

Document downloaded from:

<http://hdl.handle.net/10251/73815>

This paper must be cited as:

TENG; Gómez-Hernández, JJ. (2015). Inverse sequential simulation: Performance and implementation details. *Advances in Water Resources*. 86B:311-326.
doi:10.1016/j.advwatres.2015.04.015.



The final publication is available at

<http://dx.doi.org/10.1016/j.advwatres.2015.04.015>

Copyright Elsevier

Additional Information

Inverse Sequential Simulation: Performance and Implementation Details

Teng Xu^{a,*}, J. Jaime Gómez-Hernández^a

^a*Research Institute of Water and Environmental Engineering, Universitat Politècnica de València, 46022, Valencia, Spain*

Abstract

For good groundwater flow and solute transport numerical modeling, it is important to characterize the formation properties. In this paper, we analyze the performance and important implementation details of a new approach for stochastic inverse modeling called inverse sequential simulation (iSS). This approach is capable of characterizing conductivity fields with heterogeneity patterns difficult to capture by standard multiGaussian-based inverse approaches. The method is based on the multivariate sequential simulation principle, but the covariances and cross-covariances used to compute the local conditional probability distributions are computed by simple co-kriging are derived from an ensemble of conductivity and piezometric head fields, in a similar manner as the experimental covariances are computed in ensemble Kalman filtering. A sensitivity analysis is performed on a synthetic aquifer regarding the number of members of the ensemble of realizations, the number of conditioning data, the number of piezometers at which piezometric heads are observed, and the number of nodes retained within the search neighborhood at the moment of computing the local conditional probabilities. The results show the importance of having a sufficiently large number of all of the mentioned parameters for the algorithm to characterize properly hydraulic conductivity fields with clear non-multiGaussian features.

Keywords: Inverse modeling; normal-score transform; non-Gaussianity; simple co-kriging; data assimilation; non-stationary covariance

1. Introduction

In groundwater flow and mass transport the characterization of the formation properties is important if we wish predict the state of the aquifer, i.e., the spatiotemporal distributions of piezometric heads and solute concentrations. This characterization is generally made on the basis of a few direct (hard) measurements of

*Corresponding author. Tel: +34 963879615 Fax: +34 963879492

Email addresses: tenxu@posgrado.upv.es (Teng Xu), jaime@dihma.upv.es (J. Jaime Gómez-Hernández)

5 the parameters that control the aquifer state, such as hydraulic conductivities and porosities, some indirect
6 (soft) measurements, such as those derived from geophysical surveys, and a few observation of the state of
7 the aquifer, such as piezometric heads.

8 Incorporating the state observations to characterize the parameters of the system is the object of inverse
9 modeling, and it is a powerful but difficult task that has been the subject of extensive research during the
10 last decades. See the work by Zhou et al. [1] for a recent review on inverse methods in hydrogeology.

11 Many inverse modeling approaches have been developed. Just to name a few, there are the gradual
12 deformation method, the sequential self-calibration, variants of the Markov chain Monte Carlo method, the
13 representer method, the pilot points method, ... [e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13].

14 Besides the above mentioned methods, the ensemble Kalman filter (EnKF) [14] is the method that has
15 attracted most attention recently. Although not an inverse method by conception, the inclusion of the
16 parameters governing the state equation as part of an extended state vector, has turned the EnKF in the
17 most favored inverse method nowadays [15, 16]. The EnKF is well known for its flexibility to be applied to
18 virtually any inverse problem, for its simple implementation and usage, and for its efficiency in producing
19 realizations of the geological parameters that are consistent with the observed state variable data.

20 None of the above mentioned methods has been able to address the problem of characterizing conductivity
21 fields with clear non-multiGaussian features, including the EnKF [e.g., 17, 18]. The main reason why the
22 EnKF fails for non-multiGaussian fields is that it is optimal only for multiGaussian variates and linear state-
23 transfer functions [19]. For this reason, nowadays, the main focus of inverse modeling, at least in the fields
24 of hydrogeology and petroleum engineering, is on how to reproduce non-multiGaussian patterns.

25 Outside of the inverse modeling realm, the development of the single normal-equation simulation [20, 21]
26 has solved the problem of incorporating hard and soft data for the characterization of spatial patterns using
27 statistics higher than order two, and thus, the inclusion of the spatial features that cannot be characterized
28 simply by a covariance function. The algorithms that are capable to account for statistics higher-than-
29 order-two are broadly termed as multiple-point statistics (MPS). They rely on the existence of a training
30 image exhibiting the types of patterns to be present in the final models, from which to infer the higher-order
31 statistics. Some available algorithms and codes are SNESIM [22], FILTERSIM [23], SIMPAT [24], DS [25],
32 EnPAT [26], and others [e.g., 27, 28, 29, 30, 31]. But these algorithms were not devised for inverse modeling;
33 therefore, they cannot incorporate state variable information.

34 There have been some attempts to combine MPS and inverse modeling: Hu et al. [32] used, with mod-
35 erate success, the realizations of uncorrelated random numbers needed for the drawing of the conductivity

36 value from the local distribution function on the sequential simulation implementation of the single normal
37 equations as the state variables to be updated during the analysis step if the EnKF; Zhou et al. [27], Li et al.
38 [26] developed a new MPS algorithm (termed EnPAT) which blends direct simulation [25] and the EnKF to
39 generate inverse conditional realizations of conductivity in channelized bimodal aquifers; EnPAT works well,
40 but it is still very CPU-time consuming.

41 Some authors are against the use of MPS arguing that MPS is too dependent on the choice of the
42 parameters controlling the algorithm [33]. Other approaches to address the issue of non-multiGaussianity
43 in inverse modeling include the works by Sun et al. [17], who combined the EnKF with a Gaussian mixture
44 model, or by Liu and Oliver [34], Gu and Oliver [35], Franssen and Kinzelbach [36] who use an iterative
45 EnKF; plus a set of works who combine the normal-score transform (sometimes referred as anamorphosis)
46 and the EnKF [e.g., 18, 37, 38, 39, 40, 41, 42, 43, 44]. None of these methods can be considered as the
47 definite solution of inverse modeling for conductivity fields that display non-multiGaussian features.

48 In this paper we describe a new method for inverse stochastic modeling applicable for non-multiGaussian
49 fields. We have called this method inverse sequential simulation, and it is inspired on the standard multivari-
50 ate sequential simulation algorithm [45, 46] with normal-score transforms [47] and the Monte Carlo concept
51 of the EnKF. The paper describes the algorithm and its implementation, and then performs a sensitivity
52 analysis of the key parameters controlling the algorithm; the paper ends with a post-audit of the generated
53 ensemble of realizations to check how they would perform in a solute transport prediction exercise. The
54 algorithm has been benchmarked against the normal-score ensemble Kalman filter, with excellent results, in
55 the paper by Xu and Gómez-Hernández [48]; therefore, this paper will not focus on a comparison with other
56 methods, but on the implementation and performance of the algorithm.

57 **2. Methodology**

58 The key idea of inverse sequential simulation (iSS) is to use multivariate multi-Gaussian sequential
59 simulation [45] to generate realizations of normal scores of conductivity, conditioned on conductivity and
60 observed head data. The main difference with standard sequential simulation is that the method does not
61 use an analytical, stationary model for the auto- and cross-covariances, but rather, as in the EnKF, non-
62 stationary auto- and cross-covariances are derived from an ensemble of conductivity realizations and their
63 associated piezometric heads (obtained by solving a groundwater flow model).

64 Before describing the whole algorithm, recall the main steps in any sequential simulation algorithm:

- 65 1. Define a random path to visit all nodes of the grid on which the realization will be generated.

66 2. Visit the random path sequentially.

- 67 (a) At each node, collect the conditioning data for all variables (in our case, we will have two variables:
68 conductivity and piezometric heads) within a user-defined search neighborhood centered at the
69 point to simulate (the size and orientation of the search neighborhood, and the number of data
70 of each variable to keep within it are parameters that must be specified by the user).
- 71 (b) Compute the local conditional distribution function. If we adopt a multivariate multiGaussian
72 random function model, the local conditional distribution is a Gaussian distribution with mean
73 and variance given by the simple co-kriging estimate and the simple co-kriging variance.
- 74 (c) Draw, randomly, a value from the local conditional distribution function.
- 75 (d) Include the simulated value in the set of conditioning data for the simulation of the next nodes
76 and move to the next node.

77 Our proposal is to use this algorithm to generate conductivity fields conditioned to piezometric heads.
78 For this purpose we need the auto-covariances of both conductivity and head, and their cross-covariance.
79 These covariances, particularly the ones involving the piezometric heads, but also the conductivity auto-
80 covariance when there are conditioning conductivity data, are clearly non-stationary. Some authors have
81 developed analytical expressions relating these covariances by approximating the solution of the groundwater
82 flow equation [e.g., 49]. We propose to use experimentally-derived covariances obtained from an ensemble of
83 realizations, much like it is done in the EnKF.

84 At any time t , we could derive all necessary covariances experimentally as follows:

- 85 1. Generate an ensemble of N_e realizations of conductivity. Each realization contains N nodes. $K_i(j)$
86 refers to the conductivity value at realization i and node j .
- 87 2. Given initial and boundary conditions, sources and sinks, solve the groundwater flow equation Eq.(1)[50]
88 for each realization until time t and obtain an ensemble of piezometric heads.

$$S_s \frac{\partial H}{\partial t} - \nabla \cdot (K \nabla H) = W, \quad (1)$$

89 where S_s is specific storage coefficient [L^{-1}], H is the hydraulic head [L], K is the hydraulic conductivity
90 [LT^{-1}], W denotes sources and sinks per unit volume [T^{-1}]; t is the time [T], $\nabla \cdot$ is the divergence
91 operator, and ∇ is the gradient operator.

- 92 3. The cross-covariance between conductivity K at location j and piezometric head H at location l is

93 given by

$$C_{K,H}(j,l) = \frac{1}{N_e} \sum_{i=1}^{N_e} (K_i(j) - \overline{K}(j))(H_i(l) - \overline{H}(l)), \quad (2)$$

94 where the overbar indicates ensemble average, i.e.,

$$\overline{K}(j) = \frac{1}{N_e} \sum_{i=1}^{N_e} K_i(j). \quad (3)$$

95 The auto-covariances for heads and conductivities are computed similarly.

96 In addition, since we are planning to work with multiGaussian sequential simulation, it is more convenient
97 to work with the normal-score transform [47] of the variable of interest, in our case conductivity. Therefore,
98 the sequential simulation algorithm is performed on a new variable \tilde{K} which is obtained by the normal-score
99 transform of K according to the following expression:

$$\tilde{K}_i(j) = G^{-1}(F_j(K_i(j))) \quad (4)$$

100 where $F_j(K_i(j))$ is the local cumulative distribution at node j computed (numerically) from the N_e conduc-
101 tivity values of all realizations at node j , and $G(\cdot)$ is the standard Gaussian cumulative distribution function.
102 Auto-covariances and cross-covariances are computed, as described above, for the normal-score transformed
103 conductivities, not for the untransformed ones as in the description. These covariances will be different from
104 the ones corresponding to the untransformed conductivity.

