Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: workflow and case study

Hermans Thomas, F.R.S.-FNRS Postdoctoral Researcher, ArGEnCo Department, Applied Geophysics, University of Liege

Nguyen Frédéric, ArGEnCo Department, Applied Geophysics, University of Liege

Caers Jef, Energy Resources Department, Stanford University

Corresponding author

Hermans Thomas, Chemin des Chevreuils 1, 4000 Liège (Sart Tilman), Belgium. +32 43669263, thomas.hermans@ulg.ac.be

The final version of this paper is published in *Water Resources Research*, a publication of the AGU.

Please cite as:

Hermans, T., Nguyen, F., and Caers, J. 2015. Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: workflow and case study. *Water Resources Research*, 5, 5332-5352, doi: 10.1002/2014WR016460.

Key points

- The workflow assesses posterior uncertainty in model and geological scenario
- ERT data is used twice: to validate scenarios and condition MPS simulations
- The workflow can be adapted to many contexts and methods

Abstract

In inverse problems, investigating uncertainty in the posterior distribution of model parameters is as important as matching data. In recent years, most efforts have focused on techniques to sample the posterior distribution with reasonable computational costs. Within a Bayesian context, this posterior depends on the prior distribution. However, most of the studies ignore modeling the prior with realistic geological uncertainty. In this paper, we propose a workflow inspired by a Popper-Bayes philosophy, that data should first be used to falsify models, then only be considered for matching. We propose a workflow consisting of three steps: (1) in defining the prior, we interpret multiple alternative geological scenarios from literature (architecture of facies) and site specific data (proportions of facies). Prior spatial uncertainty is modeled using multiple-point geostatistics, where each scenario is defined using a training image. (2) We validate these prior geological scenarios by simulating electrical resistivity tomography (ERT) data on realizations of each scenario and comparing them to field ERT in a lower dimensional space. In this second step, the idea is to probabilistically falsify scenarios with ERT, meaning that scenarios which are incompatible receive an updated probability of zero while compatible scenarios receive a non-zero updated belief. (3) We constrain the hydrogeological model with hydraulic head and ERT using a stochastic search method. The workflow is applied to a synthetic and a field case studies in an alluvial aquifer. This study highlights the importance of considering and estimate prior uncertainty (without data) through a process of probabilistic falsification.

Keywords: prior uncertainty, Popper-Bayes, training image, geological scenario, electrical resistivity tomography, probability perturbation method

1. Introduction

Solving spatial inverse problems in the Earth Sciences remains a considerable challenge in particular when uncertainty quantification in the form of multiple Earth models is required. In a Bayesian framework, multiple models can be obtained by sampling a posterior distribution formulated as the product of a spatial (geostatistical) prior and a likelihood function depending on data and model errors. Many efforts have been done in recent years to propose efficient sampling techniques often based on Markov-Chain Monte Carlo [e.g., *Fu and Gomez-Hernandez*, 2009; *Mariethoz et al.*, 2010; *Hansen et al.*, 2012; *Vrugt et al.*, 2013; *Lochbühler et al.*, 2015]. However, most of these techniques become computationally prohibitive if the forward problem takes hours of computing time for one single model evaluation, such as is often the case when inverting dynamic flow and transport data.

In addition, when uncertainty is important, the proposed solutions may be strongly dependent on the formulation of the prior distribution of models. In case geostatistical algorithms are used to model complex 3D heterogeneity on large grids, such prior is rarely available analytically or of closed form or parametric expressions. Moreover, due to the nature of geological interpretation and the nature of classification of geological systems, the prior uncertainty is often hierarchical. Based on well and geophysical data, hydrogeologists speculate on the nature of the depositional system and often form scenario-type hypothesis. In reservoir geology, a scenario can be seen as alternative understanding of subsurface heterogeneity leading to alternative parameter definitions for subsurface modeling [*Martinius and Naess*, 2005]. Within each scenario, one may then define within-scenario spatial uncertainty, usually generated through geostatistical algorithms. Most methodologies are focused on inverse modeling within a single limited scenario (e.g., a multi-Gaussian with variogram parameters or a single Boolean model definition) and ignore the discrete uncertainty related to the scenario itself.

Scenario uncertainty in hydrogeological inverse problems has been extensively studied in the past decades and is generally investigated using Bayesian model averaging (BMA) [e.g., *Ye et al.*, 2004; *Li and Tsai*, 2009] or generalized likelihood uncertainty estimation (GLUE) [*Beven and Binley*, 1992, 2014 and reference therein] or a combination of both [e.g., *Rojas et al.*, 2008]. The basic idea of GLUE is to run many scenarios that reproduce equally well observed data and to compute on that basis a likelihood estimation. Monte Carlo simulations are performed through the different scenarios and a generalized likelihood measure is calculated for every proposed model according to its performance to reproduce observations. These likelihood estimations are normalized and use to build a cumulative density function expressing the uncertainty for some predictions of the models. Models with a likelihood below a threshold are generally rejected. The procedure requires a large and often practically prohibitive amount of simulations including those of dynamical data to reject some scenarios.

The BMA uses a more common Bayesian framework [*Hoeting et al.*, 1999]. To estimate the joint uncertainty, BMA combines the uncertainty within a scenario with the uncertainty regarding the scenario itself. Both uncertainty types are estimated through sampling the posterior distribution with Monte Carlo simulations. Given the high computational demand, many authors limit uncertainty analysis to the maximum likelihood BMA [*Neuman*, 2003]. Through the procedure, scenarios with low posterior probability may be rejected. As in GLUE, many simulations are required to identify inconsistent scenarios. For an overview of uncertainty analysis in hydrogeology, the readers are referred to *Refsgaard et al.* [2012].

In this paper, we propose a workflow for assessing uncertainty of the hydrogeological model that includes prior to inversion, a process of probabilistic falsification of scenarios. A Popper-Bayes philosophy proposed for the Earth Sciences by Tarantola [2006], states that data should first be used to falsify models, then only be considered for matching. The aim of this process is to maintain realistic uncertainty by first stating a very wide prior (step 1 below), then narrowing that prior by falsification (step 2 below). The proposed process requires matching data (step 3 below) after falsification, and thus reduces significantly the computational cost when compared to methods such as GLUE or BMA. In practice we proposed a strategy/workflow based on three steps:

1) Construction of a geologically informed spatial prior through the definition of alternative geological scenarios quantified as multiple-training images. Within scenario, variability (spatial uncertainty) is modeled using multiple-point geostatistics (MPS).
2) Validation of the prior with geophysical data (electrical resistivity tomography - ERT) and estimation of an updated probability assigned to each training image (with possibly some training images being rejected/falsified)
3) Matching dynamical data considering scenarios probabilities using a stochastic search method termed probability perturbation.

In the first step, we generate alternative geological scenarios from literature data as well as some site specific data to propose various plausible facies architectures scenarios. Facies or hydrofacies-based approaches are common in hydrogeology [e.g., *Fogg et al.,* 1998; *dell'Arciprete et al.*, 2012; *Zhang et al.*, 2013] and are generally used to reproduce complex geological architectures such as multimodal distributions which are difficult to reproduce with Gaussian distributions [*McKenna and Poeter*, 1995]. Multiple-point statistics (MPS) [*Strebelle*, 2002; *Caers and Zhang*, 2004; *Mariethoz and Caers*, 2015] was chosen for its ability to easily condition to data and for its ability to reproduce curvilinear and interconnected structures [*Hu and Chugunova*, 2008; *dell'Arciprete et al.*, 2012] often encountered in aquifers. MPS has already been successfully applied in groundwater studies [e.g., *Feyen and Caers*, 2006; *Ronayne et al.*, 2008; *Huysmans and Dassargues*, 2009, 2011]. The various scenarios will be quantified through a discrete set of training images. We generate, within scenario, variation (spatial uncertainty) by stochastic simulations with each training images using the SNESIM algorithm [*Strebelle*, 2000, 2002]. The method is dependent on the choice of the training image and hence its uncertainty should be considered [*Feyen and Caers*, 2006; *Park et al.*, 2013; *Scheidt et al.*, in press; *Khodabakhsi and Jafarpour*, 2013].

Ideally, Bayesian inverse models require a prior that is data-agnostic. However, this may also entail that the prior space is very large and possibly that some part of this prior is simply inconsistent with data. Therefore, in the second step, we validate prior geological scenarios (training images) using geophysical data. Geophysical methods may provide spatially distributed information on subsurface petrophysical properties and may thus be used to validate the architecture of prior scenarios. More specifically, potential-based methods such as electromagnetic or DC resistivity methods are commonly used to characterize aquifers [e.g., *Robert et al.*, 2011; *Hermans et al.*, 2012; *Doetsch et al.*, 2012]. However, geophysical techniques provide indirect information on smaller scale geological heterogeneity represented by training images. We transform prior scenarios into resistivity distribution scenarios through forward and regularization-based geophysical inverse modeling to validate them with field ERT coming from the study site. The comparison is made through distance calculation and projection into a low dimensional space to calculate the probability of each scenario given field ERT data [*Park et al.*, 2013; *Hermans et al.*, 2014]. This consistency step between prior scenarios and secondary data also ensures that geophysics can be used to constrain the stochastic simulations as soft data in the third step of our strategy. The performance of this falsification procedure is first assessed using test cases where the reference model is known.

