Figures 5e and 5f) contains 334 lithofacies distributions and is also similar to the one obtained in the absence of connectivity constraints. This reveals that the iterative process was similar and independent of the conditioning data. This is confirmed by the similarity of the cumulative sum of rejections (thin lines in Figure 6) and of the stack sizes (Figure 7). Adding connectivity constraints plays an important role mainly at early stages of the algorithm (Figure 6). In fact, if the number of iterations is large, information about connectivity can also be extracted from head data. Yet, using connectivity constraints reduces the uncertainty of the channels connecting PW with observation wells O1 to O4 (Figure 5f). The fact that the uncertainties depicted in Figure 2f (i.e., disregarding connectivity constraints) were also small confirms that head data were already informative about connectivity.

[33] From the above, we conclude that the use of connectivity constraints simply enhances the convergence rate and speeds up the process. At a given iteration, the resemblance of the reference fields is better if connectivity constraints are accounted for. Consequently, objective function values are often smaller than those in the absence of connectivity constraints. However, the global behavior of the algorithm, as measured by the minimum objective functions and the final stacks (Figures 6–8) is very similar because the number of iterations was large. We argue that the enhancement of the convergence rate is due to the fact that connectivity constraints reduce the dimension of the prior space of lithofacies distributions (connected plus disconnected) to that of the subspace of connected distributions (to which the reference solution belongs to). This issue will be further discussed later.

[34] One of the appealing features of the BMW is its simplicity. A small number of triggers control the behavior of the algorithm. Notwithstanding the parameters defining the annealing schedule (i.e., initial temperature, number of iterations per temperature step and cooling factor α), the parameter exerting a major control on the BMW is the size of the blocking moving window ("*ws*" hereinafter). Figure 7 displays the stack size obtained using 12 different window sizes. Regardless of the conditioning data, using a small window size ($ws < 24$) is not a good option. In that case, the stack is empty because only a very small portion of the domain is perturbed at a given iteration. Therefore, the probability of obtaining a significant improvement by building a new channel or creating a new connection between wells is

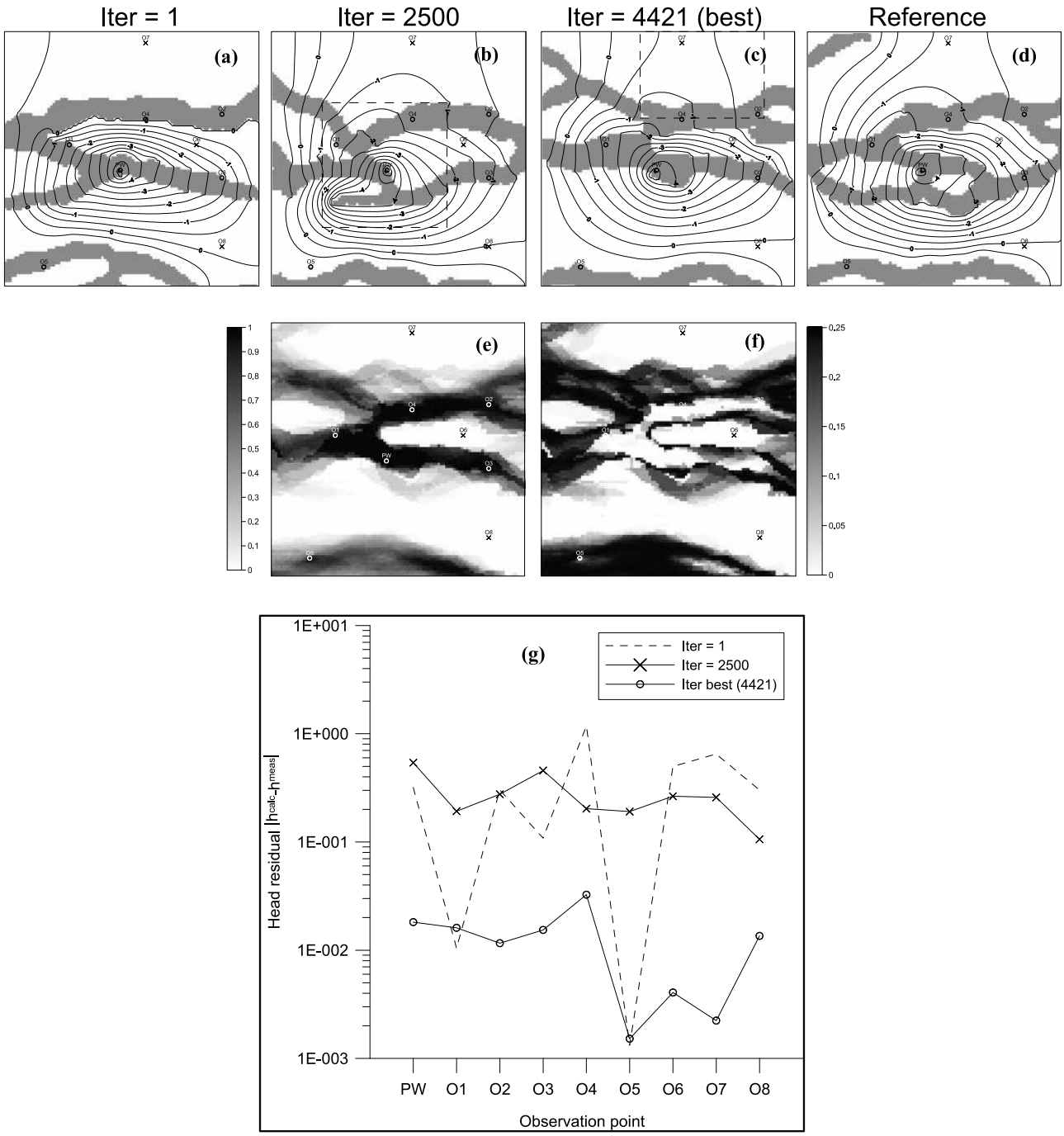

Figure 5. Algorithm evolution in the presence of connectivity constraints. (a–c) Lithofacies distributions and corresponding heads attained at early, intermediate, and final stages of the algorithm (iterations 1, 2500, and 4421, respectively; the latter leads to the minimum objective function). The blocking moving window is depicted by a dashed line. At first iteration it encompasses the whole simulation domain. (d) Reference lithofacies distribution and reference head field. (e) Mean of the final stack. This can be interpreted as the probability to find sand (p = 1) or shale (p = 0) at a given element. (f) Corresponding variance (i.e., related uncertainty). (g) Evolution of head residuals (absolute value of calculated minus measured values) at measurement locations.