105 Once the normal-scores are generated, they are transformed back to conductivity space for the solution
106 of the groundwater flow equation using the inverse expression:

$$K_i(j) = F_j^{-1}(G(\tilde{K}_i(j))). \quad (5)$$

107 Using the cross-covariances to generate normal-score transformed conductivities conditioned on piezomet-
108 ric heads is equivalent to imposing the linear relationship between normal-score transformed conductivities
109 and heads as given by the linear approximation of a Taylor expansion of the equation relating both variables
110 [51]. (This equation will be a convolution of the groundwater flow equation and the direct and inverse
111 normal-score transform functions, see [43].) The generated fields will honor the linear part of the Taylor
112 expansion, but if the equation is highly non-linear, the resulting conductivity fields may not honor the full
113 state equation, that is, the solution of the state equation using the conditioned conductivity fields may not

114 reproduce the conditioning piezometric heads. In order for the conductivity fields to be truly conditional
115 to the piezometric heads, in the sense that the solution of the groundwater flow equation matches them,
116 we must proceed iteratively and use piezometric head measurements at multiple time steps. This is an
117 iterative approach akin the iterative approximation of the minimum of a non-linear function by successive
118 linearizations.

119 Under these considerations, the iSS algorithm consists of the following steps:

- 120 1. Define the parameters controlling the sequential simulation: search neighborhood, maximum number
121 of conditioning points to retain within the search neighborhood, and number of realizations to generate.
- 122 2. Define other variables controlling the groundwater flow simulation: initial conditions, boundary con-
123 ditions, sources and sinks.
- 124 3. Generate an initial ensemble of conductivity fields conditioned to measured conductivity data, if any.
125 This ensemble contains N_e realizations discretized into N nodes. Any stochastic simulation technique
126 can be used to initialize this ensemble, whether multiGaussian-based or not. (In the example be-
127 low we will use a hybrid of multiple-point simulation and Gaussian sequential simulation to generate
128 realizations with channel-like patterns of sand and shale.)
- 129 4. Begin iteration in time.
 - 130 (a) Solve, for each conductivity realization, the transient groundwater flow equation from the previous
131 time until time t . The result is an ensemble of piezometric heads H .
 - 132 (b) Normal-score transform the conductivity realizations.
 - 133 (c) Compute the auto- and cross-covariances of \tilde{K} and $H(t)$.
 - 134 (d) Collect measured piezometric heads at the observation locations. These values will be used as
135 conditioning data.
 - 136 (e) Generate a new ensemble of normal-score conductivity fields by multivariate multiGaussian se-
137 quential simulation using both normal-score transformed conductivity and piezometric head as
138 conditioning data. (To account for either conductivity or piezometric head measurement errors,
139 add some error to each measurement for each realization according to a specified (multi-)variate
140 error distribution before using it for conditioning.)
 - 141 (f) Back-transform the normal-score conductivities into conductivities and move on to the next time
142 step.

143 After each time iteration, a new set of conductivity fields results, which is used to forecast the piezometric
144 heads to the next time. It is important to notice that after each iteration, the new set of conductivity fields will

145 have a different statistical structure than the previous one, that is, different local cumulative distributions,
146 different ensemble mean, different ensemble variance, different covariance, and different cross-covariance with
147 the piezometric heads as a result of the conditioning effect in the sequential simulation step. For those readers
148 familiar with the EnKF, there is a fundamental difference in that iSS builds from scratch the new ensemble
149 of conductivities at each time step, whereas in the EnKF, each realization of the ensemble is updated, at
150 each time step, by adding to it a smooth perturbation computed by simple co-kriging.

151 This iterative refinement of the conductivity spatial random function (as given by the ensemble of realiza-
152 tions) results in a final random function that characterizes much better the spatial patterns of conductivity
153 than the initial one. The final random function is truly conditional to the piezometric heads, as it will be
154 shown in the next example.

155 3. Synthetic Example

156 We wish to test the performance of the iSS algorithm in an aquifer with spatial heterogeneity that is not
157 suited for modeling with a multiGaussian random function. The reference field must have bimodal marginal
158 histogram of logconductivities, and conductivities must show strong connectivity for the high values. It is
159 well known that these two aspects are difficult to handle by a multiGaussian random function model [52, 53].

160 We use as reference a synthetic confined aquifer composed of sand and shale. The aquifer occupies an
161 area of 50 m by 50 m by 5 m and it is discretized into a grid of 50 by 50 by 1 cells. The proportion of sand
162 is 0.35, and thus the proportion of shale is 0.65. We impose that the facies spatial distribution be consistent
163 with the channelized training image of Figure 1 taken from Strebelle [22]. (Note that the training image
164 extends over a much larger area than the aquifer.)

165 The synthetic aquifer is built in two steps. In the first step a binary sand/shale realization is generated
166 using the training image in Figure 1 and eight conditioning points distributed as shown in Figure 2. The
167 code used in this first step is SNESIM [22]. In the second step, the facies are populated with log-conductivity
168 values generated, independently and unconditionally, by sequential Gaussian simulation. The code used in
169 this second step is GCOSIM3D [45]. The parameters used for the generation within each facies are listed
170 in Table 1. The resulting field is shown in Figure 3, where the channelized structure is clearly seen. The
171 histogram of log-conductivities in the reference aquifer is shown in Figure 4, it displays two modes at 3.5
172 $\ln(\text{m/d})$ and $-2.5 \ln(\text{m/d})$, its global mean is $-0.9 \ln(\text{m/d})$ and its global standard deviation is $2.9 \ln(\text{m/d})$.

173 Groundwater flow is solved in the aquifer assuming no flow boundary conditions, an initial head set at 8
174 m everywhere, two injection wells (coinciding with the two facies data in the sand facies along the west side

175 of the aquifer) injecting $16 \text{ m}^3/d$ (top one) and $15 \text{ m}^3/d$ (bottom one), and three pumping wells (coinciding
176 with the three facies data in the sand facies along the east side of the aquifer) pumping $7.5 \text{ m}^3/d$, $7.5 \text{ m}^3/d$
177 (top two ones) and $14.5 \text{ m}^3/d$ (bottom one). The aquifer specific storage is homogeneous and equal to 0.03
178 m^{-1} . (See Figure 3 for the location of the facies conditioning data)

179 The transient groundwater flow simulator MODFLOW [e.g., 54, 55] is used to solve the transient ground-
180 water flow equation for a period of 500 days. The time span is discretized into 100 time steps whose length
181 follows a geometric series of ratio 1.02. (The first time step results in 1.6 days.)

182 The piezometric heads are sampled at each time step at the locations indicated below, these observa-
183 tions will be used for conditioning during the first 50 time steps (135.4 days) and then used for validation
184 afterwards. We assume an uncorrelated Gaussian measurement error with zero mean and 0.01 m standard
185 deviation.

186 We have designed six scenarios to make a sensitivity analysis to several parameters. These scenarios
187 are described in Table 2. The scenarios differ between them in the number of conditioning facies data, in
188 the number of realizations in the ensemble, in the maximum number of conditioning data retained in the
189 sequential simulation step for the computation of the local conditional distribution function, and in the
190 number of piezometers at which piezometric head is observed.

191 The procedure to generate the initial set of realizations for all six scenarios is the same one used to
192 generate the reference field. They only differ in the number of facies conditional data; please note that there
193 are no conductivity conditioning data, only the facies is used to condition the realization, this amounts to
194 assume that there is a conductivity observation error equivalent to the variance of the conductivity within
195 each facies. Scenario S2 has no conditioning data, scenarios S1, S4, S5 and S6 use the same 8 conditioning
196 data employed to generate the reference field (see Figure 2), and scenario S3 uses 8 additional conditioning
197 data sampled from the reference field. The location and values of these 16 data can be seen in Figure 5.

198 It is debatable whether we should use the same random function model for the generation of the initial
199 realizations that we used to generate the reference field (in this case the same training image for the facies and
200 the same multivariate distributions for the sand and shale conductivities). However, for the purpose of this
201 paper, which is to analyze the sensitivity of the iSS to different parameters that can affect its performance, to
202 include uncertainty in the random function model would have masked the conclusions about the sensitivity
203 of the algorithm to the parameters under consideration. It is left for another study to check the capability
204 of the algorithm to correct and compensate the erroneous specification of the initial random function. (As
205 already mentioned, the random function model of the ensemble of realizations evolves as new piezometric

206 head data are included, and therefore, it is possible that if enough piezometric head data are available the
207 method may correct an initial misspecification of the random function model similarly as it happens with
208 the NS-EnKF [56].)

209 All scenarios use 25 piezometers for the observation of the piezometric head, except for scenario S5 that
210 uses only 16. The location of these piezometers is shown in Figure 6. This figure also shows the locations of
211 the two piezometers used to illustrate the performance of the method (labeled #6 and #7).

212 4. Analysis

213 We have generated six sets of conductivity realizations according to the parameters in Table 2, next we
214 are going to analyze up to which point they are truly conditional on the piezometric heads in the sense that
215 observations are matched by groundwater flow predictions. Then, we will analyze the different sets against
216 the reference field to check how close these realizations fluctuate about the reference field and how large
217 these fluctuations are. Finally, we will carry out a solute transport exercise to verify how the final fields
218 perform in reproducing information that has not been used in the conditioning process.

219 And finally, we will perform a post-audit analysis to check how these sets of realizations will perform
220 under completely different conditions, more precisely, how these realizations will reproduce solute transport
221 as observed in the reference field.

222 4.1. Conditioning to heads

223 Figure 7 shows the piezometric heads evolution at piezometers #6 and #7 of Figure 6 on the initial sets
224 of realizations, that is, those realizations that have been generated only conditional on facies data, since no
225 piezometric head information has been used yet. As expected, the scatter of the piezometric heads is very
226 wide. Figure 8 shows the same results after having conditioned to piezometric heads in the first 50 time
227 steps (up to the vertical dashed line). Recall that this means that for each scenario 50 sets of realizations
228 have been generated, one for each time step based on the auto- and cross-covariances computed on the set
229 of realizations from the previous time step. The results are quite convincing with regard to the objective
230 of iSS: the generation of realizations that match the observed piezometric heads. Comparing with Figure 7,
231 the reduction of the spread is remarkable. Comparing among the graphs in Figure 8 we can appreciate that
232 the conditioning is not equally good for all scenarios or both piezometers. It is clear that the best results are
233 obtained for the set of realizations with the largest ensemble size and the largest values for all parameters
234 (S3). Using the results for scenario S1 as benchmark we can also conclude that reducing the number of

235 conditioning nodes to be retained within the search neighborhood during the sequential simulation step also
236 reduces the quality of the conditioning to head observations. Finally, we can comment that the apparent
237 improvement noticeable for the case with the smaller ensemble is at the cost of a large bias with respect to
238 the reference field, as we will discuss later.

239 As it has already been mentioned, the set of realizations improves as time passes and new piezometric
240 head data are collected. This can be appreciated in Figure 9 where the piezometric heads predicted at
241 piezometers #6 and #7 with the ensembles of realizations generated at time zero, those generated after 10
242 time steps and those generated after 50 time steps are shown.

243 *4.2. Reproduction of conductivities*

244 From the previous section it is apparent that the final realizations are coherent, from a groundwater
245 flow perspective, with the observed piezometric heads. The question now is how well they approximate the
246 underlying hydraulic conductivity distribution.