The third step is most common in inverse modeling. We constrain the updated prior uncertainty with dynamic data, namely hydraulic heads, and geophysical data. The integration of dynamic data such as hydraulic heads or tracer breakthrough curves is not straightforward in geostatistical methods (see Zhou et al. [2014] for a review). The relationship between the simulated parameter and the dynamical data is complex and requires to solve a non-linear spatial problem including flow (and possibly transport) equations. Several methods are available to solve such problems, under some prior spatial constraints (e.g., variograms or training images) such as the pilot-point method [e.g., *de Marsily et al.*, 1984], the gradual deformation method [e.g., *Roggero and Hu*, 1998] or Markov chain Monte Carlo simulations [e.g., *Irving and Singha*, 2010]. Among them, the Probability Perturbation Method (PPM) [*Caers*, 2003] is a Bayesian stochastic search technique well-suited to integrate dynamical data in the MPS framework and successfully applied in several real-field cases [e.g., *Hoffman et al.*, 2006; *Caers et al.*, 2006; *Ronayne et al.*, 2008; *Park et al.*, 2013]. In the case of discrete variables [*Caers and Hoffman*, 2006], PPM corresponds to a stochastic search for MPS realizations that match the dynamic data. PPM is applied within each considered scenario to search for MPS realization matching hydraulic heads.

In the next section, we provide an overview of the technical components of the entire workflow. Next, the performance of the falsification/updating procedure is assessed using synthetic cases. Then, the proposed workflow is illustrated using a field example located in the alluvial aquifer of the Meuse River in Hermalle-sous-Argenteau, Belgium.

2. Technical Details of the Workflow

2.1. Modeling the posterior distribution with scenarios

We consider the inverse problem in hydrogeology of matching hydraulic heads data $\mathbf{D_{head}}$ given some uncertain prior spatial constraints. The aim of the proposed workflow is to model the posterior distribution considering jointly the uncertainty in the facies model $\mathbf{M}$ and in the geological scenario Sc. In this process, we consider the use of geophysical tomographic data (electrical resistivity tomography) $\mathbf{D_{ERT}}$.

The problem is decomposed in two parts: the first part is to assess the probability of the geological scenarios given geophysical data $P(Sc_i|\mathbf{D_{ERT}})$. This is used to determine how many realizations of each scenario should be used to build the posterior distribution. The second part is related to the pre-posterior uncertainty for any given scenario $P(\mathbf{M}|Sc_i,\mathbf{D_{ERT}},\mathbf{D_{head}})$. The latter is calculated using PPM with MPS simulations constrained with geophysical data.

Then, we combine those two terms to derive the posterior distribution considering the uncertainty in geological scenarios

$$P(\mathbf{M}, Sc \mid \mathbf{D_{ERT}}, \mathbf{D_{head}}) = \sum_{i=1}^{N} P(\mathbf{M} \mid Sc_i, \mathbf{D_{ERT}}, \mathbf{D_{head}}) P(Sc_i \mid \mathbf{D_{ERT}}) \quad \text{(equation 1)}$$

where N is the number of geological scenarios. Equation 1 corresponds to a weighted sum of individual pre-posterior distributions. This equation is similar to the BMA approach, except that the term $P(Sc_i|\mathbf{D_{ERT}})$ is calculated before inverse modeling.

In this workflow, we do not validate geological scenarios with hydraulic heads; we assume that $\mathbf{D_{ERT}}$, given its spatial distribution, is more informative about the scenario variable *Sc* than $\mathbf{D_{head}}$, because of the spatial nature of geophysical data. Note that $\mathbf{D_{ERT}}$ is used twice in the workflow: 1) to validate globally the geological scenarios 2) to constrain locally MPS simulations.

2.2. Construction of a spatial prior with multiple alternative geological scenarios

The construction of the prior with alternative geological scenarios is based on the generation of several training images representing uncertainty related to interpretation of geological heterogeneity. Hydrogeologists may postulate several scenarios, constructed from conceptual understanding and based on analog databases containing information of geometric shapes, spatial positioning and other important elements of subsurface heterogeneity [see *Eschard et* *al.*, 2002; *Kiessling and Flügel*, 2002; *Gibling*, 2006; *Kenter and Harris*, 2006; *Jung and Aigner*, 2012; *Pyrcz et al.*, 2008; *Colombera et al.*, 2012]. In the following, we will refer to a specific geological scenario as Sc_i with $i = 1, 2,\ldots, N$. We will use Boolean simulation [*Maharaja*, 2008] to generate a training image for each scenario.

To generate realizations for a given scenario, we use multiple-point geostatistics [see *Hu and Chugunova*, 2008]. The (possibly infinite) set of realizations drawn from multiple training images then constitutes our prior. In particular, we use the SNESIM algorithm [*Strebelle*, 2000, 2002] to generate realizations for a given training image. The SNESIM algorithm relies on storing frequencies into a search tree, thereby alleviating the calculation of conditional probabilities in sequential simulation. The method easily allows constraining to any facies information (drilling) from wells. In addition, should soft data in the form of facies probabilities derived from geophysical data be available, then such information models can easily be constrained to such information [e.g., *Trainor*, 2010; *Castro et al.*, 2007; *Strebelle et al.*; 2002]. However, within our strategy, such constraining is only done at the very end, after falsification of scenarios.

2.3. Electrical resistivity tomography

Electrical resistivity tomography (ERT) data $\mathbf{D}_{\mathbf{ERT}}$ are used twice in the process. First, they are used to validate globally the geological scenarios and update the prior (section 2.4). Second, they are included as soft data to constrain MPS simulations in the sampling process (section 2.5).

The electrical resistivity distribution is obtained after inversion of electrical resistance data collected on the field site. A least-square regularization procedure [*Tikhonov and Arsenin,* 1977] is used for the deterministic inversion of resistance data. For the field study, we used the model parameter covariance matrix as regularization operator and a reference model in inversion to improve the inversion results compared to the traditional smoothness constrained inversion [see *Hermans et al.*, 2012; *Caterina et al.*, 2014]. This ensures that our ERT inversions are more informative and provide better estimates of the true resistivity distribution.

For constraining MPS simulation, the electrical resistivity distribution $\mathbf{D_{ERT}}$ is transformed into conditional facies probability maps $P(\mathbf{M}|\mathbf{D_{ERT}})$. The latter is computed using the comparison of co-located values for both geophysical parameter and facies. This probabilistic approach avoids the definition of a petrophysical relationship linking the geophysical parameters and the facies or hydrogeological parameter. Several studies have shown the limitation of using such a direct link in tomographic methods to derive hydrogeological parameters due to the regularization and spatially variable resolution inherent to those methods [e.g., *Day-Lewis et al.*, 2005]. Synthetic simulations relationships were proposed to overcome those limitations [*Moysey et al.*, 2005; *Singha and Moysey*, 2006]. To avoid regularization, one has to consider coupled inversion schemes where the hydrogeological parameters are transformed to geophysical parameters, using a petrophysical relationship, to check that geophysical observations are fitted [e.g., *Hinnel et al.*, 2010; *Irving and Singha*, 2010]. Recently, it has been proposed that geophysical imaging could be improved through physically-based regularization using synthetic simulations and principal component analysis (PCA) [*Oware et al.*, 2013]. Although very promising, these techniques have been mostly demonstrated in synthetic test cases or relatively simple field cases where the processes and conceptual models are well known.

In the traditional soft data approach, each value of resistivity will correspond to a certain probability of observing the different facies. This is a conservative approach because it does not impose facies or parameter values. The derived facies probability maps integrate uncertainties related to ERT inversion, including those linked with the regularization operator.

One limitation is that the loss of resolution with depth for surface arrays is globalized. Taking into account resolution loss more accurately would require sufficient borehole data to estimate the resistivity distribution of the different facies according to depth, resolution or sensitivity.

2.4. Falsification and updating of scenario probability

The initial set of training images, defined from analog information may be incompatible with actual subsurface data, such as dynamic or geophysical data, and the initial (often equiprobable) training image probabilities need to be updated once subsurface information is considered. *Park et al.* [2013] proposes a Bayesian method for updating the initial probabilities with subsurface data (in their case dynamic flow data in an oil reservoir) and to reject training images deemed incompatible with flow data. *Scheidt et al.* [in press] used the same method to falsify scenario of turbidite reservoir using a new drilled well. The method was adapted by *Hermans et al.* [2014] to deal with geophysical data and is shortly reviewed here. The idea is to compute the probability of observing a specific training image Sc_i given some observed geophysical data $\mathbf{D_{ERT}}$: $P(Sc=Sc_i|\mathbf{D_{ERT}}) = P(Sc_i|\mathbf{D_{ERT}})$ in a lower dimensional space. The falsification procedure can be summarized in the following steps:

1. Consider N scenarios with equal probability $P(Sc_i)=1/N$. A set of unconditional geostatistical realizations are constructed for each one.
2. From field knowledge and analogs, a value of geophysical parameter is assigned to each facies. This is the responsibility of the geophysicist to choose a coherent value; otherwise, the method may be misleading.
3. The forward geophysical response is calculated.
4. Simulated and field geophysical data sets are inverted using the same inversion parameters (e.g. section 2.3. and reference therein) to generate simulated and field inverted geophysical models.