very small. This result is in agreement with early results in the McMC literature that perturb one element per iteration [*Oliver et al.*, 1997]. Still, small window sizes can be used for better profiling already existing channels. On the contrary, large window sizes (*ws* > 72) perturb a significant portion of the domain, what makes the scanning too random, and therefore, inefficient. Yet, a large perturbation can be useful if there are suspicions of local minima in the objective function. Similar results were obtained by *Fu and Gomez-Hernandez* [2009] in the context of Gaussian fields. In either case, finding stack members becomes a hard task, as revealed by the small (even null) stack sizes in

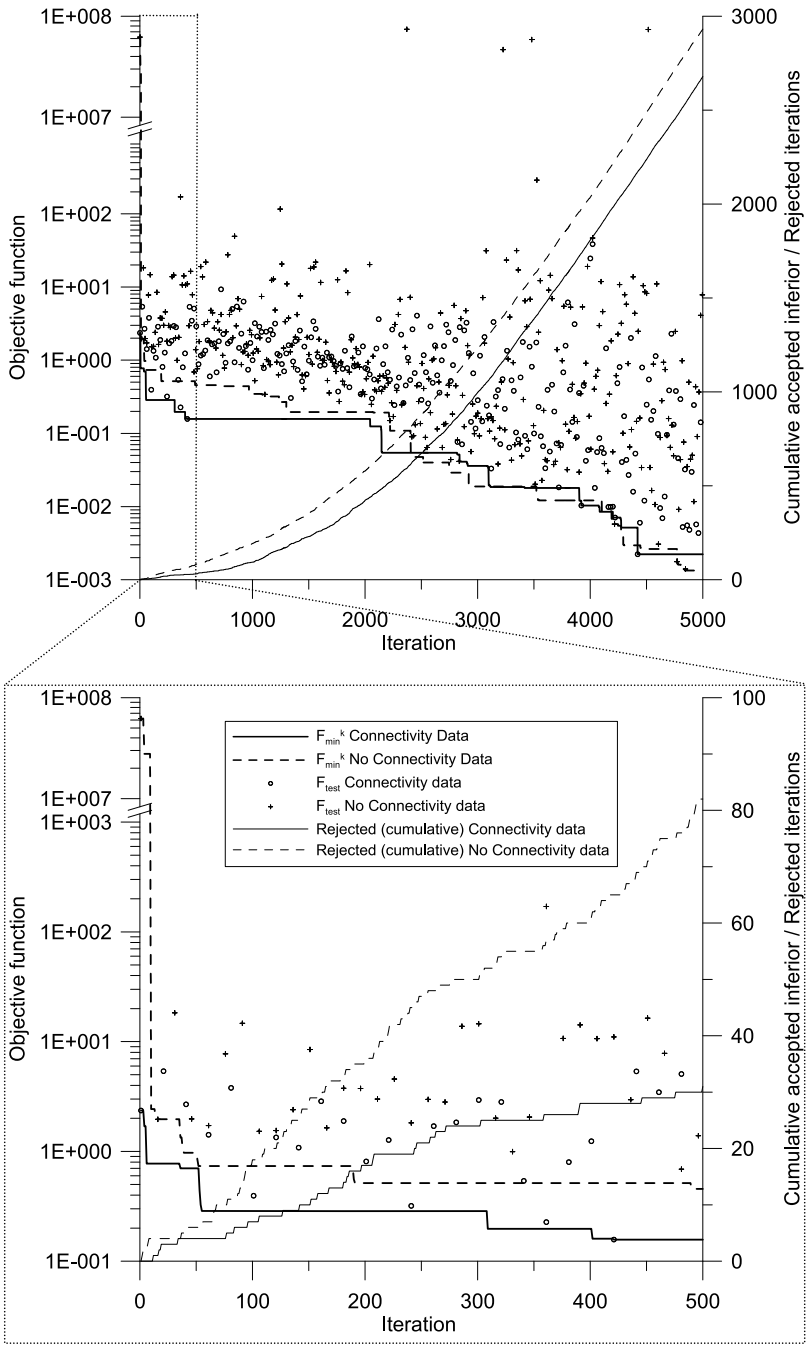

Figure 6. Evolution of the objective function in the absence/presence of connectivity data. One in every twenty values of the test objective functions are depicted. Note that the left vertical axis is broken in the range $(10^3,10^7)$.

Figure 7. This bad performance can be alleviated by increasing significantly the number of iterations. This inherently leads to CPU-intensive runs and therefore, to inefficiency. From the above, we recommend an intermediate window size (if it is constant along the iterative process). A possibility to circumvent this problem (also implemented in the BMW but not explored here) consists of defining a probability distribution of window sizes, rather than a single constant value. Large probabilities should be assigned to intermediate sizes and small probabilities (but with long tails) should be assigned to extreme sizes, either to better profile existing channels or to cause a large perturbation in the image. Yet, we fear that the choice of such a pdf (or the constant window size) is problem dependent. In this particular example, we found that appropriate window sizes are in the range of 1/4 to 1/2 of the domain size, equivalent to 24 to 48 elements. Regardless of the absence/presence of connectivity constraints, the smallest values of the minimum objective function fall within this range (Figure 7). The mean and the variance of the calculated stacks are depicted in Figure 8. The connectivity features are similar for all of them, regardless of the conditioning geological data. The main differences between the stacks are, first, the roughness of the ensemble mean and, second, the variability, as measured by the stack