247 Figure 10 shows the ensemble mean and ensemble variance computed locally through the initial ensembles
248 of realizations for the different scenarios. Recall that all sets of initial realizations are generated with the
249 hybrid method used to generate the reference field, that is, it combines multipoint geostatistics and Gaussian
250 sequential simulation. Scenario S2 is unconditional, as a consequence the ensemble mean and variance are
251 uniform and equal to the global mean and variance: even if each realization displays channel patterns as
252 in the training image, the fact that no realization has information about the channel locations renders
253 realizations that, in average, are uninformative. Scenario S1 uses five conditioning points in sand close to
254 the two vertical boundaries (where injection and production wells are placed) plus three more in shale in the
255 center; we can notice the effect of conditioning locally in the vicinity of the conditioning points, but we are
256 still far from identifying the continuous channels present in the reference field. Scenario S6 yields virtually
257 the same results as S1 with an ensemble of only 100 realizations. Scenario S3 has double the number of
258 conditioning points well spread over all the aquifer producing realizations that, in average, start to identify
259 the channels present in the reference.

260 Figure 11 shows the ensemble means and ensemble variances for all six scenarios after conditioning to
261 piezometric heads for 50 time steps. As expected, the best results occur for scenario S3, the one with the
262 largest amount of conditional information for both facies and piezometric head. It is the only scenario for
263 which the ensemble mean captures the three channels and their geometry almost perfectly. Scenario S1 with
264 a smaller number of facies conditional data does a good job, too, but introduces a connection between the
265 bottom two channels that does not exist in the reference and the edges of the channels are not as well defined

266 as in S3. Scenario S6, which is the same as S1 but using only 100 realizations yields a channel distribution
267 almost identical with S1 but it tends to overestimated the conductivity values, and the channels edges are
268 even more noisy. The ensemble variance maps support the findings about the ensemble means. The variance
269 for S3 indicates that there is some residual uncertainty in the edges of the channels, but otherwise the
270 patterns are well captured in all realizations. This uncertainty about channel edges is larger in S1, for which
271 there is also large uncertainty regarding the connection between the two bottom channels. Whereas scenario
272 S6 shows an overconfident ensemble variance, with values close to zero almost everywhere. This extreme
273 reduction of the ensemble variance is due to an underestimation of the experimental covariance caused by
274 the use of a relatively small number of realizations. (This phenomenon is similar to the filter inbreeding
275 observable in ensemble Kalman filtering when the ensemble size is small [36].)

276 The ensemble means and variances for scenarios S2 and S5 yield the same conclusions, after 50 time
277 steps, the smaller number of conditioning data than in S1 (whether facies or piezometers) produces sets of
278 realizations that start to delineate loosely the channels but that still display a lot of uncertainty.

279 Finally, for implementation purposes, it is very important to retain a sufficiently large number of condi-
280 tioning points within the search neighborhood for the algorithm to work. Reducing this number from 24 (for
281 S1) to 8 (for S4) results in a mean map that misses the connectivity and a variance map with large values
282 everywhere.

283 Figure 12 shows the evolution of the ensemble mean and ensemble variance maps for S1 for the initial
284 realizations, for those obtained after 10 time steps, and for those obtained after 50 time steps. We can
285 appreciate how the channels become better delineated as times passes, and how the ensemble variance
286 reduces in the areas in which the mean map coincides with the reference.

287 The normal-score transform that is applied to transform the conductivities into Gaussian variates before
288 the sequential simulation step of the algorithm is applied ensures that the final sets of conductivities have
289 the bimodal characteristics of the reference field, as can be corroborated with the histograms of the final
290 realizations displayed in Figure 13. Overall all scenarios show the two modes, with means ranging from
291 $-0.6 \ln(\text{m/s})$ to $0.8 \ln(\text{m/s})$ (reference $-0.9 \ln(\text{m/s})$) and standard deviations from $2.7 \ln(\text{m/s})$ to $2.9 \ln(\text{m/s})$
292 (reference $2.9 \ln(\text{m/s})$) with the best reproduction of the reference histograms for those scenarios that we
293 have already seen that perform best.

294 The normal-score transform is, particularly in this case, a non-linear transformation which not only
295 changes explicitly the marginal histograms of the variables involved, but in this case, it also affects, indirectly,
296 the connectivity, as it can be seen in Figure 14. It is quite astonishing to compare the two columns of this

297 figure; the mean of the normal-score transforms hardly displays any hint of channeling, and its variance is
 298 also quite distant in relative values and patterns from the variance map of the transformed log-conductivities.
 299 This can be explained because the local cumulative distribution functions are different from point to point,
 300 and because of the conditioning effect of the piezometric heads.

301 We can quantify the goodness of the final sets of realizations by computing a single scalar measuring the
 302 accuracy with which the ensemble average matches the reference field and its precision as measured by the
 303 ensemble standard deviation. For this purpose, we define two metrics, the root mean square error (RMSE)
 304 and the average ensemble spread (ES) as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N \left(\ln(K_{ref}(j)) - \overline{\ln(K(j))} \right)^2}, \quad (6)$$

305

$$ES = \sqrt{\frac{1}{N} \sum_{j=1}^N \sigma_j^2}, \quad (7)$$

306 where N is the number of model elements; $\ln(K)_{ref}(j)$ is the value of logconductivity in the reference field
 307 at node j , $\overline{\ln(K(j))}$ is the logconductivity ensemble mean at node j ; σ_i is the ensemble variance at node i .

308 Figure 15 shows the evolution with time step of both the RMSE and the ES. Regarding the RMSE we
 309 can notice that, at the 50th time step, the best scenario is S3, as it was appreciated previously, and the
 310 worst scenario is S4, pinpointing the importance of a good choice of the parameters controlling the sequential
 311 simulation step. As the number of facies conditioning data reduces, the RMSE increases (see the progression
 312 from S3 to S1 to S2). The second worst scenario is the one with the reduced number of piezometers used
 313 for the simulation, indicating the importance of having enough transient piezometric information for the
 314 algorithm to work. (Something that was also observed by Xu et al. [56] in the context of the normal-score
 315 ensemble Kalman filter.) Regarding the ensemble spread we notice how it decreases steadily in time for all
 316 scenarios, indicating a progressively less uncertain set of realizations; although, scenario S6, the one with
 317 only 100 realizations, shows a too dramatic decrease of the ES, which, in this case, indicates overconfidence
 318 associated with a high bias, noticeable for the high value of the RMSE. This behavior of S6 is due to an
 319 underestimation of the experimental covariances, which are computed on a small number of realizations.

320 4.3. Postaudit

321 We have seen how the iSS is capable of generating ensembles of realizations which are conditioned to both
 322 conductivity data and piezometric head data. As a final check on the performance of the algorithm we carry

323 out a post-audit of all realizations and solve a solute transport problem, whose information had not been
324 used for the generation of the realizations. For this purpose, keeping the same groundwater flow conditions
325 as in the rest of the study, we model the release of solute at time zero from the vertical segment labeled
326 “particle injection” in Figure 3 and located at $x = 2.5$ m, and we monitor its arrival at the two control planes
327 displayed in the same figure, located at $x = 15$ m and $x = 45$ m. The total solute is distributed in 1000
328 particles, which are uniformly distributed along the segment and whose movement is tracked in the aquifer
329 using the random-walk particle-tracking code RW3D by Fernández-García et al. [57]. The solute is assumed
330 to be inert, and the porosity of the aquifer is assumed to be uniform and equal to 0.3.

331 Figure 16 shows a summary of the breakthrough curves (BTC) at the two control planes as computed in
332 the initial sets of realizations used for the different scenarios; it shows, in red, the BTC for the reference field,
333 in green the median of the BTCs computed timewise, and in black the 90% confidence interval as delimited
334 by the 5% and 95% percentiles. The reference BTC has a very long tail, associated with all those particles
335 which are introduced in the shale part of the release segment. Overall the BTCs show a large spread for
336 the 90% confidence interval, which, in all cases, encloses the reference BTC. The spread is the smallest for
337 scenario S3, what is coherent with the largest number of conditioning facies data used.

338 Figure 17 shows a summary of the same BTCs as the previous figure but now computed in the ensembles of
339 realizations generated for all scenarios after 50 time steps. The spread of the BTCs has reduced considerably,
340 although now some of the scenarios do not produce a 90% confidence interval fully enclosing the reference
341 BTC.

342 At control plane 1, the one closest to the release segment, scenarios S1, S2, S3 and S4 behave similarly.
343 Scenario S5 fails to capture the fast breakthrough of the first 75% of particles, what is related to the low
344 conductivity values simulated in the upper right corner (see Figure 11), and which is due to the lowest
345 number of conditioning piezometric head data overall, and specifically in the upper part of the aquifer. The
346 results for S6 show the overconfident prediction already observed for conductivities (see Figure 11) and the
347 bias associated to the highest departure from the reference field as measured by the RMSE.

348 At control plane 2 we can see that scenarios S1 and S2 behave similarly and with a good reproduction
349 of the reference BTC. It is somehow surprising that scenario S3, which has been the best so far, produces
350 BTCs that depart from the reference with a clear arrival in two bursts, a first one, which is faster than the
351 reference, and a second one, slower than the reference. At this time we do not have an explanation for this
352 behavior in relation with the proposed method. Scenarios S4 and S5 behave similarly (with slower travel
353 times for S5) and display a larger spread. Finally, scenario S6 shows again its overconfidence and bias.

354 Figure 18 shows the evolution of the BTCs for scenario S1 in the initial realizations, in those generated
355 after 10 time steps, and in those generated after 50 time steps. As before, we can see the reduction in their
356 spread with time as a result of the conditioning to the observed piezometric heads.

357 *4.4. Discussion*

358 From the analysis of the results for the different scenarios we can conclude that the iSS is a method
359 that can generate non-multiGaussian realizations of conductivity inverse conditioned on piezometric heads,
360 that is, to generate conductivity realizations that when used in a groundwater flow model will reproduce the
361 observed heads. The method is not based in any optimization but rather in a covariance-based sequential
362 simulation paradigm.

363 From the different scenarios analyzed we can reach several conclusions, some affect parameters that can
364 be decided by the user, and some parameters that are given. The first parameter that the user must define is
365 how many realizations to generate: it is important to use a sufficiently large set of realizations to avoid the
366 underestimation of the experimental covariances and cross-covariances. In this example 600 realizations were
367 enough; 100 realizations resulted in the collapse of the final ensemble about a single realization. The problem
368 of underestimation of the experimental covariances is also present in the implementation of the EnKF, and
369 there it is solved using techniques of covariance inflation [e.g., 58, 59, 60, 61], those techniques could be a
370 solution here, too. The need to compute the non-stationary covariances from an ensemble of realizations
371 implies that iSS cannot be used to generate a single conditional realization (as it can be done with standard
372 sequential simulation algorithms that use user-defined analytical expressions for the covariances), there is
373 always the need to generate a sufficiently large ensemble of realizations.