5. The pair-wise Euclidean distance between any two simulated inverted models and between any simulated inverted model and field inverted model is calculated and stored in a distance matrix D.
6. The simulated and field inverted models are projected in a lower *d*-dimensional space, using multidimensional scaling (MDS) [*Borg and Groenen*, 2006; *Caers*, 2011]. Multi-dimensional scaling approximates the above Euclidean distance with a lower dimensional Euclidean distance in Cartesian space using the eigen-value decomposition of D. MDS therefore reduces the dimension of the data variable $\mathbf{D_{ERT}}$ to a new *d*-dimensional variable $\mathbf{D^{*}_{ERT}}$ of much lower dimension. The actual observed field data $\mathbf{D^{*}_{ERT,obs}}$ can also be mapped into this lower dimension. If it falls outside the distribution of simulated models, this indicated that none of the training image is consistent with the data. Because a Cartesian space is now constructed and mapped, density estimation can proceed directly in that lower dimensional space.
7. Adaptive kernel smoothing [see *Park et al.*, 2013] is applied in the *d*-dimension space to estimate the probability density of the data variable for each training image $f(\mathbf{D^{*}_{ERT}}|Sc_i)$. This allows calculating the probability $P(Sc_i|\mathbf{D_{ERT}})$ using Bayes' rule:

$$P\left(Sc_i|\mathbf{D}_{\mathrm{ERT}}\right)=P\left(Sc_i|\mathbf{D}_{\mathbf{ERT,obs}}\right)\simeq P(Sc_i\mid\mathbf{D}^{*}_{\mathbf{ERT}}=\mathbf{D}^{*}_{\mathbf{ERT,obs}})=\frac{f(\mathbf{D}^{*}_{\mathbf{ERT}}=\mathbf{D}^{*}_{\mathbf{ERT,obs}}\mid Sc_i)P(Sc_i)}{\sum_{\mathrm{i=1}}^{\mathrm{N}}f(\mathbf{D}^{*}_{\mathbf{ERT}}\mid Sc_i)P(Sc_i)}$$

(equation 2)

The scenarios for which this probability is very low are falsified by the data.

The main idea of this method is to reduce dimension based on a distance defined between multiple geophysical inversions and the actual field data. Then, we calculate scenario probability. Note that at no point does the method call for matching models (**M**) to data $\mathbf{D_{ERT,obs}}$. At this point, some scenarios Sc_i with a very low probability can be rejected due to their inconsistency with available subsurface data. Note that this step does not require the simulation of dynamical data, which leads to a significant gain of computing time.

For some geophysical methods, step 4 can be avoided and the distance calculation can be made directly on the geophysical data. However, for ERT, the voltages or apparent resistivities are highly dependent on the resistivity of the very-shallow subsurface. Consequently, two identical models differing only by the "first" layer resistivity would have a large distance in the apparent resistivity even if their true resistivity distribution is relatively close.

2.5. Sampling with the Probability Perturbation Method

We now focus on the pre-posterior term $P(\mathbf{M}|Sc_i,\mathbf{D}_{\mathbf{ERT}},\mathbf{D}_{\mathbf{head}})$ in equation 1. At this stage one could opt for sampling methods [e.g., *Fu and Gomez-Hernandez*, 2009; *Mariethoz et al.*, 2010; *Hansen et al.*, 2012; *Vrugt et al.*, 2013] but given the subjective nature of the prior, it is our opinion that accurately sampling from a posterior distribution which itself relies on considerable (subjective) geological prior interpretation is not desirable. In addition, sampling requires the evaluation of 1000s of forward model runs which is impossible when the forward models takes hours of computing time.

Instead, we opt for a stochastic search method termed probability perturbation method (PPM) [*Caers*, 2003]. The aim is not for a rigorous sampling but for a broad search of the prior space for realizations that match the hydraulic head data. In short PPM, much like gradual deformation [*Caers*, 2007], allows for perturbation of an initial model **M** into a new model **M'**, without destroying the prior geological scenario. In other words, the perturbation is a sample of the prior. What is ignored in PPM is the transition probability associated with this perturbation, hence the less rigorous sampling.

At this stage, samples of the prior are generated with MPS sequential simulations with ERT probability maps used as soft data.

PPM is an iterative process which stops when the objective function ϕ reaches the targeted level ε:

$$\phi = \sqrt{\frac{\sum_{k=1}^{K}(h_k^{obs} - h_k^{calc})^2}{K}} \leq \varepsilon \quad \text{(equation 3)}$$

where K is the number of observation points, h_k^{obs} is the kth observed hydraulic head and h_k^{calc} is the kth calculated hydraulic head. We performed groundwater flow modeling with HydroGeoSphere [*Therrien et al.*, 2010].

3. Synthetic study

In this section, we propose 4 synthetic experiments (case 1 to case 4 below) to assess the falsification/updating procedure in controlled set-ups. In contrasts with *Park et al.* [2014] and *Scheidt et al.* [in press] who validated their procedure with the rejection sampler, we propose here to analyze the performance considering a large number of reference truths (true Earth models). The aim is to analyze the sensitivity of the method to identify the training image belonging to the reference truth as well evaluate the updated probabilities as calculated from our method.

Within this synthetic study, we will consider 8 different training images representing alluvial deposits (Figure 1). They are all based on a background facies made of sand and a combination of gravel channels and/or clay lobes. Three sizes of channels (small (SC), medium (MC) and big (BC)) and two sizes of lobes (small (SL) and big (BL)) are considered. For example, the scenario with small channels of gravels and small clay lobes will be identified as SC/SL. If not specified, the proportions of gravel and clay facies in the training image are respectively 20% and 22%. The facies were assigned a value of logarithm of resistivity (in Ohm.m) equal to 1.95 for the clay facies, 2.2 for the sand facies and 2.65 for the gravel facies for calculating their ERT response.

The set-up of the synthetic case mimics the field case (see sections 4 and 5): a 10 m thick alluvial aquifer with cells 0.5m thick and 1m wide. ERT data are simulated using profiles of 64 electrodes with 2m spacing (126 m length) and a dipole-dipole configuration. Noise was added on the resistance data to a level similar to the one encountered on the field (0.25%) before inversion.

100 different models are considered for each training image/geological scenario. For each experiment, all the models are subsequently used as a reference truth model and the updated probabilities are computed. We assess the ability of the method by computing a Bayesian confusion matrix. This matrix states how many models of Sc_i are classified as Sc_j. An identity matrix would correspond to a perfect classification. Similarly, we compute the mean of the updated probabilities over models from the same TI to assess the performance of the method.

3.1. Case 1: Prior containing distinctive geological scenarios

For this case, we consider 4 different training images: one with three facies and three with two facies: SC/SL, BC, BL1 (proportion of 30%) and BL2 (proportion of 50%). The 4 considered scenarios are clearly different in terms of facies geometry and resistivity distribution; hence we use this case to test how well the procedure can identify the true scenario. The corresponding 2D MDS map (Figure 2A) shows that the different scenarios are clearly identified in as few as two dimensions (representing more than 95 % of the variance). Table 1 summarizes the Bayesian confusion matrix and the mean updated probability.

The confusion matrix for this case illustrates that more than 90% of the models are correctly classified using the updating procedure, with a mean probability over 75%. It also illustrates that scenarios with very differencing facies geometries (channels vs. lobes) can be falsified using the procedure. This shows how the falsification with geophysical data is able to indicate which scenarios should be rejected.

3.2. Case 2: Prior containing geological scenarios with similarities

For this case, we consider 4 different training images with three facies and similar proportions: SC/SL, SC/BL, MC/SL, MC/BL. The difference between these 4 scenarios lies only in the geometry of the facies. Because of these similarities, this represents a more challenging case for the falsification procedure. The corresponding 2D MDS plot (Figure 2B, the crosses are used later in Case 4) represents only 49% of the total variance. Considerable overlap between scenarios with the MDS plot can now be observed. Table 2 shows the confusion matrix and the probabilities for 5 dimensions (73 % of the variance).

In this case, the methodology, on average identifies correctly the training image used to generate the reference model. However, given the similarities among the training images, the misclassification is more abundant. The mean probabilities are around 60% when calculated in 5D. When the scenarios are more alike, to allow for a good discrimination, it is necessary to consider higher dimensions for calculating updated probabilities. In this specific case, calculating in 10 dimensions (90% of the variance) allows to discriminate the scenarios as well as in Case 1.

Geophysical data is not always able to identify the correct training image because of spatial uncertainty. Due to the particular arrangement of geological bodies in space, one scenario may look like another based on the limited resolution geophysical data. This justifies the idea of considering several scenarios since retaining a single scenario may yield too small uncertainty in the modeling and hence later in forecasting.