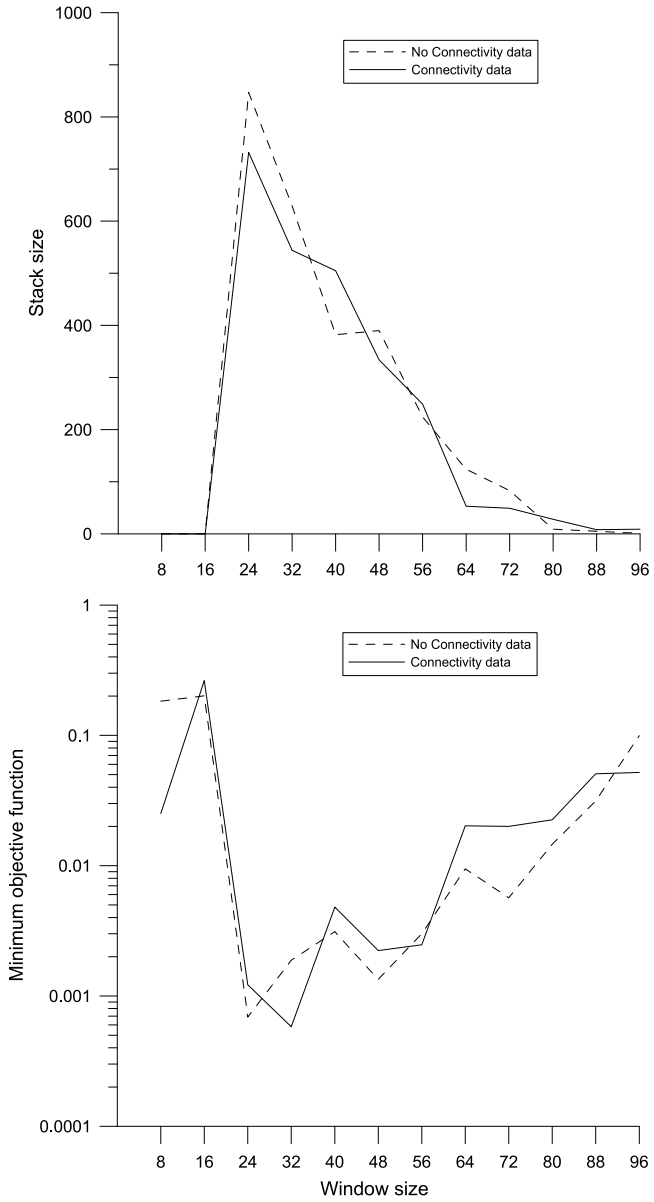

Figure 7. Stack size and minimum objective function (F_{min}^{5000}) versus blocking moving window size in the absence/presence of connectivity data.

variance. These can be partly explained by the different stack sizes (Figure 7). Populated stacks (*ws* = 24) contain a number of similar lithofacies distributions attained around the optimum one. This reduces the uncertainty and smoothes the mean. Instead, when the stack contains a few members (*ws* = 48) a channel that was attained a few times renders a large contribution to the final stack. Therefore, the stack mean is rougher and the corresponding variance larger.

[35] Figures 9 and 10 display the statistics of the stack members. Figure 9 displays the box plots of the objective functions of stack members. Owing to our definition of stack member, the maximum value is always 0.1. Regardless of the conditioning geological data, the window size and the stack population, the mean value of the objective function is always small. This confirms the goodness of the fits of available head data. The variability of the results, as measured by the size of the box plots, diminishes with the window size, but not monotonically. This is an effect of the stack size. When it is small (i.e., the window size is large) a few lithofacies distributions lead to objective function values below the threshold of 0.1. This occurs at late stages of the algorithm and a large number of iterations would be necessary to populate the stack better. Intermediate window sizes (24 to 48) yield the best results, as measured by the smallest mean values and variabilities. For most window sizes, adding connectivity constraints reduces the value of the objective function of stack members, but not substantially. This confirms that the stack is equivalent if the number of iterations is sufficiently large and that, in such case, adding connectivity constraints only enhances the convergence, but not the quality of the stack. The same conclusions can be observed in Figure 10, which depicts the box plots describing the reproduction of reference fields. The proportion of elements that match the reference lithofacies value is always above 65%. It is worth to mention that we used 9 lithofacies data only. The mean is around 75% in all cases, regardless of the conditioning data and the window size. Box plots of $RMSE_h$ are all centered on a small value and their variability is small. This confirms that the characterizations fit not only the available head data, but also yield a good overall reproduction of the reference head field.

[36] With regard to numerical performance, a run of this example using the BMW kernel in its current state (not yet parallelized) takes about 400 s of CPU time on a single Intel processor. While this might not look rather time consuming, the CPU requirements will increase nonlinearly with the complexity of the problem. This might be an issue for complex 3-D transient models. It is well known that traditional sensitivity-based optimization methods are far more efficient than Monte Carlo samplers if the objective function is smooth and continuous. Yet, this advantage is lost when the objective function displays discontinuities and an erratic behavior as in this case. Therefore, even if no exhaustive tests have been made so far, we expect that the BMW will obtain a stack of simulations conditioned to available data as efficiently as other established techniques such as the probability perturbation method. We believe that an important factor to accelerate the process is the use of as many proxies (prior rejection criteria) as possible. This will minimize the number of forward simulations, which often takes the largest part of the CPU time per iteration. This should not change the results but simply avoid the scanning of parts of the prior distribution of aquifer architectures that are far from the optimum one and, therefore, have little chances to belong to the posterior distribution.