374 The standard sequential simulation algorithm has three key aspects that must be decided by the user,
375 which are of importance for the iSS, too: the size and shape of the search neighborhood used for the
376 building of the local conditional probability distribution, the maximum number of points (and their spatial
377 distribution) to retain within the search neighborhood, and the variables to simulate. Regarding the size
378 and shape of the search neighborhood, this is related with the correlation structure of the parameter field,
379 in our example, the overall horizontal correlation of conductivity is much larger than the size of the aquifer
380 due to the presence of the horizontal channels, while the vertical correlation is smaller. We chose a large
381 search neighborhood, of 25 m radius, so that, if centered in the aquifer it covers most of it. Regarding
382 the maximum number of points to retain within the search neighborhood, the results show that a large
383 number is necessary for a good reproduction of the conductivity patters. We have not restricted their spatial
384 disposition and we have chosen always the closest points, but it would be interesting to analyze the impact

385 of distributing these data by quadrants, or by type of conditioning variable. And regarding the variables to
386 simulate, we can choose to generate only conductivities conditioned on both conductivities and piezometric
387 heads, or to generate both conductivities and piezometric heads simultaneously. If we choose to generate
388 only one variable (conductivity) then as the grid is being generated and the number of simulated nodes
389 increases, the conditioning data retained within the search neighborhood contains less and less number of
390 piezometric heads (since the number of conditioning piezometric heads is only 16 or 25 at all times, while the
391 number of conditioning conductivity values reaches 2499 for the simulation of the last node); whereas if we
392 generate both variables simultaneously we can argue that the cross-correlation between conductivities and
393 piezometric heads are persistent throughout all the simulation since until the simulation of the last nodes
394 there will always be data within the search neighborhood of both variables. In this study, we have generated
395 only conductivities; however, we have tried the alternative of generating both variables at the same time,
396 and although the results are not shown here we can say that they are very similar. Therefore, it is preferable
397 to generate only the variable of interest since it is faster.

398 There are other inputs that cannot be controlled by the user, such as the number of conductivity and
399 piezometric head conditioning data. The previous analysis gives the obvious result that the larger the number
400 of conditioning data of either variable, the better the final outcome.

401 The method is still in its early development stages and the code used to test it is a suboptimal research
402 code; for this reason, it does not make much sense to report CPU times. However, we can make an analysis
403 of the algorithm to detect those parts in which iSS will be slower than the EnKF. Both methods are very
404 similar in their implementation: an initial ensemble of realizations is generated, heads are forecasted in
405 time, and, after observing new piezometric heads, the ensemble of conductivities is modified, in the EnKF
406 by updating the fields with a smooth perturbation computed by simple co-kriging, and in the iSS new
407 fields are generated conditioned on the newly observed piezometric heads by sequential simulation. In this
408 respect, the initial generation of the ensemble of conductivities and the forecast step takes the same time
409 for both algorithms. The difference lies in the updating. The EnKF relies in co-kriging, whereas iSS relies
410 on sequential simulation. For co-kriging, the only non-stationary cross-covariances needed are between the
411 piezometric head observation locations and the conductivities/piezometric heads at all nodes; if there are N_h
412 piezometers and N discretization cells, the number of cross-covariance values to compute is $2N_h \times N$. For
413 sequential simulation, at each node we have to solve a co-kriging equation in which all nodes (and all variables)
414 within the search neighborhood may be involved; therefore, for sequential simulation we have to precompute
415 the auto- and cross-covariances between each node and all nodes within the search neighborhood; if the search

416 neighborhood contains N_{search} nodes, the number of covariance values to calculate are $N_{search} \times N$, with
417 N_{search} generally much larger than N_h . (Do not confuse the number of cells within the search neighborhood
418 with the maximum number of conditioning points to be retained to perform kriging, which is always limited
419 to a small number.) Therefore, in precomputing the covariances, iSS can be one-order of magnitude slower
420 than the EnKF. Then, kriging is always faster than sequential simulation; therefore, the EnKF will be faster
421 than iSS in the same proportion as the kriging algorithm chosen is faster than the sequential simulation
422 algorithm used.

423 Finally, from an implementation point of view, there is a critical parameter that we have not mentioned
424 before but which controls the feasibility of the method, which is the variance of an uncorrelated error that
425 must be added to the experimental covariances and cross-covariances. It is well known that the experimental
426 covariances are not positive definite and, therefore, its direct use in the simple kriging equations to compute
427 the kriging mean and kriging variance during the sequential simulation algorithm could yield singular kriging
428 matrices or, if the kriging matrix is not singular and the kriging equations can be solved, the resulting kriging
429 variance may be negative. To solve this problem we must first yield the experimental covariances positive
430 definite. Yao and Journel [62] propose a method based in fast Fourier transforms to solve this problem. A
431 less elegant, but simpler approach, is to add a nugget effect (the variance of an uncorrelated error) to the
432 computed experimental covariances. In our study we had to add a nugget effect equal to 1% of the total
433 variance to the diagonal of the experimental covariance matrices to render them positive definite. This is a
434 parameter that must be fine-tuned, searching for the smallest value that will make the sequential simulation
435 possible.

436 **Acknowledgements** The first author acknowledges the financial support from the China Scholarship
437 Council (CSC). Financial support to carry out this work was also received from the Spanish Ministry of
438 Economy and Competitiveness through project CGL2011-23295. We thank the three reviewers for their
439 thorough review and their insightful comments, which have helped to improve the final manuscript.

- 440 [1] Zhou, H., Gómez-Hernández, J.J., Li, L.. Inverse methods in hydrogeology: evolution and recent
441 trends. *Advances in Water Resources* 2014;63:22–37.
- 442 [2] Capilla, J.E., Llopis-Albert, C.. Gradual conditioning of non-gaussian transmissivity fields to flow and
443 mass transport data: 1. theory. *Journal of hydrology* 2009;371(1):66–74.

- 444 [3] Hu, L.. Gradual deformation and iterative calibration of gaussian-related stochastic models. *Mathe-*
445 *matical Geology* 2000;32(1):87–108.
- 446 [4] Gómez-Hernández, J., Sahuquillo, A., Capilla, J.. Stochastic simulation of transmissivity fields con-
447 ditional to both transmissivity and piezometric data–1. theory. *Journal of Hydrology(Amsterdam)*
448 1997;203(1):167–174.
- 449 [5] Wen, X.H., Capilla, J.E., Deutsch, C.V., Gómez-Hernández, J.J., Cullick, A.S.. A program to create
450 permeability fields that honor single-phase flow rate and pressure data. *Computers & Geosciences*
451 1999;25(3):217–230.
- 452 [6] Fu, J., Jaime Gómez-Hernández, J.. Uncertainty assessment and data worth in groundwater flow and
453 mass transport modeling using a blocking markov chain monte carlo method. *Journal of Hydrology*
454 2009;364(3):328–341.
- 455 [7] Oliver, D., Cunha, L., Reynolds, A.. Markov chain monte carlo methods for conditioning a permeability
456 field to pressure data. *Mathematical Geology* 1997;29(1):61–91.
- 457 [8] Alcolea, A., Carrera, J., Medina, A.. Pilot points method incorporating prior information for solving
458 the groundwater flow inverse problem. *Advances in Water Resources* 2006;29(11):1678–1689.
- 459 [9] Wen, X., Deutsch, C., Cullick, A.. Construction of geostatistical aquifer models integrating dynamic
460 flow and tracer data using inverse technique. *Journal of Hydrology* 2002;255(1):151–168.
- 461 [10] RamaRao, B., LaVenue, A., De Marsily, G., Marietta, M.. Pilot point methodology for automated
462 calibration of an ensemble of conditionally simulated transmissivity fields: 1. theory and computational
463 experiments. *Water Resources Research* 1995;31(3):475–493.
- 464 [11] Franssen, H., Gómez-Hernández, J., Sahuquillo, A.. Coupled inverse modelling of groundwater flow
465 and mass transport and the worth of concentration data. *Journal of Hydrology* 2003;281(4):281–295.
- 466 [12] Gordon, N., Salmond, D., Smith, A.. Novel approach to nonlinear/non-gaussian bayesian state
467 estimation. In: *Radar and Signal Processing, IEE Proceedings F*; vol. 140. IET; 1993, p. 107–113.
- 468 [13] Losa, S., Kivman, G., Schröter, J., Wenzel, M.. Sequential weak constraint parameter estimation in
469 an ecosystem model. *Journal of Marine Systems* 2003;43(1):31–49.

- 470 [14] Evensen, G.. The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean*
471 *Dynamics* 2003;53(4):343–367.
- 472 [15] Nowak, W.. Best unbiased ensemble linearization and the quasi-linear kalman ensemble generator.
473 *Water Resources Research* 2009;45(4).
- 474 [16] Van Leeuwen, P.J., Evensen, G.. Data assimilation and inverse methods in terms of a probabilistic
475 formulation. *Monthly Weather Review* 1996;124(12):2898–2913.
- 476 [17] Sun, A., Morris, A., Mohanty, S.. Sequential updating of multimodal hydrogeologic parameter fields
477 using localization and clustering techniques. *Water Resources Research* 2009;45(7):W07424.
- 478 [18] Simon, E., Bertino, L.. Application of the gaussian anamorphosis to assimilation in a 3-d cou-
479 pled physical-ecosystem model of the north atlantic with the enkf: a twin experiment. *Ocean Science*
480 2009;5(4):495–510.
- 481 [19] Aanonsen, S., Nævdal, G., Oliver, D., Reynolds, A., Vallès, B.. The ensemble kalman filter in
482 reservoir engineering—a review. *SPE Journal* 2009;14(3):393–412.
- 483 [20] Guardiano, F., Srivastava, R.. Multivariate geostatistics: beyond bivariate moments. In: Soares, A.,
484 editor. *Geostatistics-Troia*. Dordrecht: Kluwer Academic Publ; 1993, p. 133–144.
- 485 [21] Strébel, S.. Sequential simulation drawing structures from training images. Ph.D. thesis; Stanford
486 University. 187pp; 2000.
- 487 [22] Strebelle, S.. Conditional simulation of complex geological structures using multiple-point statistics.
488 *Mathematical Geology* 2002;34(1):1–21.
- 489 [23] Zhang, T., Switzer, P., Journel, A.. Filter-based classification of training image patterns for spatial
490 simulation. *Mathematical Geology* 2006;38(1):63–80.
- 491 [24] Arpat, G.B., Caers, J.. Conditional simulation with patterns. *Mathematical Geology* 2007;39(2):177–
492 203.
- 493 [25] Mariethoz, G., Renard, P., Straubhaar, J.. The direct sampling method to perform multiple-point
494 geostatistical simulations. *Water Resources Research* 2010;46(11):W11536.

495 [26] Li, L., Srinivasan, S., Zhou, H., Gómez-Hernández, J.J.. Simultaneous estimation of geologic and
496 reservoir state variables within an ensemble-based multiple-point statistic framework. *Mathematical*
497 *Geosciences* 2013;:1–27.

[27] Zhou, H., Gómez-Hernández, J.J., Li, L.. A pattern-search-based inverse method. *Water Resources*
498 *Research* 2012;48(3).

500 [28] Chugunova, T.L., Hu, L.Y.. Multiple-point simulations constrained by continuous auxiliary data.
501 *Mathematical Geosciences* 2008;40(2):133–146.