3.3. Case 3: Prior containing geological scenarios similar to the reference TI

For this case, we consider the seven training images of Cases 1 and 2 and an additional training image with small channels (SC). We only consider as reference scenario the scenario SC/SL. The idea is to test the behavior of the methodology when the reference scenario is not included in the prior, but consistent ones are. 2D MDS map (Figure 2C) shows that some scenarios are clearly falsified while others seem consistent. Table 3 summarizes the classification performance and mean probability when the reference TI (SC/SL) is included, or not, for five dimensions (94% of the variance). For the latter case, models from SC/SL were taken out of the prior for MDS map and updated probability calculation.

The methodology correctly identifies the reference training image if it is included in the prior. When this is not the case, the highest probabilities are assigned to those training images sharing at least one element in common with the reference training image: SC, MC/SL and to a lesser extent SC/BL. The methodology is thus able to identify geological scenarios consistent with the reference truth model. In both cases, the falsification procedure rejects inconsistent training images.

3.4. Case 4 : Prior containing geological scenarios distinct from the reference TI

The last case considers the four training images from Case 2 as part of the prior but uses models from the training image BL2 as reference models. We test here what happens when the geological scenarios of the prior are all inconsistent with the reference truth model. Figure 2B shows an example of the resulting 2D MDS map including 2 reference truth models (crosses). One of the models lies outside the distribution of the prior, this is an indication that the prior is not consistent with geophysical data. For the second model, the inconsistency only appears in 3D (Figure 2D). It is now up to the modeler to decide whether such training images should be excluded. In such a case, calculating updated probabilities is worthless. A new prior should be drawn with consistent geological scenarios.

4. Field site

Field data used in this study are from an experimental site of University of Liege located in the alluvial aquifer of the Meuse River, in Hermalle-sous-Argenteau (Belgium) near the Dutch-Belgian boarder (Figure 3A and B), between the Meuse River and the Albert Canal.

4.1. Building prior geological scenarios

According to geological and hydrogeological investigations [*Haddouchi*, 1987; *Rentier*, 2003; *Battle-Aguilar*, 2007], the deposits of the Meuse River are mostly representative of braided systems but structures characteristics of meandering systems are also possible. Deposition is mostly composed of sandy gravel. Heterogeneity in the deposits is characterized by zones of clean gravel (and pebble) having a higher hydraulic conductivity and zones composed of loam, clay and clayey gravel of lower hydraulic conductivity. The latter are remaining of old and eroded floodplain deposits, crevasse splays or old channels filled with fine sediments.

A facies description is available for 23 boreholes on the site (Figure 3C). Alluvial deposits are 10 m thick and lie on a bedrock composed of Visean and Houiller shales and schists. The boreholes were drilled down to the bedrock.

Globally, the deposits are divided in three main units (layers). The first unit is 0.5 to 5 m thick and is composed of fluvial loams. The second unit is composed of sandy gravel and the third unit is mainly made of clean gravel with large decimetric pebbles (Figure 3D). According to borehole logs, one of the two last units may not exist and their thickness varies with their location. However, previous studies made on the site with solute tracer tests [*Brouyère*, 2001] and heat tracing experiments [*Hermans et al.*, 2015; *Wildemeersch et al.*, 2014] have shown that heterogeneity exists inside these predefined units and that a simple model with three horizontal layer is not sufficient to catch heterogeneity realistically. Therefore, we used training image-based scenarios to model the prior.

Three facies are defined: a clay/loam facies corresponding to low hydraulic conductivity deposits, a sand/sandy gravel facies having an intermediate hydraulic conductivity and a gravel facies with high hydraulic conductivity. The analysis of borehole shows that the proportions of these facies in Hermalle-sous-Argenteau are respectively 18, 40 and 42%.

From geological descriptions of the alluvial deposits of the Meuse River [*Haddouchi*, 1987; *Rentier*, 2003; *Battle-Aguilar*, 2007] in the area of Liege (Belgium) and from interpretation of the borehole data, several training images are proposed. Considering that the gravel facies, with its higher hydraulic conductivity, has the most influence on groundwater flow, we propose two types of training images: one with the gravel facies represented as long continuous channels, the second one with gravel as shorter, but elongated bars. For each type, we considered two different sizes for the gravel elements, leading to a total of 4 different training images (Figure 4). The clay/loam facies is represented by lobes of various sizes.

4.2. Geophysical data

The geophysical data set is composed of 12 ERT parallel profiles (Figure 3C). The profiles are 126 m long (except for the northern profile which is shorter) and are separated in the perpendicular direction by 4 m. They were collected with 64 electrodes (2 m spacing between electrodes) using a dipole-dipole configuration (dipole size $a < 9$ and dipole separation $n < 7$). The noise level was estimated using reciprocal measurements and a linear error model was used to weight data during inversion [*Slater et al.*, 2000], which should avoid the creation of artifacts in the inverted sections.

The profiles were inverted as explained in section 2.3. The model parameter covariance matrix was computed based on a spherical variogram with a vertical range of 4.4m (determined using electromagnetic logs performed in the boreholes) and an anisotropy ratio of 2.5 [see *Hermans et al.*, 2012]. The reference model was divided in two horizontal zones. The first zone represents alluvial deposits (0 to 10 m depth) and has a resistivity value of 160 Ohm.m. The second zone corresponds to the bedrock and has a resistivity value of 300 Ohm.m. The two zones are disconnected during inversion, i.e. values of parameters lying in different zones are not correlated.

Figure 5A shows one typical profile collected on the site. In the north-eastern part of the profile, a low electrical resistivity zone corresponds to thick, clayey and loamy deposits. Below, the deposits are characterized by two layers of different resistivity. The first one is composed of sand, the second one is made of gravel. Lateral heterogeneity is visible in both layers showing that the division of the deposits in homogeneous layers is not satisfying. Nevertheless, it is expected that the gravel facies is preferentially located at the bottom part of the deposits.

4.3. Relationship between electrical resistivity and facies

By comparing electrical resistivity and facies at the position of borehole (Figure 5B) a histogram of resistivity for each of the three facies was constructed. Generally, a higher resistivity is observed for gravel facies due to the absence, or smaller amount, of fine sediments having a relatively high surface conductivity [*Bersezio et al.*, 2007; *Doetsch et al.*, 2010]. In this case, the sand facies globally has a higher resistivity (240 Ohm.m) than the gravel facies (140 Ohm.m). The reason lies in the nature of the gravel facies, which is composed of large pebbles and has much higher water content than the sand facies whereas the amount of fine sediments with non-negligible surface conductivity is low for both of them. The clay facies is the only one characterized by resistivity values below 90 Ohm.m. However, due to the limited resolution of ERT, the clay facies also displays resistivity values in the same range that the gravel facies.

The histograms (Figure 5B) are then use to compute the facies conditional probability as function of electrical resistivity (Figure 5C). To constrain MPS simulations, resistivity distributions in the subsurface (Figure 5A) are transformed in facies probability maps that can be used as soft data.

4.4. Hydrogeological data

The hydrogeological data set consists in drawdowns measured in 9 of the boreholes (Old piezometers in Figure 3C) screened on the whole thickness of the aquifer after reaching steady-state conditions during a pumping test. The boundaries of the hydrogeological model are drawn in Figure 3C. The model is 167 m x 93 m x 10 m. MPS simulations were drawn on a grid size of 1 m x 1 m X 0.5 m (310620 cells). The boundary conditions are imposed hydraulic heads extracted from a regional flow model [*Brouyère*, 2001]. A recharge of 300 mm/year is considered. The hydraulic conductivity of each facies is chosen according to our prior knowledge of the site and a sensitivity analysis; they remain constant during PPM. The hydraulic conductivity of the gravel facies is the most sensitive parameter. We imposed a value of 5.10^{-2} m/s for the gravel facies, 10^{-4} m/s for the sand facies and 10^{-6} m/s for the clay facies.

5. Application of the workflow to the field case

5.1.Updating of training-image scenarios

The four training image scenarios proposed for the field site (Figure 4) have the same prior probability of 0.25. To compute the MDS plot and perform kernel density estimation, we simulated 24 sections for each training image, leading to a total of 96 simulated facies models from the prior. The number of simulated models is a compromise between the time to produce the MDS map and its representativeness. The number of models must be sufficient to estimate the density distribution $f(\mathbf{D_{ERT}}|Sc_i)$ in the chosen dimension. The larger the amount of models used, the more precise the estimate is.

According to field observations (Figure 5B and C), a value of resistivity was assigned to each facies: 100 Ohm.m for the clay facies, 140 Ohm.m for the gravel facies and 240 Ohm.m for the sand facies. Then, for each simulated resistivity model, electrical resistances were simulated. Noise was added to the data according to the level measured on the field with reciprocal measurements. Simulated data sets were inverted with the same procedure as field data sets leading to 96 geophysical models.

Based on the distance matrix calculated with those 96 geophysical models, an MDS plot is drawn in 2 dimensions (Figure 6). In this plot, the four scenarios are characterized by a color code. Field models are represented by black squares. In this 2D projection, the field cases fall in the distribution of training images-based scenario cases, hence none of training images can be visually falsified with the ERT data. We observe the effect of varying the training image: models from Sc_4 occupy the bottom part of the plot, whereas from Sc_2 are concentrated in the right part. Sc_1 and Sc_3 have a higher density in the middle of the plot.