5. Conclusions

[37] The BMW couples a multiple-point simulator for generating plausible lithofacies distributions conditioned to geological data with a groundwater flow simulator for the conditioning to heads. The final output is a stack of MP simulations of plausible aquifer architectures that (1) resemble a prior geological model depicted by a training image (i.e., belong to the prior space of models), (2) are anchored to available lithofacies data, (3) correlate with secondary information such as seismic data, (4) honor connectivity constraints (i.e., point "*i*" connected with point "*j*"), and

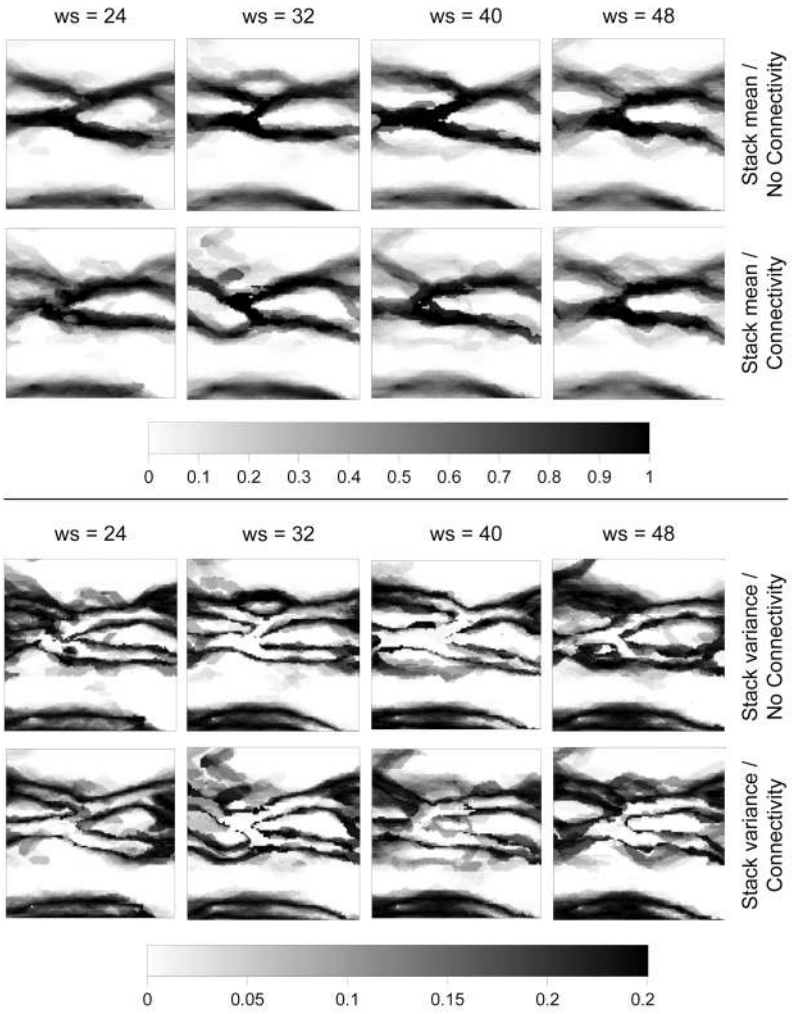

Figure 8. Stack mean and variance after 5000 iterations obtained using intermediate blocking moving window sizes "*ws*." Accepted lithofacies distributions leading to values of the objective function smaller than 0.1 are included in the stack. Corresponding stack sizes are displayed in Figure 7.

(5) fit state variable data well. In this work, we use a synthetic example to explore the role of the conditioning data and the sensitivity of the algorithm to its dominant parameter, that is, the size of the blocking moving window. The main conclusions arising from this analysis are as follows:

[38] 1. Notwithstanding other triggers of the algorithm (e.g., those controlling the cooling schedule), the size of the blocking moving window has revealed as the dominant parameter. Using a small window size can be useful for better profiling already existing channels. However, this makes the process too greedy and it is hardly possible to find stack members. Actually, 5000 iterations were not sufficient for populating the stack with a single member. On the contrary, using a large window size makes the process too random and, therefore, inefficient. Yet, it can be a good option to escape local minima of the objective function. In either case, a huge number of iterations would be required to populate the stack, what makes the computational effort unaffordable. Therefore, intermediate window sizes are recommended. Similar results are obtained by *Fu and Gomez-Hernandez* [2009] in the context of Gaussian fields. For the moment, we did not find a smart way to define the optimum size a priori. In fact, this depends likely on the size and the shape of the geological structures, the configuration of the groundwater flow field and the amount of available data. A possibility to circumvent this problem consists of using a probability distribution of window sizes, rather than a constant, user-defined value.

[39] 2. As already demonstrated by many authors [*Ronayne et al.,* 2008] in the context of MP geostatistics, state variable data such as heads contain important information about aquifer architecture and connectivity. Conditioning to state variable data enhances dramatically the initial lithofacies distribution (i.e., anchored to geological data only).

[40] 3. Adding connectivity constraints as conditioning data does not improve the quality of the results if the number of iterations is large enough. However, including them enhances the convergence of the algorithm because it reduces the dimension of the search space. This concept can be implemented in several ways, either as direct geostatis-