502 [29] Mirowski, P.W., Tetzlaff, D.M., Davies, R.C., McCormick, D.S., Williams, N., Signer, C.. Stationar-
503 ity scores on training images for multipoint geostatistics. *Mathematical Geosciences* 2009;41(4):447–474.

504 [30] Boucher, A.. Considering complex training images with search tree partitioning. *Computers & Geo-*
505 *sciences* 2009;35(6):1151–1158.

506 [31] Honarkhah, M., Caers, J.. Stochastic simulation of patterns using distance-based pattern modeling.
507 *Mathematical Geosciences* 2010;42(5):487–517.

508 [32] Hu, L., Zhao, Y., Liu, Y., Scheepens, C., Bouchard, A.. Updating multipoint simulations using the
509 ensemble kalman filter. *Computers & Geosciences* 2013;51:7–15.

510 [33] Mustapha, H., Dimitrakopoulos, R.. High-order stochastic simulation of complex spatially distributed
511 natural phenomena. *Mathematical Geosciences* 2010;42(5):457–485.

512 [34] Liu, N., Oliver, D.. Critical evaluation of the ensemble kalman filter on history matching of geologic
513 facies. In: *SPE Reservoir Simulation Symposium*. 2005,.

514 [35] Gu, Y., Oliver, D.. An iterative ensemble kalman filter for multiphase fluid flow data assimilation.
515 *SPE Journal* 2007;12(4):438–446.

516 [36] Franssen, H., Kinzelbach, W.. Real-time groundwater flow modeling with the ensemble kalman filter:
517 Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*
518 2008;44(9):W09408.

519 [37] Simon, E., Bertino, L.. Gaussian anamorphosis extension of the denkf for combined state parameter
520 estimation: Application to a 1d ocean ecosystem model. *Journal of Marine Systems* 2012;89(1):1–18.

521 [38] Bertino, L., Evensen, G., Wackernagel, H.. Sequential data assimilation techniques in oceanography.
522 International Statistical Review 2003;71(2):223–241.

523 [39] Bertino, L., Hollard, A., Evensen, G., Wackernagel, H.. An ensemble kalman filter for non-gaussian
524 variables. In: EGS-AGU-EUG Joint Assembly; vol. 1. 2003, p. 5171.

525 [40] Bocquet, M., Pires, C., Wu, L.. Beyond gaussian statistical modeling in geophysical data assimilation.
526 Monthly Weather Review 2010;138(8):2997–3023.

527 [41] Béal, D., Brasseur, P., Brankart, J., Ourmières, Y., Verron, J., et al. Characterization of mixing
528 errors in a coupled physical biogeochemical model of the north atlantic: implications for nonlinear
529 estimation using gaussian anamorphosis. Ocean Science 2010;(6).

530 [42] Schöniger, A., Nowak, W., Franssen, H.. Parameter estimation by ensemble kalman filters with
531 transformed data: Approach and application to hydraulic tomography. Water Resources Research
532 2012;48(4):W04502.

533 [43] Zhou, H., Gómez-Hernández, J., Hendricks Franssen, H., Li, L.. An approach to handling non-
534 gaussianity of parameters and state variables in ensemble kalman filtering. Advances in Water Resources
535 2011;34(7):844–864.

536 [44] Li, L., Zhou, H., Hendricks Franssen, H., Gómez-Hernández, J.. Groundwater flow inverse mod-
537 eling in non-multigaussian media: performance assessment of the normal-score ensemble kalman filter.
538 Hydrology and Earth System Sciences 2012;16(2):573.

539 [45] Gómez-Hernández, J.J., Journel, A.G.. Joint sequential simulation of Multi-Gaussian fields. In: Soares,
540 A., editor. Geostatistics Tróia '92; vol. 1. Dordrecht: Kluwer Academic Publishers; 1993, p. 85–94.

541 [46] Gómez-Hernández, J.J., Cassiraga, E.F.. Theory and practice of sequential simulation. In: Armstrong,
542 M., Dowd, P., editors. Geostatistical Simulations. Kluwer Academic Publishers; 1994, p. 111–124.

543 [47] Deutsch, C.V., Journel, A.G.. Geostatistical software library and user&s guide; vol. 1996. Oxford
544 university press New York; 1992.

545 [48] Xu, T., Gómez-Hernández, J.J.. Inverse sequential simulation: A new approach for the characterization
546 of hydraulic conductivities demonstrated on a non-gaussian field. Water Resources Research 2015;.

- 547 [49] Fiori, A., Indelman, P., Dagan, G.. Correlation structure of flow variables for steady flow toward
548 a well with application to highly anisotropic heterogeneous formations. *Water Resources Research*
549 1998;34(4):699–708.
- 550 [50] Bear, J.. *Dynamics of fluids in porous media*. New York, 764pp: American Elsevier Pub. Co.; 1972.
551 ISBN 9780444001146.
- 552 [51] Gómez-Hernández, J.J., Froidevaux, R., Biver, P.. Exact conditioning to linear constraints in kriging
553 and simulation. In: Lueanghton, O., Deutsch, C., editors. *Geostatistics, Banff 2004, Proceedings of the*
554 *Seventh International Geostatistics Congress, September 26- October 1, 2004, Banff, Canada*. Kluwer
555 Academic Publishers, Dordrecht, Holland; 2005, p. 999–1005.
- 556 [52] Gómez-Hernández, J., Wen, X.H.. Probabilistic assessment of travel times in groundwater modeling.
557 *Stochastic Hydrology and Hydraulics* 1994;8(1):19–55.
- 558 [53] Gómez-Hernández, J.J., Wen, X.H.. To be or not to be multiGaussian: A reflection in stochastic
559 hydrogeology. *Advances in Water Resources* 1998;21(1):47–61.
- 560 [54] McDonald, M., Harbaugh, A.. *A modular three-dimensional finite-difference ground-water flow model*.
561 Scientific Publications Company; 1984.
- 562 [55] Harbaugh, A., et al. *MODFLOW-2000, the US Geological Survey modular ground-water model: User*
563 *guide to modularization concepts and the ground-water flow process*. US Geological Survey; 2000.
- 564 [56] Xu, T., Jaime Gómez-Hernández, J., Zhou, H., Li, L.. The power of transient piezometric head
565 data in inverse modeling: an application of the localized normal-score enkf with covariance inflation in
566 a heterogenous bimodal hydraulic conductivity field. *Advances in Water Resources* 2013;54:100–118.
- 567 [57] Fernández-García, D., Illangasekare, T., Rajaram, H.. Differences in the scale dependence of
568 dispersivity and retardation factors estimated from forced-gradient and uniform flow tracer tests in
569 three-dimensional physically and chemically heterogeneous porous media. *Water Resources Research*
570 2005;41(3):W03012.
- 571 [58] Wang, X., Bishop, C.. A comparison of breeding and ensemble transform kalman filter ensemble
572 forecast schemes. *Journal of the atmospheric sciences* 2003;60(9):1140–1158.
- 573 [59] Anderson, J.. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*
574 *A* 2007;59(2):210–224.

- 575 [60] Li, H., Kalnay, E., Miyoshi, T.. Simultaneous estimation of covariance inflation and observa-
576 tion errors within an ensemble kalman filter. Quarterly Journal of the Royal Meteorological Society
577 2009;135(639):523–533.
- 578 [61] Kurtz, A., Lugolobi, F., Salvucci, G.. Germanium-silicon as a flow path tracer: Application to the
579 rio icacos watershed. Water Resources Research 2011;47(6):W06516.
- 580 [62] Yao, T., Journel, A.G.. Automatic modeling of (cross) covariance tables using fast fourier transform.
581 Mathematical Geology 1998;30(6):589–615.

Table 1: Parameters of the random functions describing the spatial continuity of the sand and shale log-conductivities

Facies	Proportion	Mean [ln [m/d]]	Std.dev [ln [m/d]]	Variogram type	λ_x [m]	λ_y [m]	sill
Sand	0.35	3.5	1.0	exponential	20	20	1
Shale	0.65	-2.5	0.6	exponential	20	20	0.35

Table 2: Definition of scenarios

Scenario	S1	S2	S3	S4	S5	S6
Number of observation piezometers	25	25	25	25	16	25
Number of maximum search nodes	24	24	24	8	24	24
Size of the ensemble of realizations	600	600	600	600	600	100
Number of conditional facies data	8	0	16	8	8	8

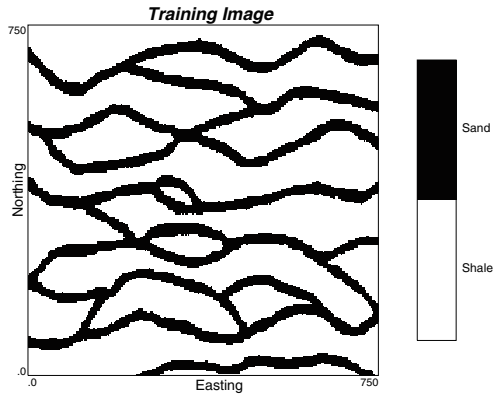


Figure 1: Training image used to generate the ensemble of binary facies realizations

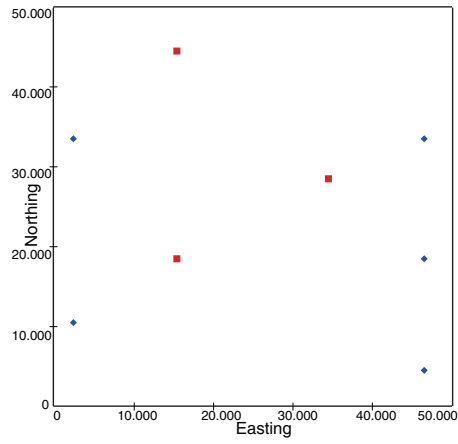


Figure 2: Location of the facies data, red nodes for shale and green nodes for sand