The analysis of the eigen-values spectrum of the distance matrix obtained with MDS enables to select the dimensions in which kernel density estimation will be performed. The higher the dimensions d, the higher the considered variance is and the closest the distance in the d-dimensions is to the real distance. The 2D projection represents slightly more than half the total variance. More than 85% of the total variance is reached at the fifth dimension. Then the contributions of eigenvalues decrease significantly. Note that we do not aim to reach 100% of the variance, mainly because not only the geological scenario influences the results (what we want to quantify) but also the methodology itself (undesirable effect): the choice of the distance metrics, the noise on the data, the geophysical parameter values and the number of simulated cases.

The probabilities of scenarios were calculated from kernel density estimation. We aggregated the contribution of individual profiles (each black square) to come up with a single probability for each scenario. The results are summarized in Table 4. According to ERT data, the lowest probability is assigned to Sc_2. However, its probability of 14% is not sufficiently low to falsify this geological scenario. The three other scenarios have quite similar probabilities with 29, 27 and 30% for Sc_1, Sc_3 and Sc_4 respectively. Using the 2D map would have led to an overestimation of the probability of Sc_2 and Sc_4 and an underestimation of Sc_1.

Prior uncertainty is represented by generating multiple realizations from each scenario, but taking into account the updated probabilities of Table 4. Figure 7 (top) shows the probability of the gravel facies considering those updated probabilities. It consists in 100 independent realizations. The number of realizations per scenario is given by $P(Sc_i \mid \mathbf{D}_{\mathbf{ERT}}) \times N_M$ with N_M equals the number of desired realizations, 100 in this case. This updated prior takes into account the uncertainty related to training image based scenarios and the facies observed in boreholes on the field which are considered here as certain.

This updated prior is then constrained with soft data from ERT, $\mathbf{D}_{\mathbf{ERT}}$. This constitutes the updated/constrained prior or pre-posterior distribution $P(\mathbf{M}|Sc,\mathbf{D}_{\mathbf{ERT}})$. Models going into PPM for matching hydraulic heads are sampled from this distribution. As can be seen from Figure 7 (bottom), adding spatially distributed information, such as geophysical data, reduces prior uncertainty where boreholes are not available. Where no specific data is available (hard or soft), the probability of gravel is close to the expected proportions (42%).

5.2. Matching hydrogeological data

In matching the dynamic head data with PPM, the targeted level of the objective function was set to 0.015. This level is easily reached in a few iterations through the PPM process for all the scenarios (Figure 8).

Models drawn from the updated prior distribution and matching data belongs to the posterior distribution. Individual realizations have different geometrical characteristics depending on the training image used for simulation (Figure 9). For Sc_1 and Sc_2, there are continuous channel-shaped gravel bodies crossing the model. Sc_2 and Sc_4 have larger gravel zones.

The pre-posterior distribution $P(\mathbf{M}|Sc_i,\mathbf{D_{ERT}},\mathbf{D_{head}})$ can be calculated by averaging the facies indicators (0 or 1) over multiple realizations that match the hydraulic head data (see Figure 10). These 3D probability cubes differ for each training image-based scenario. This shows that the training image uncertainty is important and influences strongly the results. In this case, we observe how scenarios with big elements (Sc_2 and Sc_4) lead to wider and more continuous zones where the probability of gravel is high.

5.3. Computing the posterior distribution

The posterior distribution is computed using equation 1 (Figure 11) considering a number of realizations coherent with the value of $P(Sc_i|\mathbf{D_{ERT}})$ from Table 4. Based on the posterior probability distribution, a classification model is proposed just for visualization. The most probable facies is assigned to each cell (Figure 11). This result confirms that the proportion of gravel tends to increase with depth whereas sand is more abundant in the upper part of the deposits, but that lateral heterogeneity exists within the deposits. Except near the surface, where a large clay zone is observed, clay only appears as small anomalies in the proposed classification model. This is also an effect of the lower proportion of clay compared to the two other facies.

6. Discussion and Conclusion

We propose a workflow in three steps to solve the inverse problem of matching hydraulic heads and model the posterior distribution considering jointly the uncertainty in the facies model and in the geological scenarios:

1) Construction of a geologically informed spatial prior with multiple scenarios
2) Validation/falsification of the prior with geophysical data
3) Matching dynamical data considering scenarios probabilities and geophysical data using a stochastic search method

The originality of the method lies in the use of geophysical data both to validate/falsify geological scenarios and to constrain geostatistical realizations and to perform this in a manageable computational time for practical field cases. The method is sensitive to the geophysical parameters used to produce simulated models in the falsification procedure. They should be chosen carefully to avoid eliminating consistent scenarios. We have successfully assessed the validity of the method on 4 synthetic test cases and demonstrated its applicability on a case study in an alluvial aquifer using MPS and PPM. However, one of the workflow's strength is its adaptability. The workflow is not limited to training image-based scenarios, step 1 can be based on variogram-based scenarios or any other geostatistical methods.

The falsification procedure causes an additional computational cost, but it is relatively small compared to the time required by the hydrogeological inverse procedure. The additional cost is related to the forward and inverse modeling of synthetic geophysical data. As an example, the falsification procedure for the field case required 96 synthetic models. It was run on a 3.07 GHz computer with 8 GB RAM in 8.5 hours without parallelization. The MDS procedure only takes a few minutes. As a comparison, the time to calibrate one model with PPM on the same computer, given the size of the 3D model, takes about 3 hours. The cost increases when more synthetic models are used and is highly dependent on the type of data used for falsification. It can be easily estimated knowing the time needed for one synthetic model.

Similarly, the application of MDS mapping for the calculation of probability in step 2 is not limited to the use of geophysical data. One may combine both geophysical and informative hydrogeological data sets, to assess the probability of scenarios $P(Sc|\mathbf{D}_{\mathbf{ERT}},\mathbf{D}_{\mathbf{hydro}})$ or even use hydrogeological data alone. In this study, we used geophysical data only because there are more informative on the spatial distribution of facies than head data. The fast calibration of models from any scenario suggests that head data are not very useful to discriminate the considered scenarios in this case. Transient data would likely be more informative. Actually, any method relevant to assess the pertinence of geological scenarios could be applied. Combining different data types would require the definition of a distance balancing between several terms and should result in some refinement in the falsification/updating procedure to identify more precisely the most probable scenarios. However, it would be at the cost of more computations, hence be more time consuming. Other geophysical techniques, such as the spectral induced polarization method could also be used, given its high potential for discriminating facies based on their permeability. Step 2 strongly depends on the definition of the distance metrics, the latter could be adapted depending on the context.

The application of the falsification procedure on several test cases highlights that the method is suitable for the rejection of inconsistent training images because the reference or field model lies outside the distribution of the prior. In most cases, the methodology correctly identifies the reference training image (highest updated probability). However, misclassification due to spatial uncertainty may occur resulting in an incorrect training image being favored. This illustrates that the geophysical data is not always sufficient to reveal geological scenarios and that a large prior (several scenarios) is needed to avoid underestimating uncertainty. The number of simulated models should always be sufficient to estimate the density distribution in the MDS map. Otherwise, updated probability could be badly estimated.

In the last step of the workflow, any sampling technique can be considered as long as it can integrate the probability of geological scenarios. It means that McMC or any other techniques can be used to sample the pre-posterior distributions and then combine the results using equation 1.

In many hydrogeological problems, the model relies on relatively sparse data generally limited to observed borehole facies. Potential-based methods such as electromagnetic or DC resistivity methods are commonly used to characterize aquifers but, they are rarely considered directly in the modeling step. While the MPS framework proposes an established way to integrate soft data, very few hydrogeological studies have considered the use of geophysics as spatially distributed conditioning data in MPS simulations [*Trainor*, 2010]. Our work shows that the use of ERT as soft data is twofold. First, constraining MPS simulations with geophysical data reduces the number of possible models that are proposed to the PPM algorithm. It reduces the variability in the prior and subsequently reduces the variability in the posterior distribution, too. Second, the convergence of the PPM is increased. This is a non-negligible advantage regarding the high computational demand for solving non-linear flow and transport models.

Acknowledgement

The PPM code and a corresponding data set is available at https://github.com/SCRFpublic/PPM. Other material can be obtained on demand, please contact the corresponding author.

We thank the F.R.S.-FNRS and the Fondation Roi Baudouin — Prix Ernest Dubois (grant No. 2013-8126501-F002) for their financial support of Thomas Hermans. We are grateful to Alain

Dassargues and the Hydrogeology and Environmental Engineering unit of the University of Liege for giving us access to the experimental site of Hermalle-sous-Argenteau.

References

Battle-Aguilar, J. (2008) Groundwater flow and contaminant transport in an alluvial aquifer: in-situ investigations and modelling of a brownfield with strong groundwater-surface interactions, Ph. D. thesis, University of Liege, Belgium.

Bersezio, R., M. Giudici, and M. Mele (2007), Combining sedimentological and geophysical data for high resolution 3D mapping of fluvial architectural elements in the Quaternary Po plain (Italy), Sedimentary Geology, 202, 230-248.

Beven, K., and A. Binley (1992), The future of distributed models : model calibration and uncertainty prediction. Hydrogeological Processes, 6, 279-298.