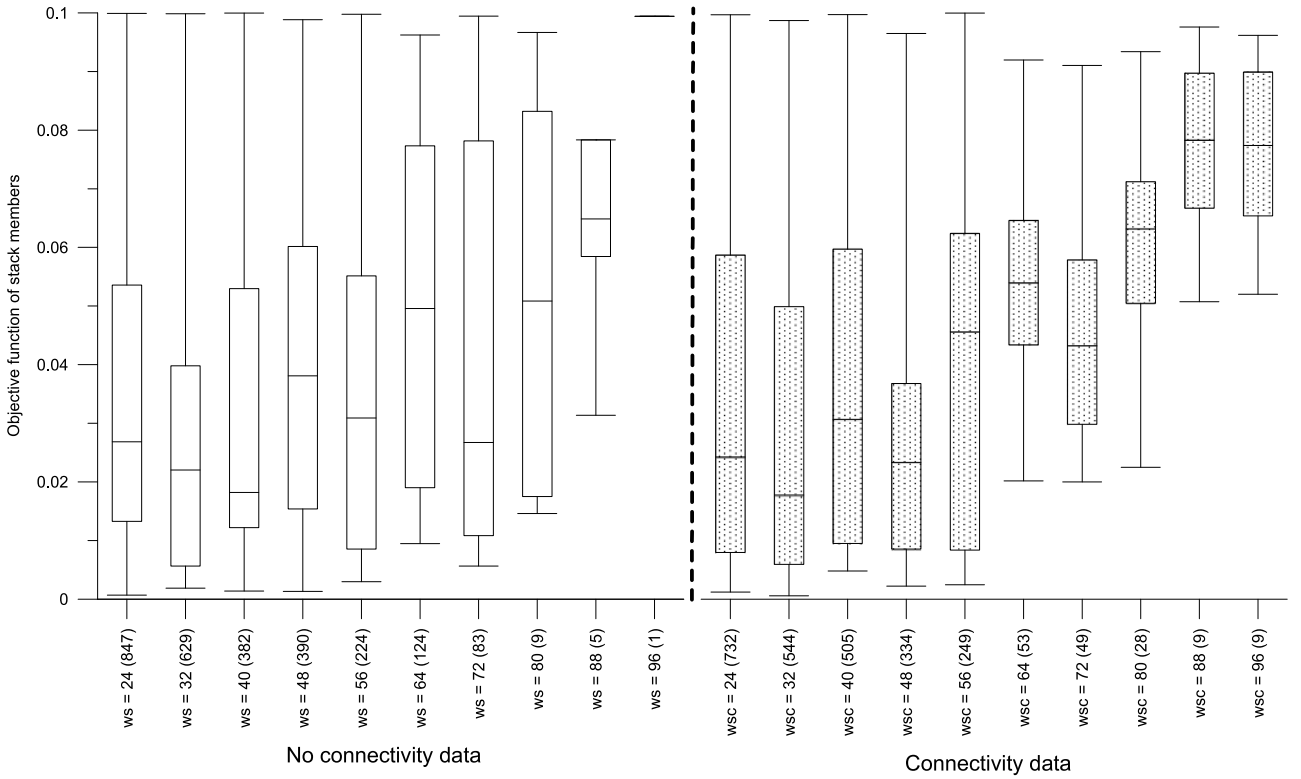

Figure 9. Box plots of the objective function corresponding to lithofacies distributions belonging to the final stack. The horizontal axis contains different blocking moving window sizes in the absence/presence of connectivity data ("*ws*" and "ws_c" at left and right, respectively). The numbers in brackets represent the stack size depicted in Figure 7. Window sizes leading to empty stacks ($ws < 24$) are not displayed.

tical conditioning (as in this work) or by means of prior rejection criteria.

[41] Much remains to be done. Here we do not consider uncertainties in the prior geological model. These may have a large impact on the results because different training images will lead to different stacks. A comparison of the statistics of those stacks could address the problem of prior uncertainty of the geological model. In view of the sensitivity analysis, a smart way for the prior definition of the range of optimum window sizes is required. Alternatively, our implementation allows the window size to vary throughout the iterative process. In that case, the blocking moving window size is randomly sampled from a user-supplied pdf (not in this work). This allows us to cause large perturbations in the aquifer architecture from time to time (i.e., when a large window size is sampled), what generates new channels and connections. At the same time, we better profile these new structures by using a small window size. Yet, the definition of the pdf of window sizes is not easy and, we fear, is problem dependent. Self-adaptive schemes may circumvent this problem.

[42] In spite of these difficulties, the strengths of the proposed method lie in its simplicity and flexibility. On the one hand, a few triggers control the behavior of the algorithm. On the other hand, the BMW can make use of any geostatistical simulation tool (2-D, 3-D, accounting for multiple-point statistics or not). It can accommodate easily any type of direct problem (maybe highly nonlinear). This allows us to condition to different types of state variables reporting on aquifer architecture and connectivity in different ways. For instance, travel times could be used to inform about the length of a given connection between observation wells. The BMW circumvents the problem of continuity and derivability of the objective function in discontinuous geological scenarios (i.e., we calculate the value of the objective function, but not its derivatives with respect to model parameters). Furthermore, the algorithm guarantees that all simulated lithofacies distributions lie in the prior space of plausible simulations defined by the prior geological model. This may not be satisfied by other traditional sensitivity-based inversion techniques that modify the parameter fields iteratively. This may be accomplished by other simulated annealing based methods, but the main advantage of the BMW is that it yields directly an ensemble of samples of the posterior distribution of parameters (lithofacies distribution in this case) instead of a single optimal solution. Yet, the sampling is not exhaustive, as opposed to the classical McMC approach. In fact, the BMW becomes a McMC exhaustive sampler by removing the cooling schedule.

[43] Overall, the main drawback of our approach is that it is computationally demanding. A number of prior rejection criteria can be evaluated before simulating the state variables (actually, the most CPU-intensive part). These should be based on a fast evaluation of the lithofacies distribution being tested. Examples are local and directional connectivity of the training image or the lithofacies proportions. We foresee that future improvements will focus on the parallelization of the algorithm and on the search of suitable prior rejection criteria. The results presented here should be