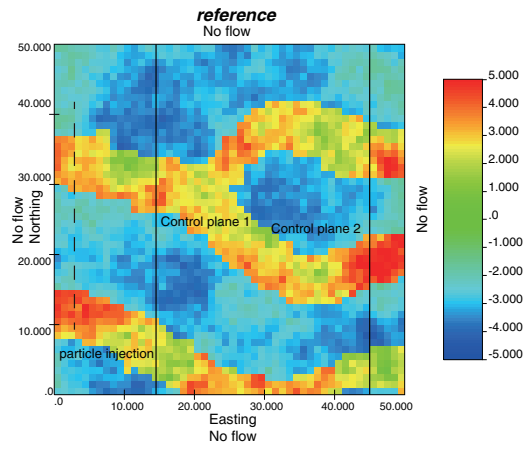


Figure 3: Reference field

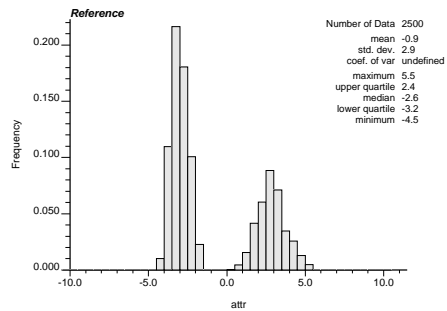


Figure 4: Histogram of the reference field

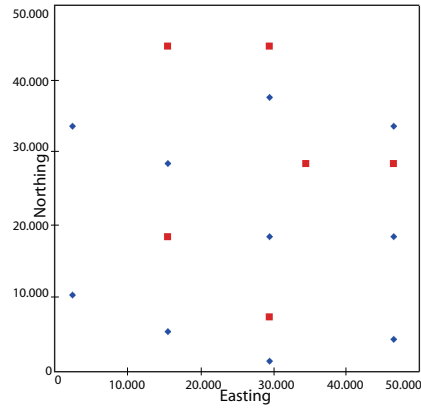
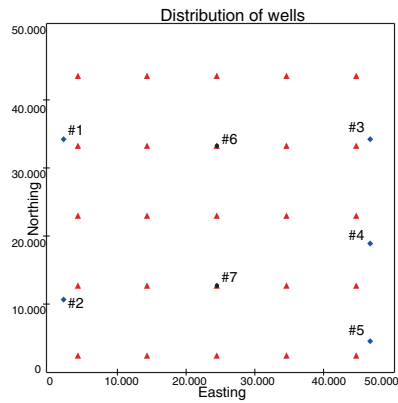
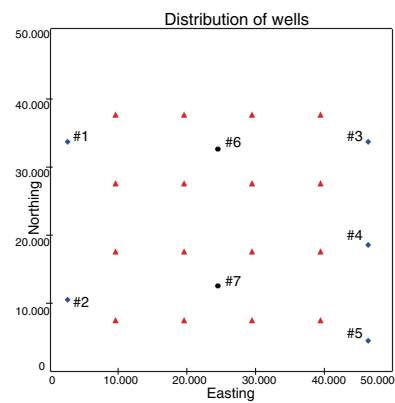


Figure 5: Location of the facies data for scenario S3, the one with a larger number of conditioning points; red nodes for shale and green nodes for sand



a



b

Figure 6: (a) Distribution of piezometers of scenarios S1, S2, S3, S4 and S6; (b) Distribution of piezometers for scenario S5. For reference, the injection (#1, #2) and pumping wells (#3, #4 and #5) are also shown. The piezometric head evolution at piezometers #6, #7 is used in the analysis.

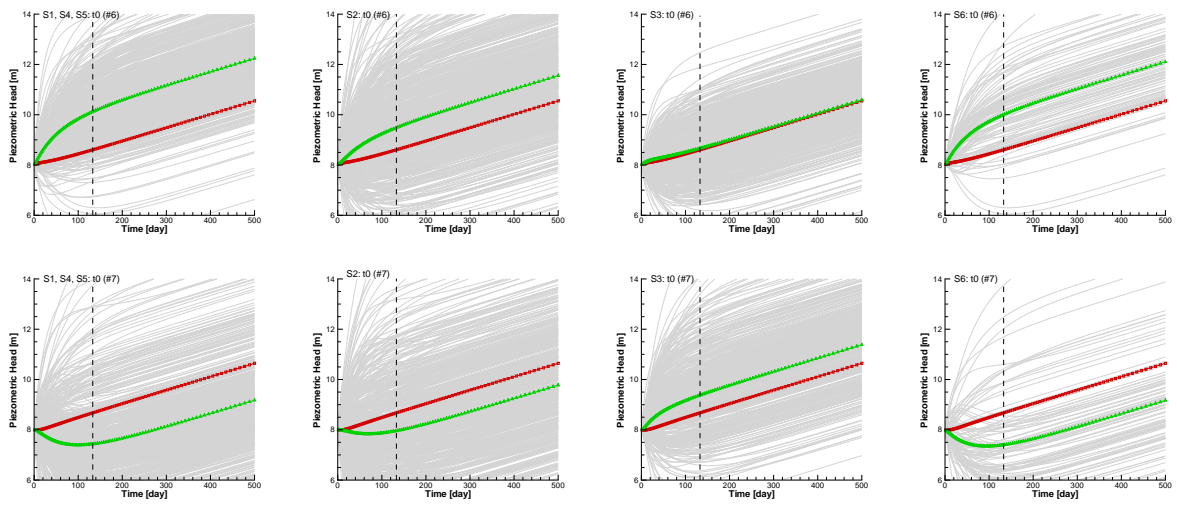


Figure 7: Evolution of piezometric head at piezometers #6 (top row) and #7 (bottom row) on the initial sets of realizations for the different scenarios. Scenarios S1, S4 and S5 share the same set of initial realizations (first column), scenarios S2, S3 and S6 are shown in the second, third and fourth columns, respectively. The gray lines correspond to the evolution in each realization. The red line is the piezometric evolution in the reference, and the green line is the time average of all gray lines. The vertical dashed lines indicate the end of the conditioning data.

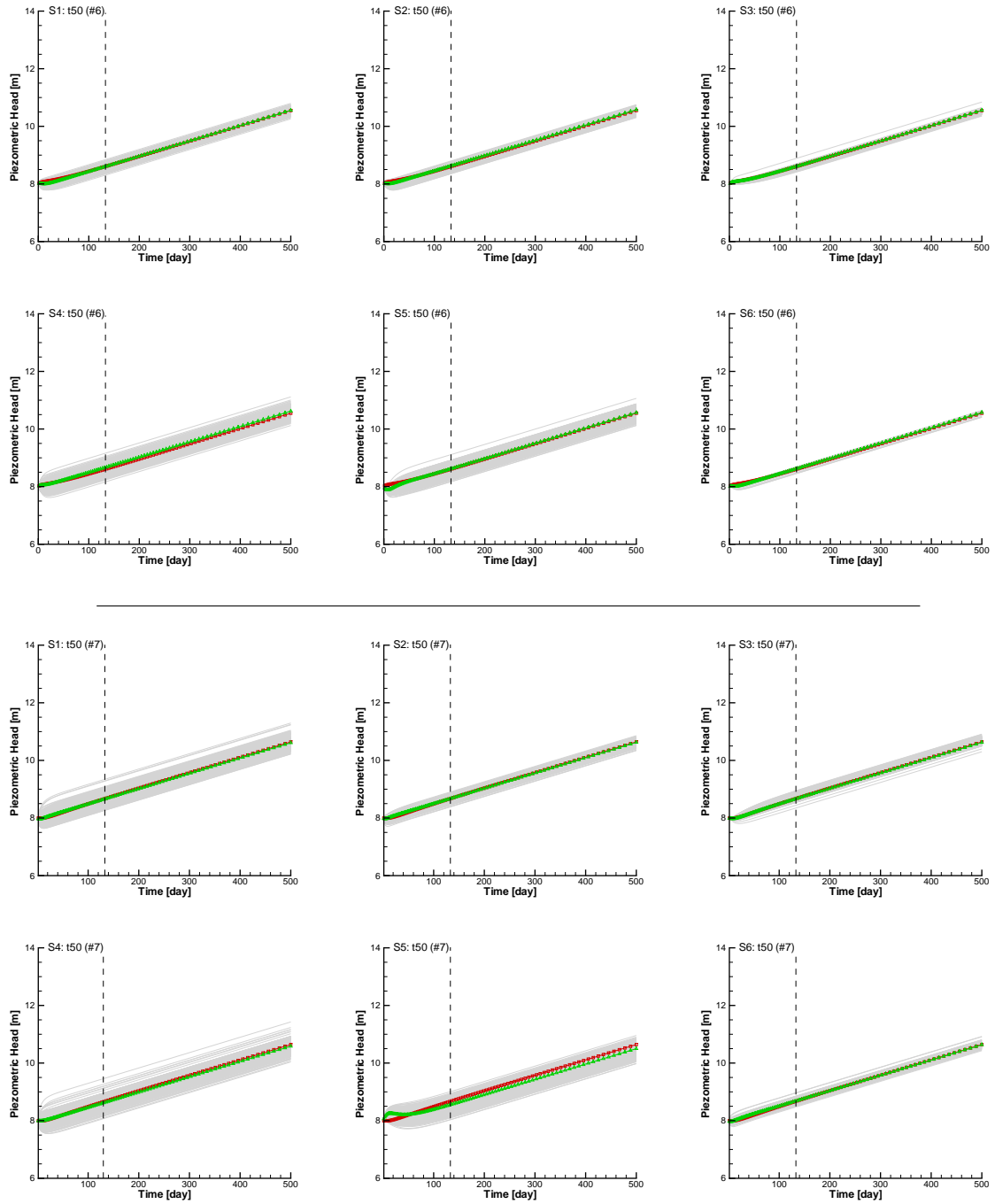


Figure 8: Evolution of piezometric head at piezometers #6 (top two rows) and #7 (bottom two rows) on the sets of realizations generated by iSS after conditioning to piezometric heads during the first 50 time steps (vertical dashed line) for the different scenarios. Meaning of lines is the same as previous figure.

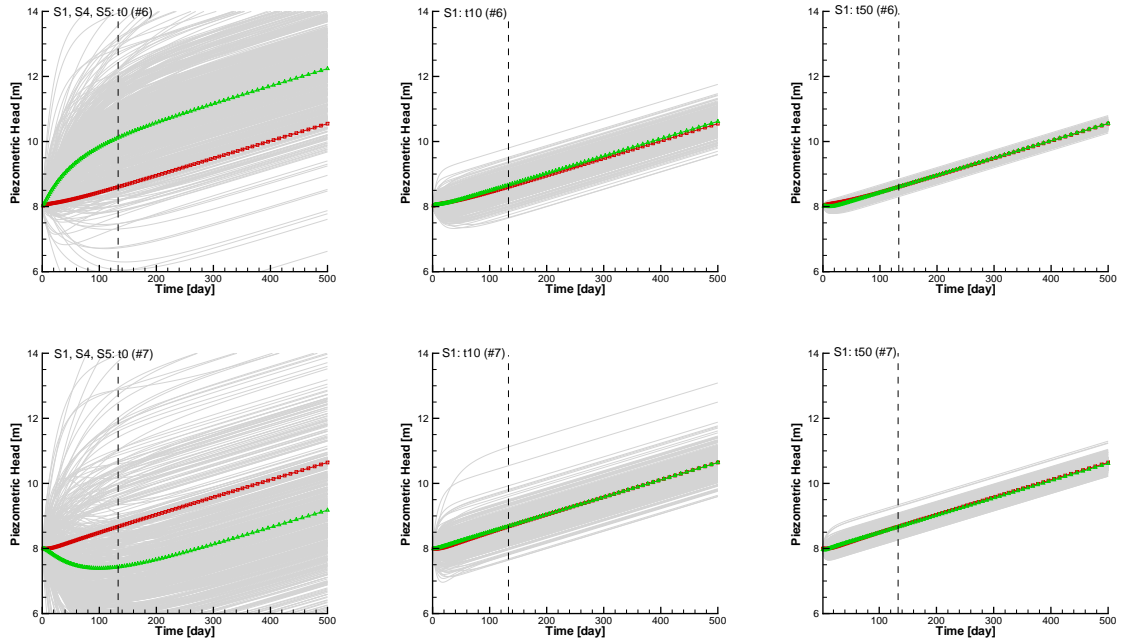


Figure 9: Evolution of piezometric head at piezometers #6 (top row) and #7 (bottom row) on the sets of realizations generated by iSS for scenario S1 at three different times: initial, after 10 time steps, and after 50 time steps. Meaning of lines is the same as previous figure.