Beven, K., and A. Binley (2014), GLUE: 20 years on, Hydrological Processes, 28, 5897–5918.

Borg, I., and P. J. F Groenen (2005), Modern Multidimensional Scaling Theory and Applications, New York, Springer.

Brouyère, S. (2001), Etude et modélisation du transport et du piégeage des solutés en milieu souterrain variablement saturé, Ph. D. thesis, University of Liege, Belgium, in French.

Caers, J. (2003), Geostatistical history matching under training-image based geological model constraints, SPE Journal, 8(3), 218-226.

Caers, J. (2007), Comparing the gradual deformation with the probability perturbation method for solving inverse problems. Mathematical geology, 39(1), 27-52.

Caers, J. (2011), Modeling uncertainty in Earth Sciences, Wiley-Blackwell, Chichester, UK.

Caers, J., and T. Hoffman (2006), The Probability Perturbation Method: A New Look at Bayesian Inverse Modeling, Mathematical Geology, 38(1), 81–100.

Caers, J. and T. Zhang (2004), Multiple-point geostatistics: a quantitative vehicle for integrating geological analogs into multiple reservoir models. In: Grammer, M., P.M. Harris, and G.P. Eberli (eds.), Integration of outcrop and modern analogs in reservoir modeling, AAPG Memoir, 80, 383-394.

Caers, J., T. Hoffman, S. Strebelle, and X.H. Wen (2006), Probabilistic integration of geologic scenarios, seismic and production data – A West Africa turbidite reservoir case study, The Leading Edge (Tulsa, OK), 25, 240-244.

Castro, S.A., J. Caers, and C. Otterlei (2009), Incorporating 4D seismic data into reservoir models while honoring production and geologic data: a case study. The Leading Edge, 28(12), 1499-1506.

Caterina, D., T. Hermans, and F. Nguyen (2014), Case studies of incorporation of prior information in electrical resistivity tomography: comparison of different approaches, Near Surface Geophysics, 12, 451-465, doi: 10.3997/1873-0604.2013070.

Colombera, L., F. Felletti, N.P. Mountney, and W.D. McCaffrey (2012), A database approach for constraining stochastic simulations of the sedimentary heterogeneity of fluvial reservoirs, AAPG bulletin, 96(11), 2143-2166.

Day-Lewis, F.D., K. Singha, and A.M. Binley (2005), Applying petrophysical models to radar travel time and electrical tomograms: resolution-dependent limitations, Water Resources research, 110, B08206. doi:10.1029/2004JB003569

dell' Arciprete, D., R. Bersezio, F. Felletti, M. Giudici, A. Comunian, and P. Renard (2012), Comparison of Three Geostatistical Methods for Hydrofacies Simulation: A Test on Alluvial Sediments, Hydrogeology Journal, 20, 299–311.

de Marsily, G., G. Lavedan, M. Boucher, and, G. Fasanino (1984), Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, In. Verly, G., M. David, A. Journel, A. Marechal (Eds.), Proceedings of Geostatistics for natural resources characterization, D. Reildel

Pub. Co., 831-849.

Doetsch, J., N. Linde, I. Coscia, S. A Greenhalgh, and A. G. Green (2010), Zonation for 3D Aquifer Characterization Based on Joint Inversions of Multimethod Crosshole Geophysical Data, Geophysics, 75(6), G53–G64, doi:10.1190/1.3496476.

Doetsch, J., N. Linde, M. Pessognelli, A.G. Green, and T. Günther (2012), Constraining 3-D Electrical Resistance Tomography with GPR Reflection Data for Improved Aquifer Characterization, Journal of Applied Geophysics, 78, 68–76, doi:10.1016/j.jappgeo.2011.04.008.

Eschard, R., B. Doligez, and H. Beucher (2002), Using quantitative outcrop database as a guide for geological reservoir modelling. In: Armstrong, M., C. Bettini, N. Champigny, A. Galli, and A. Remacre (eds.), Geostatistics Rio 2000, Doordrecht, Netherlands, Kluwer, 7-17.

Feyen, L., and J. Caers (2006), Quantifying Geological Uncertainty for Flow and Transport Modeling in Multi-Modal Heterogeneous Formations, Advances in Water Resources 29(6), 912–929.

Fogg, G. E., C.D. Noyes, and S.F. Carle (1998), Geologically based model of heterogeneous hydraulic conductivity in an alluvial setting, Hydrogeology Journal, 6, 131–143.

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

Gibling, M.R. (2006), Width and thickness of fluvial channel bodies and valley fills in the geological record: a literature compilation and classification, Journal of Sedimentary Research, 76, 731-770.

Haddouchi, B. (1987), Etude géologique et hydrogéologique de la plaine alluviale de la Meuse en Belgique, Ph.D. thesis, University of Liege, in French.

Hansen, T.M., K.S. Cordua, and K. Moosegaard (2012), Inverse problems with non trivial priors: efficient solution through Gibbs sampling, Computational Geosciences, 16, 593-611, doi:10.1007/s10596-011-9271-1.

Hermans, T., A. Vandenbohede, L. Lebbe, and F. Nguyen (2012), Imaging Artificial Salt Water Infiltration Using Electrical Resistivity Tomography Constrained by Geostatistical Data, Journal of Hydrology, 438–439, 168–180. doi:10.1016/j.jhydrol.2012.03.021.

Hermans, T., J. Caers, and F. Nguyen (2014), Assessing the probability of training image-based geological scenarios using geophysical data, In: Pardo-Iguzquiza, E., C. Carolina-Albert, J. Heredia, L. Moreno-Merino, J.J. Dura, and J.A. Vargas-Guzman. Mathematics of Planet Earth – Proceedings of the 15th Annual meeting of the IAMG, Madrid Spain, 679-682.

Hermans, T., S. Wildemeersch, P. Jamin, P. Orban, S. Brouyère, A. Dassargues, and F. Nguyen (2015), Quantitative temperature monitoring of a heat tracing experiment using cross-borehole ERT, Geothermics, 53, 14-26, doi: 10.1016/j.geothermics.2014.03.013

Hinnel, A.C., T.P.A. Ferre, J.A. Vrugt, J.A. Huisman, S. Moysey, J. Rings, and M. Kowalsky (2011), Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion, Water Resources Research, 46, W00D40.

Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky (1999), Bayesian model averaging: a tutorial, Statistical Science, 14(4), 382-417.

Hoffman, B.T., J.K. Caers, X.H. Wen, and S. Strebelle (2006), A practical data-integration approach to history-matching: application to a deepwater reservoir, SPE Journal, 11, 464-479.

Hu, L.Y., and T. Chugunova (2008), Multiple-Point Geostatistics for Modeling Subsurface

Heterogeneity: A Comprehensive Review, Water Resources Research, 44, W11413, doi:10.1029/2088WR006993.

Huysmans, M., and A. Dassargues (2009), Application of Multiple-Point Geostatistics on Modelling Groundwater Flow and Transport in a Cross-Bedded Aquifer (Belgium), Hydrogeology Journal, 17(8), 1901–1911, doi:10.1007/s10040-009-0495-2.

Huysmans, M., and A. Dassargues (2011), Direct Multiple-Point Geostatistical Simulation of Edge Properties for Modeling Thin Irregularly Shaped Surfaces, Mathematical Geosciences, 43, 521–536, doi:10.1007/s11004-011-9336-7.

Irving, J., and K. Singha (2010), Stochastic Inversion of Tracer Test and Electrical Geophysical Data to Estimate Hydraulic Conductivities, Water Resources Research, 46, W11514.

Journel, A.G. (2002), Combining Knowledge from Diverse Sources: An Alternative to Traditional Data Independence Hypotheses, Mathematical Geology, 34(5), 573–596.

Jung, A., and T. Aigner (2012), Carbonate geobodies: hierarchical classification and database – a new workflow for 3D reservoir modelling, Journal of Petroleum Geology, 35(1), 49-65.

Li, X., and F.T.-C. Tsai (2009),. Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod, Water Resources Research 45, W09403.

Kenter, J.A.M., and P.M. Harris (2005), Web-based outcrop digital analog database (WODAD): a public relational database archiving carbonate platform margins in the geological record. In: Geophysical research Abstract, European Geosciences Union.