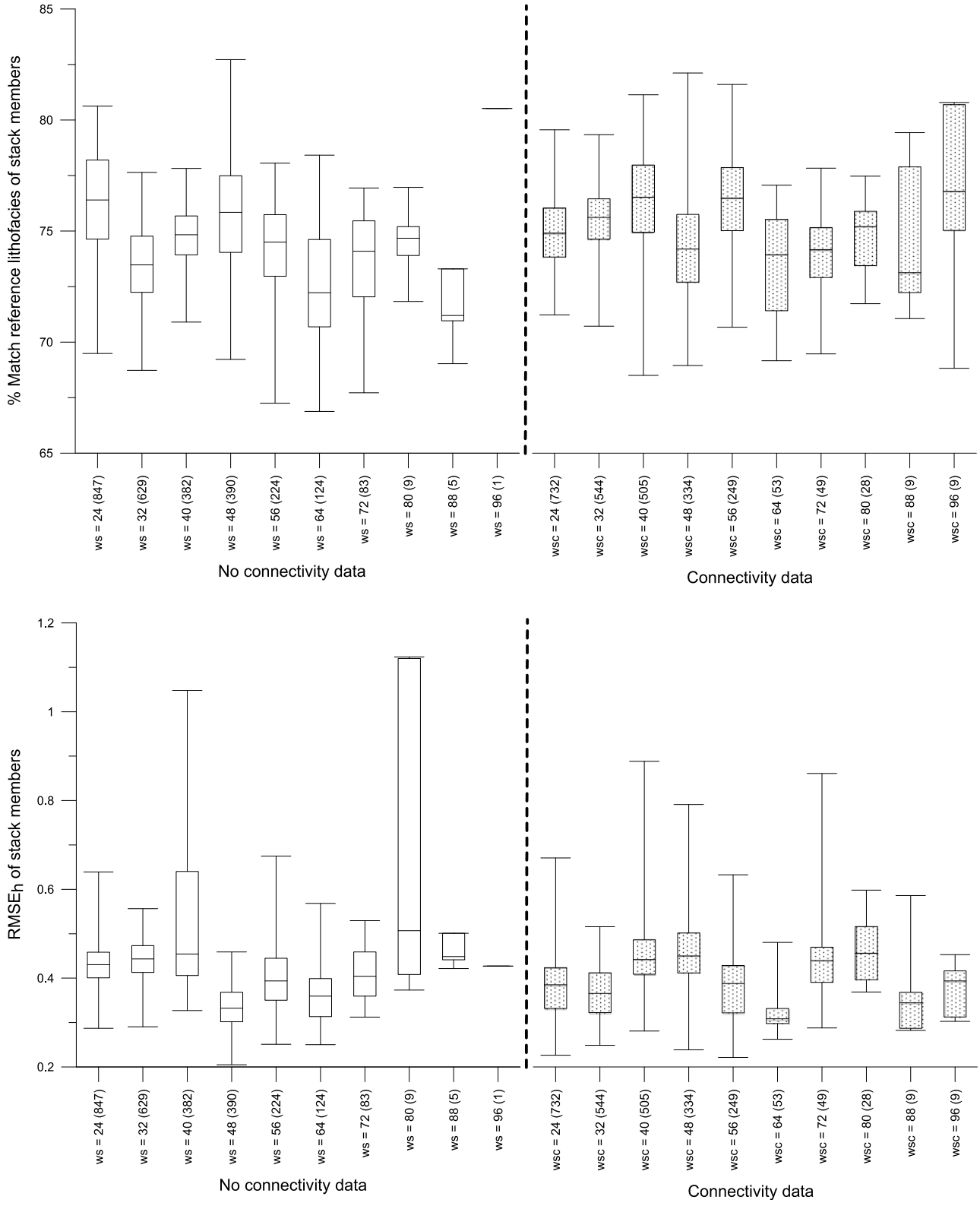

Figure 10. Box plots depicting (top) the proportion of elements matching the reference lithofacies of the stack members matching the reference lithofacies and (bottom) the root mean square error of heads $RMSE_h$ of the stack members. The horizontal axis contains different blocking moving window sizes in the absence/presence of connectivity data ("*ws*" and "ws_c" at left and right, respectively). The numbers in brackets represent the stack size depicted in Figure 7. Window sizes leading to empty stacks ($ws < 24$) are not displayed.

viewed as a hopeful step in the direction of linking multiple-point geostatistics and hydrogeology.

[44] **Acknowledgments.** Funding for this work was provided by the Swiss National Science Foundation through contract PP002-106557. We want to thank Jesus Carrera (Institut Jaume Almera, Consejo Superior de Investigaciones Cientificas, Spain) and Jaime Gomez-Hernandez (Technical University of Valencia, Spain) for their technical assessment. We also want to thank three anonymous reviewers for their helpful contributions.

References

Alcolea, A., J. Carrera, and A. Medina (2000), A hybrid Marquardt-Simulated Annealing method for solving the groundwater inverse problem, paper presented at ModelCare'99. Calibration and Reliability in Groundwater Modelling: Coping with Uncertainty, Int. Assoc. of Hydrol. Sci., Zurich, Switzerland.

Alcolea, A., J. Carrera, and A. Medina (2006a), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Adv. Water Resour.*, *29*, 1678–1689, doi:10.1016/j.advwatres.2005.12.009.

Alcolea, A., J. Carrera, and A. Medina (2006b), Inversion of heterogeneous parabolic-type equations using the pilot points method, *Int. J. Numer. Methods Fluids*, *51*(9–10), 963–980, doi:10.1002/fld.1213.

Alcolea, A., J. Carrera, and A. Medina (2008), Regularized pilot points method for reproducing the effect of small scale variability: Application to simulations of contaminant transport, *J. Hydrol.*, *355*(1–4), 76–90, doi:10.1016/j.jhydrol.2008.03.004.

Allard, D. (1994), Simulating a geological lithofacies with respect to connectivity information using the truncated Gaussian model, in *Geostatistical Simulations: Proceedings of the Geostatistical Simulation Workshop*, edited by M. Armstrong and P. A. Dowd, Kluwer Acad., Norwell, Mass., pp. 197–211.

Balakrishnan, S., A. Roy, M. G. Ierapetritou, G. P. Flach, and P. G. Georgopoulos (2003), Uncertainty reduction and characterization for complex environmental fate and transport models: An empirical Bayesian framework incorporating the stochastic response surface method, *Water Resour. Res.*, *39*(12), 1350, doi:10.1029/2002WR001810.

Caers, J. (2003), Geostatistical history matching under training-image based geological model constraints, *SPE J.*, *8*(3), 218–226, doi:10.2118/74716-PA.

Caers, J. (2007), Comparing the gradual deformation with the probability perturbation method for solving inverse problems, *Math. Geol.*, *39*(1), 27–52, doi:10.1007/s11004-006-9064-6.

Caers, J., and B. T. Hoffman (2006), The probability perturbation method: A new look at Bayesian inverse modeling, *Math. Geol.*, *38*(1), 81–100, doi:10.1007/s11004-005-9005-9.

Caers, J., and T. Zhang (2004), Multiple-point statistics: A quantitative vehicle for integration of geologic analogs into multiple reservoir models. Integration of outcrop and modern analog data in reservoir models, *AAPG Mem.*, *80*, 383–394.

Caers, J., S. Strebelle, and K. Payrazyan (2003), Stochastic integration of seismic data and geologic scenarios: A West Africa submarine channel saga, *Leading Edge*, *22*(3), 192–196, doi:10.1190/1.1564521.

Capilla, J. E., J. Rodrigo, and J. J. Gomez-Hernandez (1999), Simulation of non-Gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information, *Math. Geol.*, *31*(7), 907–927, doi:10.1023/A:1007580902175.