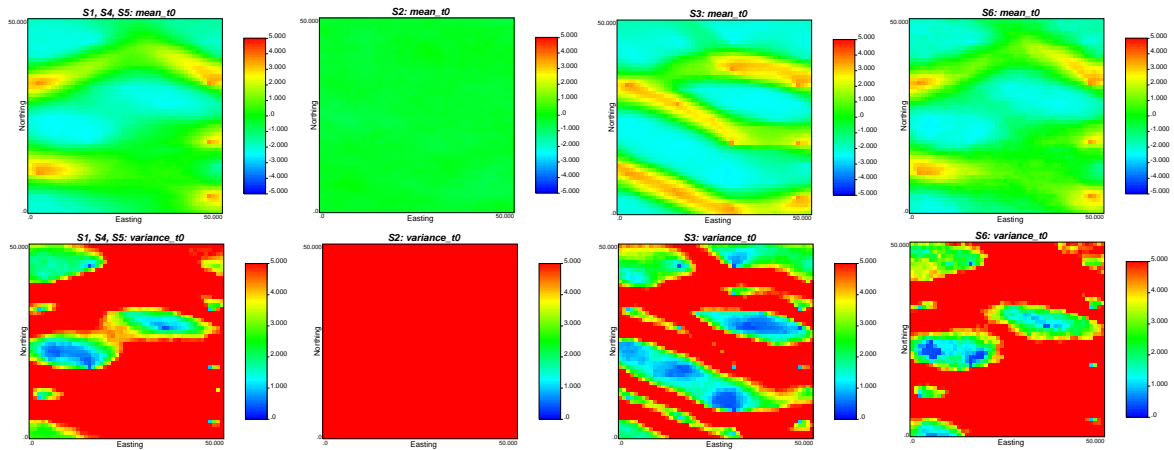


Figure 10: Ensemble mean (top row) and ensemble variance (bottom row) of logconductivity for the sets of initial realizations for the different scenarios. First column is for scenarios S1, S4 and S5, second column for S2, third column for S3, and fourth column for S6.

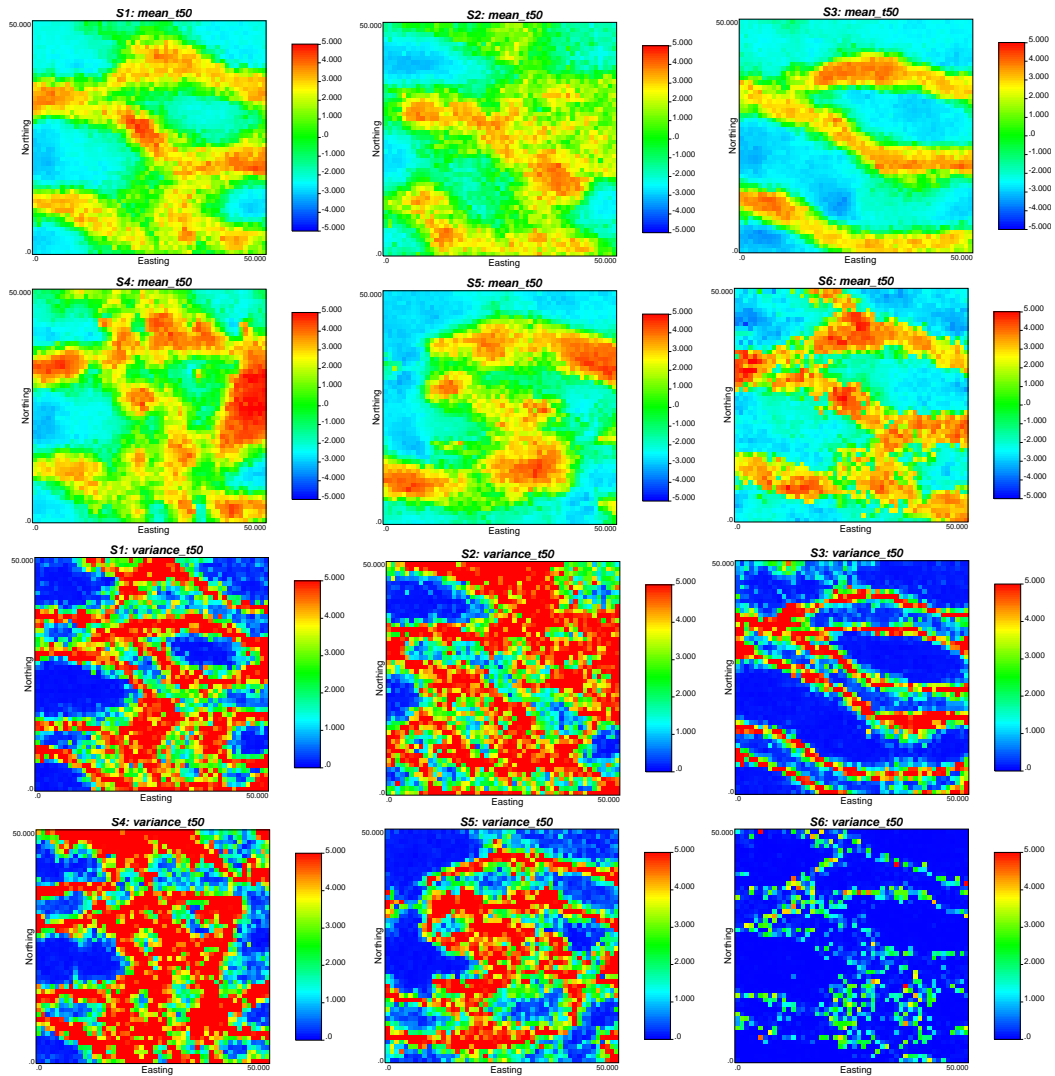


Figure 11: Ensemble mean (top two rows) and ensemble variance (bottom two rows) of logconductivity for the sets of realizations for the different scenarios obtained after 50 time steps.

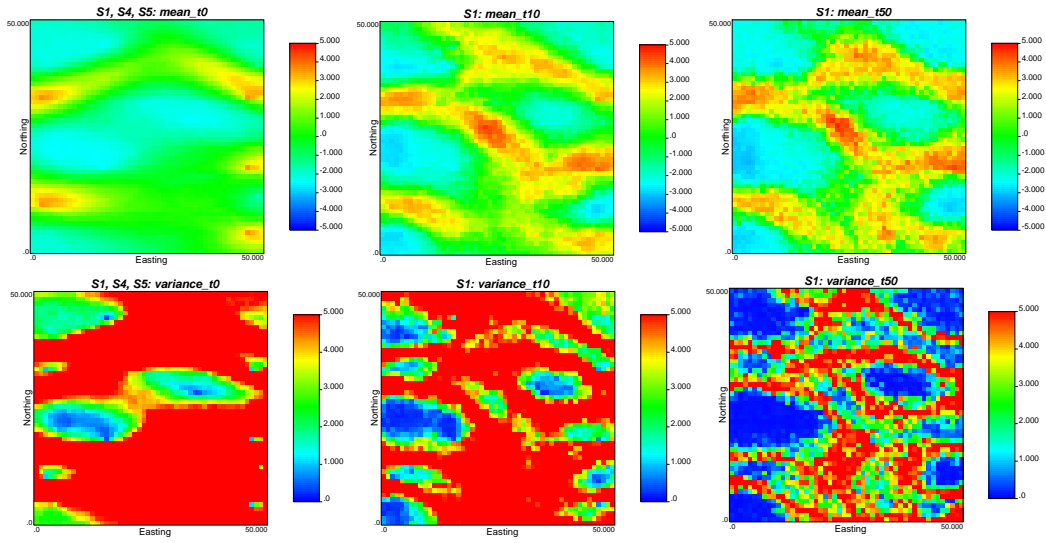


Figure 12: Ensemble mean (top row) and ensemble variance (bottom row) of logconductivity for the sets of realizations generated by iSS for scenario S1 at three different times: initial, after 10 time steps, and after 50 time steps.

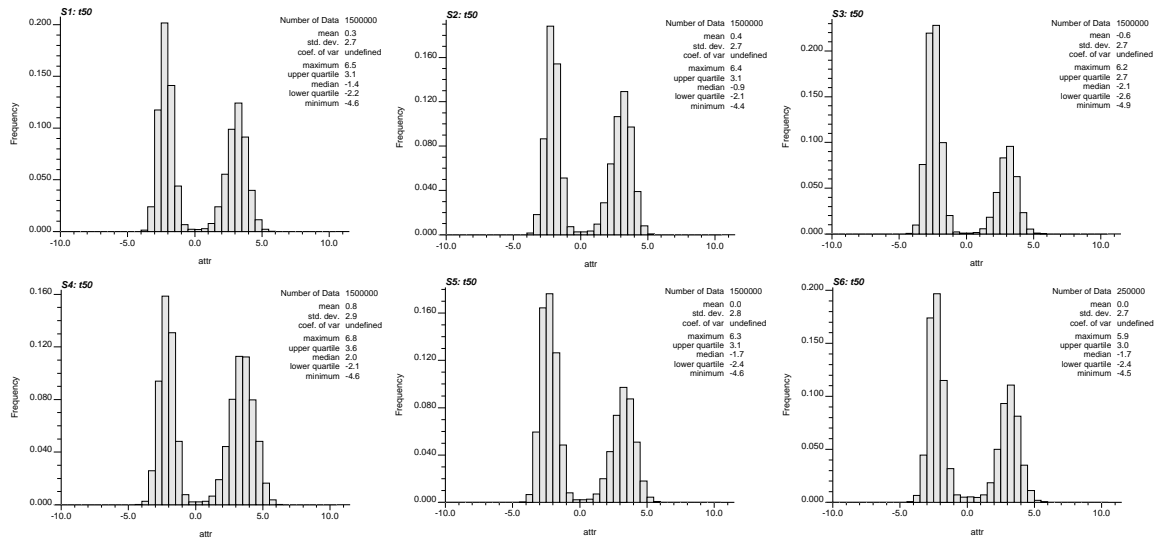


Figure 13: Histograms of logconductivity for the sets of realizations for the different scenarios obtained after 50 time steps.

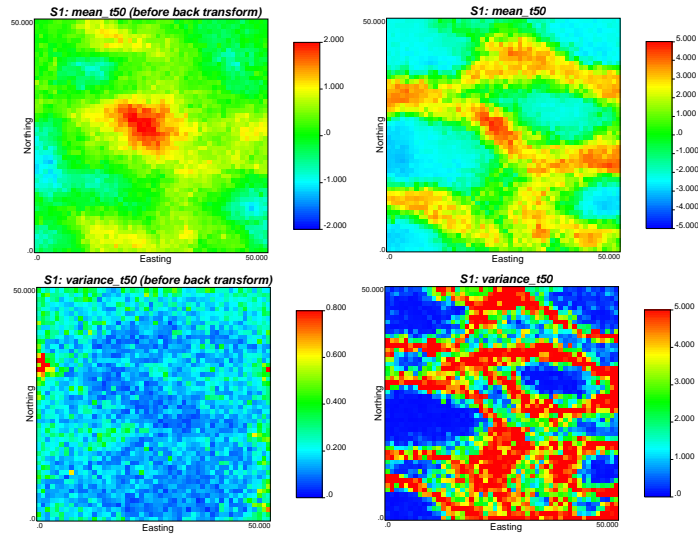


Figure 14: Ensemble mean (top row) and ensemble variance (bottom row) of the ensemble of normal scores of logconductivity as generated by the sequential simulation algorithm for scenario S1 after 50 time steps, and of the corresponding log-conductivity fields after backtransforming the normal-scores.