Khodabakhshi, M., and B. Jafarpour (2013), A Bayesian Mixture-Modeling Approach for Flow-Conditioned Multiple-Point Statistical Facies Simulation from Uncertain Training Images, Water Resources Research, 49(1), 328-342,

doi:10.1029/2011WR010787.
Kiessling, W., and E. Flügel (2002), Paleo-reefs – a database on Phanerozoic reefs. In: Phanerozoic Reef Patterns, SEPM Special Publication, 77-92.
Lochbühler, T., J.A. Vrugt, M. Sadegh, and N. Linde (2015), Summary statistics from training images as prior information in probabilistic inversion, Geophysical Journal International, 201(1), 157-171.
Maharaja, A. (2008), TiGenerator: Object-based training image generator. Computer & Geosciences, 34, 1753-1761.
Mariethoz, G., and J. Caers (2015), Multiple-point geostatistics: stochastic modeling with training images, Wiley-Blackwell, 374 p.
Mariethoz, G., P. Renard, and J. Caers (2010), Bayesian inverse problem and optimization with iterative spatial resampling, Water Resources Research, 46, W11530, doi:10.1029/2008WR007621.
Martinius, A.W., and A. Naess (2005), Uncertainty analysis of fluvial outcrop data for stochastic reservoir modelling, Petroleum Geoscience, 11, 203-214.
McKenna, S.A., and E.P. Poeter (1995), Field example of data fusion in site characterization, Water Resources Research, 31(12), 3229-3240.
Moysey, S. (2005), A framework for inferring field-scale rock physics relationships through numerical simulation, Geophysical Research Letters, 32,L08304.
Neuman, S.P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, Stochastic Environmental Research and Risk Assessment, 17, 291-305.
Oware, E. K., S. Moysey, and T. Khan, (2013), Physically based regularization of hydrogeophysical inverse problems for improved imaging of process-driven systems, Water Resources Research, 49, 6238–6247 (2013).
Park, H., C. Scheidt, D. Fenwick, A. Boucher, and J. Caers (2013), History matching and

uncertainty quantification of facies models with multiple geological interpretations, Computer and Geosciences, 17, 609-621.

Pyrcz, M.J., J.B. Boisvert, and C.V. Deutsch (2008), A library of training images for fluvial and deepwater reservoirs and associated code, Computer and Geosciences, 34(5), 542-560.

Refsgaard, J.C., S. Christensen, T. Sonnenborg, D. Seifert, A.L. Hojberg, and L. Troldborg (2012), Review of strategies for handling geological uncertainty in groundwater flow and transport modeling, Advances in Water Resources, 36, 36-50.

Rentier, C. (2003), Méthode stochastique de délimitation des zones de protection autour des captages d'eau - Conditionnement par des mesures de conductivité hydraulique K, de hauteur piézométrique h et de résistivité ρ, Ph. D. Thesis, University of Liege, Belgium, In French.

Robert, T., A. Dassargues, S. Brouyère, O. Kaufmann, V. Hallet, and F. Nguyen (2011), Assessing the contribution of electrical resistivity tomography (ERT) and self-potential (SP) methods for a water well drilling program in fractured/karstified limestones, Journal of applied geophysics, 75, 42-53.

Roggero, F. and, L.Y. Hu (1998), Gradual deformation of continuous geostatistical models for history matching, In: Proceedings Society of Petroleum Engineers 49004, Annual technical Conference, New Orleans, L.A.

Rojas, R., L. Feyen, and A. Dassargues, (2008), Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging, Water Resources Research, 44, W12418.

Ronayne, M.J., S.M. Gorelick, and J. Caers (2008), Identifying Discrete Geologic Structures That Produce Anomalous Hydraulic Response: An Inverse Modeling Approach, Water Resources Research, 44, W08426, doi:10.1029/2007WR006635.

Scheidt, C., P., Tahmasebi, M. Pontiggia, A. Da Pra and J. Caers (2015), Updating joint uncertainty in trend and depositional scenario for reservoir exploration and early appraisal, Computational Geosciences, In press.

Singha, K., and S. Moysey (2006), Accounting for spatially variable resolution in electrical resistivity tomography through field-scale rock-physics relations, Geophysics, 71, A25–A28.

Slater, L., A. Binley, W. Daily, and R. Johnson (2000), Cross-hole electrical imaging of a controlled saline tracer injection, Journal of applied geophysics, 44, 85-102.

Strebelle, S. (2000), Sequential Simulation Drawing Structures from Training Images, Stanford University, 2000.

Strebelle, S. (2002), Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics. Mathematical Geology, 34(1), 1–21.

Tarantola, A. (2006), Popper, Bayes and the inverse problems. Nature Physics, 2, 492-494.

Therrien, R., R. McLaren, E. Sudicky, and S. Panday. (2010), HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Groundwater simulations group.

Tikhonov, A. and, V. Arsenin (1977), Solution of ill-posed problems, Winston and Sons, Washington, 1977.

Trainor, W. (2010), On the value of information for spatial problems in the earth Sciences, Stanford University, 2010.

Vrugt, J.A., C.J.F. ter Braak, C.G.H. Diks, and G. Schoups (2013), Hydrologic data assimilation using particle Markov chain Monte Carlo simulations: Theory, concepts and applications, Advances in Water Resources, 51, 457-478, doi:10.1016/j.advwatres.2012.04.002.

Wildemeersch, S., P. Jamin, P. Orban, T. Hermans, F. Nguyen, S. Brouyère, A. Dassargues

(2014), Coupling heat and chemical tracer experiments for the estimation of heat transfer parameters in shallow alluvial aquifers, Journal of contaminant hydrology, 169, 90-99, doi: 10.1016/j.jconhyd.2014.08.001

Ye, M., S.P. Neuman, and P.D. Meyer, (2004) Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, Water Resources Research, 40, W05113.

Zhou, H., J.J. Gómez-Hernández, and L. Li (2014). Inverse methods in hydrogeology: evolution and recent trends, Advances in Water Resources, 63, 22-37.

Zhang, Y., C.T. Green, and G.E. Fogg (2013), The impact of medium architecture of alluvial settings on non-Fickian transport, Advances in Water Resources, 54, 78–99.

TABLE

	Confusion matrix (%)				Mean updated probability (%)			
Scenario	**SC/SL**	**BL1**	**BL2**	**BC**	***P*(SC/SL|ref)**	***P*(BL1|ref)**	***P*(BL2|ref)**	***P*(BC|ref)**
SC/SL	100	0	0	0	88.3	1.8	0.8	9.1
BL1	3	96	1	0	3	75.1	21.7	0.2
BL2	3	15	82	0	1.7	21.6	76.6	0.1
BC	14	0	0	86	14.5	0.1	0	85.4

Table 1. Confusion matrix and updated probability for Case 1 in 2 dimensions.

	Confusion matrix (%)				Mean updated probability (%)			
Scenario	**SC/SL**	**SC/BL**	**MC/SL**	**MC/BL**	***P*(SC/SL|ref)**	***P*(SCLBL|ref)**	***P*(MC/SL|ref)**	***P*(MC/BL|ref)**
SC/SL	93	4	2	1	62.9	18.2	15.4	3.5
SC/BL	11	88	1	0	18.3	68.2	10.6	2.9
MC/SL	29	10	51	10	24.2	12	42.7	21.1
MC/BL	8	6	19	67	9.9	6.1	30	54

Table 2. Confusion matrix and updated probability for Case 2 in 5 dimensions.

Scenario	**SC/SL**	**SC/BL**	**MC/SL**	**MC/BL**	**BL1**	**BL2**	**BC**	**SC**
Classification								
with SC/SL	70	2	4	1	0	0	0	23
without SC/SL		4	32	1	0	0	0	63
Mean updated probability								
with SC/SL	46.2	7.7	14.7	7	0.8	0.4	2.2	21
without SC/SL		17	26.7	8.3	2	1.4	4	40.6

Table 3. Classification performance and updated probability for Case 3 in 5 dimensions.

Scenario	**Gravel geometry**	$P(Sc_i)$	$P(Sc_i \mid \mathbf{D_{ERT}})$ – 2D	$P(Sc_i \mid \mathbf{D_{ERT}})$ – 5D
Sc_1	Small channels	0.25	0.2	0.29
Sc_2	Medium channels	0.25	0.18	0.14
Sc_3	Small bars	0.25	0.27	0.27
Sc_4	Big bars	0.25	0.35	0.30

Table 4: Scenario and updated probability

FIGURE CAPTIONS

Figure 1: The eight training images used for the synthetic study are composed of gravel channels and clay lobes of different sizes, SC stands for small channels, MC for medium channels, BC for big channels, SL for small lobes, BL for big lobes. BL1 and BL2 have respectively 30 and 50 % of clay instead of 22%.

Figure 2: 2D MDS maps for the synthetic cases 1 (A), 2 (B), 3 (C) and 3D MDS map for the case 4 (D). Each color represents a specific prior scenario (TI). The percentages in parenthesis quantify the part of the variance represented by the dimension. When the scenarios are relatively different (A and C), the colors are well differentiated, showing that the methodology is able to falsify bad scenarios. When scenarios are more similar (B and D), more dimensions are needed to discriminate reliably the scenarios.

Figure 3: Location of the field site in the Meuse River alluvial aquifer (A and B), of the boreholes on the site together with the position of ERT profiles and model boundaries (C) and typical geological logs on the site (D).

Figure 4: The four training image scenarios considered at Hermalle-sous-Argenteau. Sc_1 = small channels, Sc_2= medium channels, Sc_3 = small bars, Sc_4 = big bars.

Figure 5: (A) Inversion results of one typical profile collected on the site. The global structures are perpendicular to the profiles. The histograms of resistivity (B) show that the sand facies is the most resistive and that low resistivity values correspond to clay. The histograms are used to compute the conditional probability (C) of facies given the resistivity.