Carle, S. F., and G. E. Fogg (1997), Modeling spatial variability with one multidimensional continuous-lag Markov chain, *Math. Geol.*, *29*(7), 891–918, doi:10.1023/A:1022303706942.

Carrera, J. (1987), State of the art of the inverse problem applied to the flow and solute transport problems, in *Groundwater Flow and Quality Modeling*, edited by N. Asi, pp. 549–585, NATO, Brussels.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeol. J.*, *13*, 206–222, doi:10.1007/s10040-004-0404-7.

Chugunova, T. L., and L. Y. Hu (2008), Multiple-point simulations constrained by continuous auxiliary data, *Math. Geosci.*, *40*(2), 133–146, doi:10.1007/s11004-007-9142-4.

de Marsily, G., J. P. Delhomme, F. Delay, and A. Buoro (1999), 40 years of inverse problems in hydrogeology, *C. R. Acad. Sci., Ser. IIa Sci. Terre Planetes*, *329*(2), 73–87.

de Marsily, G., F. Delay, J. Gonçalvès, P. Renard, V. Teles, and S. Violette (2005), Dealing with spatial heterogeneity, *Hydrogeol. J.*, *13*(1), 161–183, doi:10.1007/s10040-004-0432-3.

de Vries, L. M., J. Carrera, O. Falivene, O. Gratacos, and L. J. Slooten (2009), Application of multiple point geostatistics to non-stationary images, *Math. Geol.*, *41* (1), 29–42.

Dubuisson, M. P., and K. J. Jain (1994), A modified Hausdorff distance for object matching, in *Pattern Recognition*, Conference A: Computer Vision & Image Processing, Proceedings of the 12th IAPR International Conference, vol. 1, pp. 566–568.

Falivene, O., P. Arbues, A. Gardiner, G. Pickup, J. A. Munoz, and L. Cabrera (2006), Best practice stochastic facies modeling from a channel-fill turbidite sandstone analog (the Quarry outcrop, Eocene Ainsa basin, northeast Spain), *AAPG Bull.*, *90*(7), 1003–1029, doi:10.1306/02070605112.

Feyen, L., and J. Caers (2006), Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations, *Adv. Water Resour.*, *29*, 912–929, doi:10.1016/j.advwatres.2005.08.002.

Fu, J., and J. Gomez-Hernandez (2008), Preserving spatial structure for inverse stochastic simulation using blocking Markov chain Monte Carlo method, *Inverse Probl. Sci. Eng.*, *16*(7), 865–884, doi:10.1080/17415970802015781.

Fu, J., and J. Gomez-Hernandez (2009), A blocking Markov chain Monte Carlo method for inverse stochastic hydrogeological modeling, *Math. Geosci.*, *41*(2), 105–128, doi:10.1007/s11004-008-9206-0.

Gomez-Hernandez, J. J., and R. M. Srivastava (1990), ISIM3D: An ANSI-C three-dimensional multiple indicator conditional simulation model, *Comput. Geosci.*, *16*(4), 395–440, doi:10.1016/0098-3004(90)90010-Q.

Gomez-Hernandez, J. J., and X. H. Wen (1998), To be or not to be multi-Gaussian?: A reflection on stochastic hydrogeology, *Adv. Water Resour.*, *21*, 47–61, doi:10.1016/S0309-1708(96)00031-0.

Gomez-Hernandez, J. J., A. Sahuquillo, and J. E. Capilla (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 1. Theory, *J. Hydrol.*, *204*(1–4), 162–174.

Guardiano, F., and R. M. Srivastava (1993), Multivariate geostatistics: Beyond bivariate moments, in *Geostatistics Troia 1992*, edited by A. Soares, pp. 133–144, Kluwer Acad., Dordrecht, Netherlands.

Haldorsen, H. H., and L. W. Lake (1984), A new approach to shale management in field scale simulation models, *SPEJ Soc. Pet. Eng. J.*, *24*(4), 447–457, doi:10.2118/10976-PA.

Hassan, A. E., H. M. Bekhit, and J. B. Chapman (2009), Using Markov Chain Monte Carlo to quantify parameter uncertainty and its effect on predictions of a groundwater flow model, *Environ. Modell. Software*, *24*(6), 749–763, doi:10.1016/j.envsoft.2008.11.002.

Hendricks Franssen, H.-J., A. Alcolea, M. Riva, M. Bakr, N. van der Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse modelling of groundwater flow and the characterisation of well catchments, *Adv. Water Resour.*, *32*, 851–872, doi:10.1016/j.advwatres.2009.02.011.

Holloman, C. H., H. K. H. Lee, and D. M. Higdon (2006), Multiresolution genetic algorithms and Markov chain Monte Carlo, *J. Comput. Graphical Stat.*, *15*(4), 861–879, doi:10.1198/106186006X157423.

Hu, L. Y. (2008), Extended probability perturbation method for calibrating stochastic reservoir models, *Math. Geosci.*, *40*(8), 875–885, doi:10.1007/s11004-008-9158-4.

Hu, L. Y., and T. Chugunova (2008), Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review, *Water Resour. Res.*, *44*, W11413, doi:10.1029/2008WR006993.

Hu, L. Y., M. Le Ravalec, and G. Blanc (2001), Gradual deformation and iterative calibration of truncated Gaussian simulations, *Petrol. Geosci.*, *7*, S25–S30.

Jenni, S., L. Y. Hu, R. Basquet, G. De Marsily, and B. Bourbiaux (2007), History matching of a stochastic model of field-scale fractures: Methodology and case study, *Oil Gas Sci. Technol. 62*(2), 265–276, doi:10.2516/ogst:2007022.

Journel, A. G. (2002), Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses, *Math. Geol.*, *34*(5), 573–596, doi:10.1023/A:1016047012594.

Journel, A., and C. V. Deutsch (1993), Entropy and spatial disorder, *Math. Geol.*, *25*(3), 329–355, doi:10.1007/BF00901422.

Kerrou, J., P. Renard, H.-J. Hendricks Franssen, and I. Lunati (2008), Issues in characterizing heterogeneity and connectivity in non-multiGaussian media, *Adv. Water Resour.*, *31*, 147–159, doi:10.1016/j.advwatres.2007.07.002.