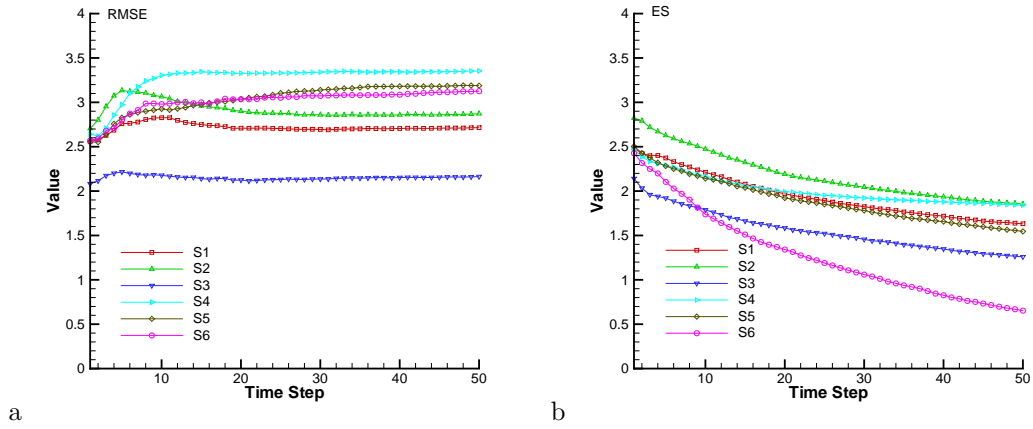


Figure 15: RMSE and ES of all the scenarios.

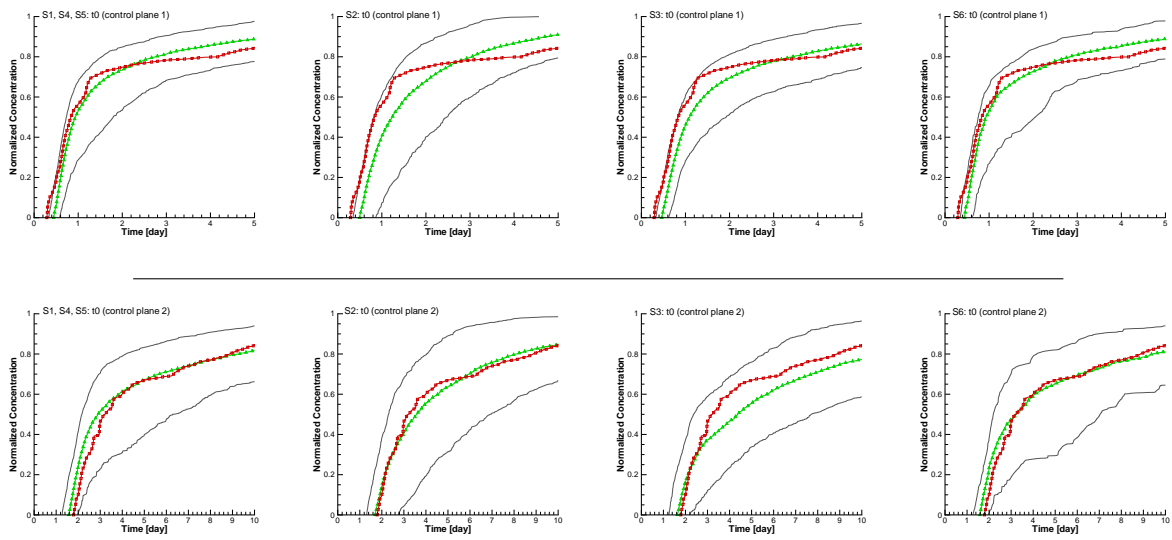


Figure 16: Solute breakthrough curves (BTC) at control planes 1 (top row) and 2 (bottom row) on the initial sets of realizations used in the different scenarios. The red line is the BTC in the reference, the green line is the median (computed timewise) of the BTCs from the different realizations, and the two black lines mark the 90% confidence interval bounded by the 5% and 95% percentiles.

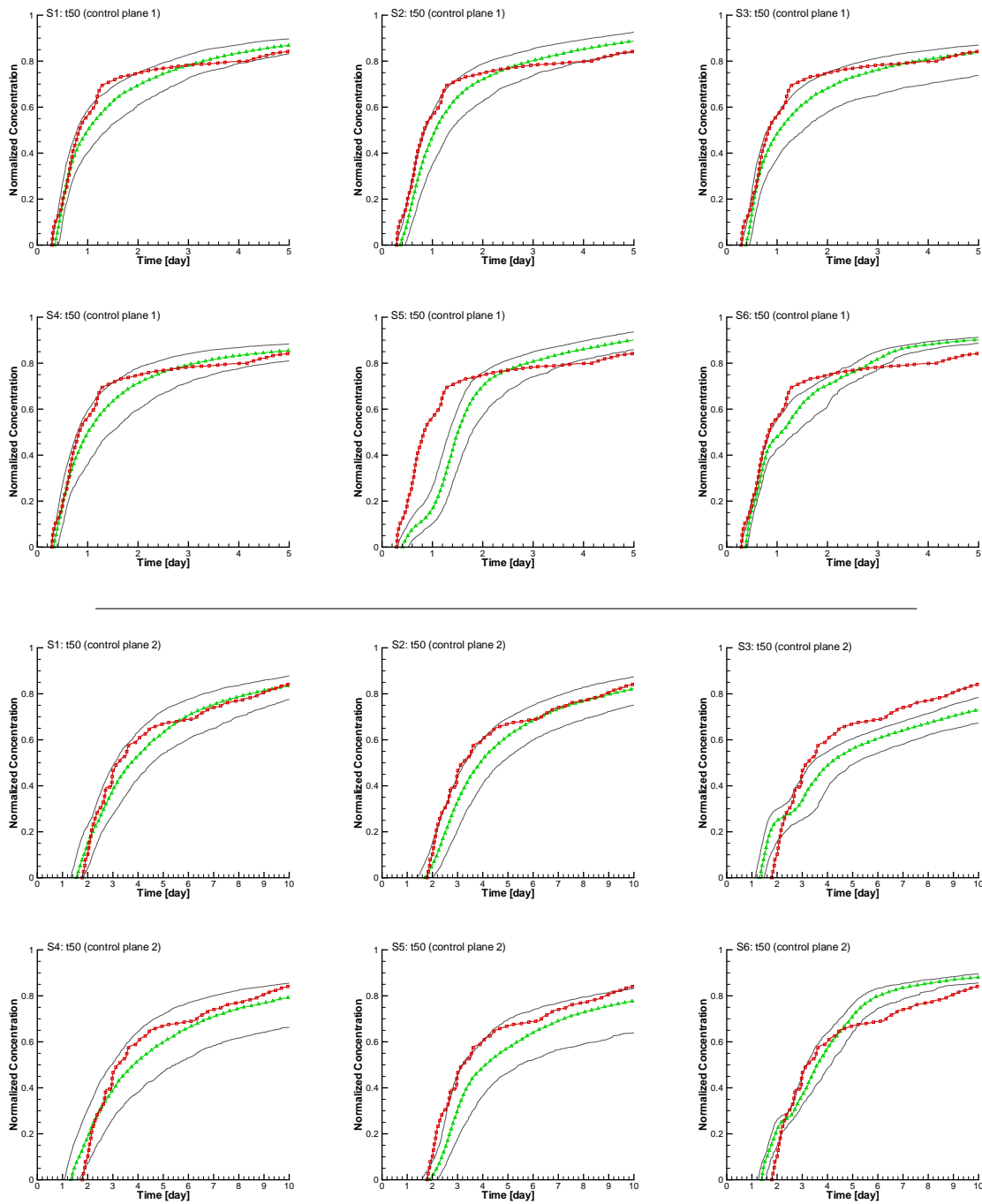


Figure 17: Solute breakthrough curves (BTC) at control planes 1 (top two rows) and 2 (bottom two rows) on the sets of realizations obtained for the different scenarios generated after 50 time steps. Meaning of lines is the same as in previous figure.

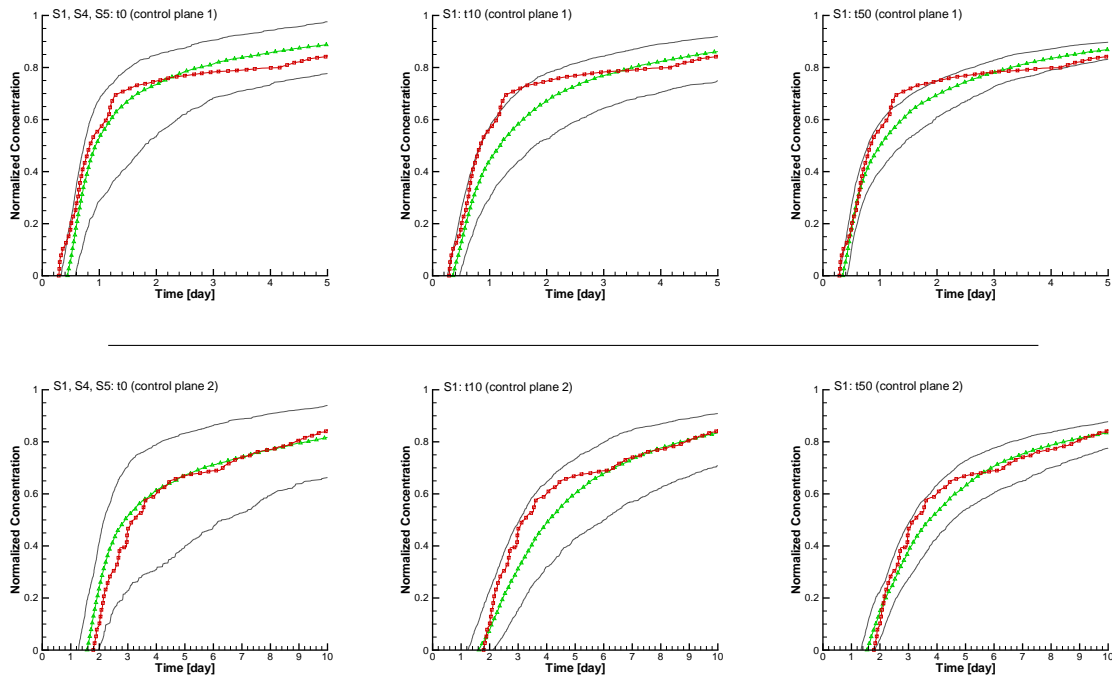


Figure 18: Solute breakthrough curves (BTC) at control planes 1 (top row) and 2 (bottom row) for scenario S1 on the initial set of realizations, those generated after 10 time steps, and those generated after 50 time steps. Meaning of lines is the same as in previous figure.