Kirkpatrick, S., C. D. Gelatt Jr., and M. P. Vecchi (1983), Optimization by simulated annealing, *Science*, *220*, 671–680, doi:10.1126/science.220.4598.671.

Knudby, C., and J. Carrera (2005), On the relationship between indicators of geostatistical, flow and transport connectivity, *Adv. Water Resour.*, *28*, 405–421, doi:10.1016/j.advwatres.2004.09.001.

Kool, J. B., J. C. Parker, and M. T. Van Genuchten (1987), Parameter estimation for unsaturated flow and transport models: A review, *J. Hydrol.*, *91*, 255–293, doi:10.1016/0022-1694(87)90207-1.

Le Loc'h, G., and A. Galli (1997), Truncated plurigaussian method: Theoretical and practical points of view, in *Geostatistics Wollongong 1996*, vol. 1, edited by E. Y. Baafi and N. A. Schofield, pp. 211–222, Kluwer Acad., Dordrecht, Netherlands.

Liu, N., and D. S. Oliver (2003), Evaluation of Monte Carlo methods for assessing uncertainty, *SPE J.*, *8*(2), 188–195, doi:10.2118/84936-PA.

Liu, Y., A. Harding, W. Abriel, and S. Strebelle (2004), Multiple-point simulation integrating wells, three-dimensional seismic data and geology, *AAPG Bull.*, *88*, 905–921, doi:10.1306/02170403078.

McLaughlin, D., and L. R. Townley (1996), A reassessment of the groundwater inverse problem, *Water Resour. Res.*, *32*, 1131–1161, doi:10.1029/96WR00160.

Medina, A., and J. Carrera (2003), Geostatistical inversion of coupled problems: Dealing with computational burden and different types of data, *J. Hydrol.*, *281*, 251–264, doi:10.1016/S0022-1694(03)00190-2.

Meier, P. M., A. Medina, and J. Carrera (2001), Geostatistical inversion of cross-hole pumping tests for identifying preferential flow channels within a shear zone, *Ground Water*, *39*(1), 10–17, doi:10.1111/j.1745-6584.2001.tb00346.x.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines, *J. Chem. Phys.*, *21*(6), 1087–1092, doi:10.1063/1.1699114.

Oliver, D. S., L. B. Cunha, and A. C. Reynolds (1997), Markov chain Monte Carlo methods for conditioning a log-permeability field to pressure data, *Math. Geol.*, *29*(1), 61–91, doi:10.1007/BF02769620.

Renard, P., and J. Caers (2008), Conditioning facies simulations with connectivity data, in *Proceedings of the Eighth International Geostatistics Congress*, vol. 2, edited by J. M. Ortiz and X. Emery, Gecamin Publ., Santiago, Chile, pp. 597–606.

Robert, C. P., and G. Gasella (1999), *Monte Carlo Statistical Methods*, 507 pp., Springer, New York.

Ronayne, M. J., S. M. Gorelick, and J. Caers (2008), Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach, *Water Resour. Res.*, *44*, W08426, doi:10.1029/2007WR006635.

Strebelle, S. (2002), Conditional simulation of complex geological structures using multiple-point statistics, *Math. Geol.*, *34*(1), 1–21, doi:10.1023/A:1014009426274.

Strebelle, S. (2006), Sequential simulation for modeling geological structures from training images, in *Stochastic Modelling and Geostatistics: Principles, Methods and Case Studies*, edited by J. M. Yarus and R. L. Chambers, *Am. Assoc. Petrol. Geol.*, pp. 139–149.

Suzuki, S., and J. Caers (2006), History matching with an uncertain geological scenario, paper 102154-MS, SPE Annual Technical Conference and Exhibition, 24–27 Sept. 2006, San Antonio, Tex., doi:10.2118/102154-MS.

Suzuki, S., G. Caumon, and J. Caers (2008), Dynamic data integration for structural modeling: Model screening approach using a distance-based model parameterization, *Comput. Geosci.*, *12*(1), 105–119, doi:10.1007/s10596-007-9063-9.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Parameter Estimation*, 342 pp., Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Trinchero, P., X. Sanchez-Vila, and D. Fernandez-Garcia (2008), Point-to-point connectivity: An abstract concept or a key issue for risk assessment studies?, *Adv. Water Resour.*, *31*, 1742–1753, doi:10.1016/j.advwatres.2008.09.001.

Van Laarhoven, P. J. M., and E. H. L. Aarts (1987), *Simulated Annealing: Theory and Applications*, Kluwer Acad., Dordrecht, Netherlands.

Wen, X. H., T. T. Tran, R. A. Behrens, and J. J. Gomez-Hernandez (2000), Production data integration in sand/shale reservoirs using sequential self-calibration and GeoMorphing: A comparison, paper presented at Annual Technical Conference and Exhibition, Soc. of Pet. Eng., Dallas, Tex., 1–4 Oct.

Xu, W. (1996), Conditional curvilinear stochastic simulation using pixel-based algorithms, *Math. Geol.*, *28*(7), 937–949, doi:10.1007/BF02066010.

Yeh, W. W. G. (1986), Review of parameter identification procedures in groundwater hydrology, *Water Resour. Res.*, *22*, 95–108, doi:10.1029/WR022i002p00095.

Zanini, A., and P. K. Kitanidis (2009), Geostatistical inversing for large-contrast transmissivity fields, *Stochastic Environ. Res. Risk Assess.*, *23*(5), 565–577, doi:10.1007/s00477-008-0241-7.

Zinn, B., and C. F. Harvey (2003), When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields, *Water Resour. Res.*, *39*(3), 1051, doi:10.1029/2001WR001146.

A. Alcolea, TK Consult AG, Seefeldstrasse 287, CH-8008 Zürich, Switzerland. (andres.alcolea@tkconsult.ch)

P. Renard, Centre for Hydrogeology and Geothermics, University of Neuchâtel, Rue Emile Argand, 11, CP-158, CH-2009 Neuchâtel, Switzerland. (philippe.renard@unine.ch)