Figure 6: On the 2D MDS map, the field models fall inside the distribution of simulated models, showing that all training images are visually consistent with geophysical data

Figure 7: The updated prior uncertainty was computed using MPS simulations with borehole data (hard data only) and the probabilities deduced from the MDS map (top). Then, we introduce ERT data as soft data to deduce the updated and constrained prior uncertainty (bottom). The latter is used for matching hydraulic heads with PPM.

Figure 8: The data misfit is similar for all the scenarios showing that on the hydrogeological point of view, all scenarios can explain the observed hydraulic heads.

Figure 9: Individual realizations display various geometrical characteristics corresponding to their training image-based geological scenario.

Figure 10: The pre-posterior distributions obtained for the four different training show variations due to the architecture of facies relative to each training image

Figure 11: Posterior distribution (top) and classification (bottom) of the facies considering the use of geophysical data and the probability of training image scenarios.

FIGURES

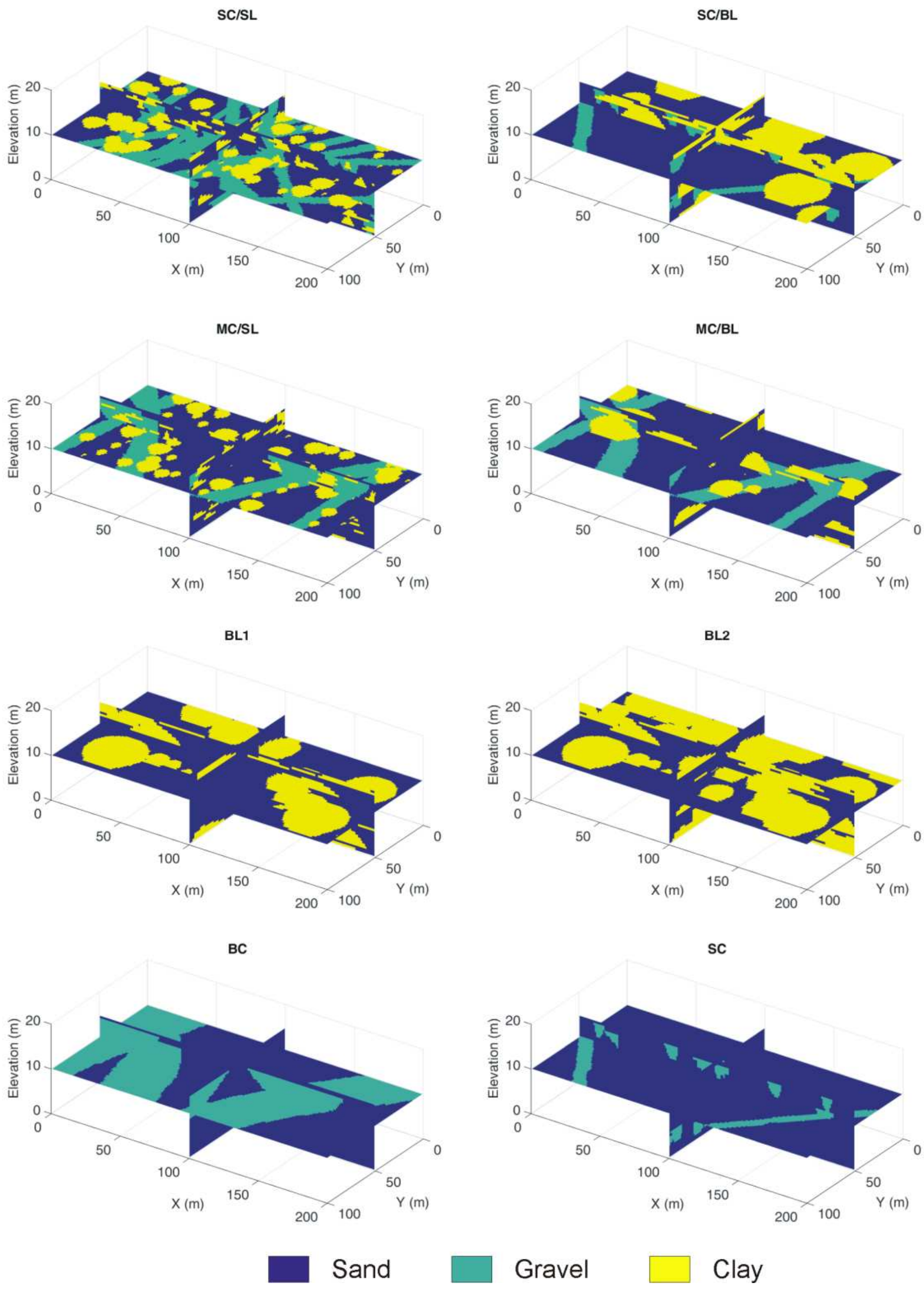

Figure 1: The eight training images used for the synthetic study are composed of gravel channels and clay lobes of different sizes, SC stands for small channels, MC for medium channels, BC for big channels, SL for small lobes, BL for big lobes. BL1 and BL2 have respectively 30 and 50 % of clay instead of 22%.

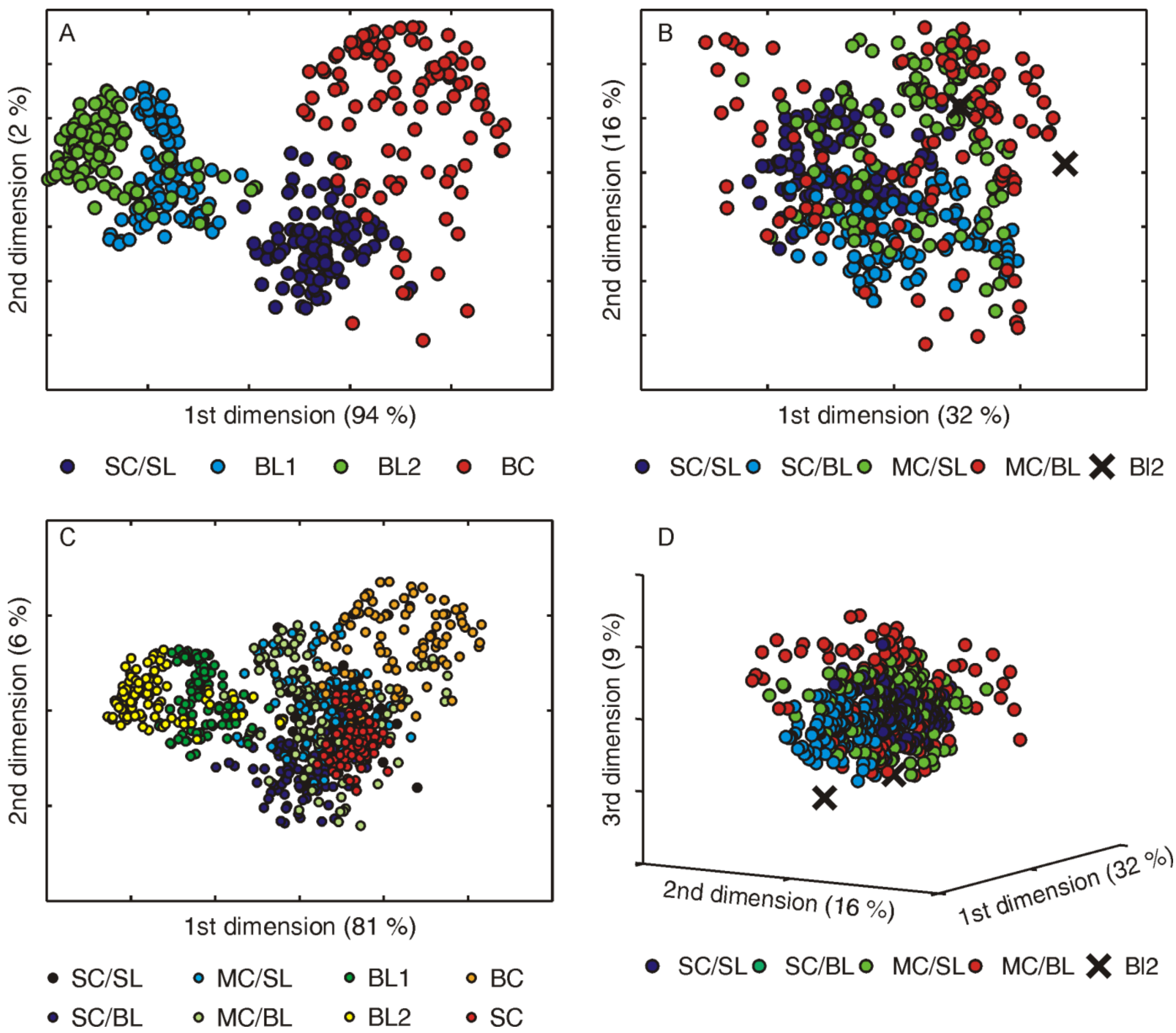

Figure 2: 2D MDS plots for the synthetic cases 1 (A), 2 (B), 3 (C) and 3D MDS plot for the case 4 (D). Each color represents a specific prior scenario (TI). The percentages in parenthesis quantify the part of the variance represented by the dimension. When the scenarios are relatively different (A and C), the colors are well differentiated, showing that the methodology is able to falsify bad scenarios. When scenarios are more similar (B and D), more dimensions are needed to discriminate reliably the scenarios.

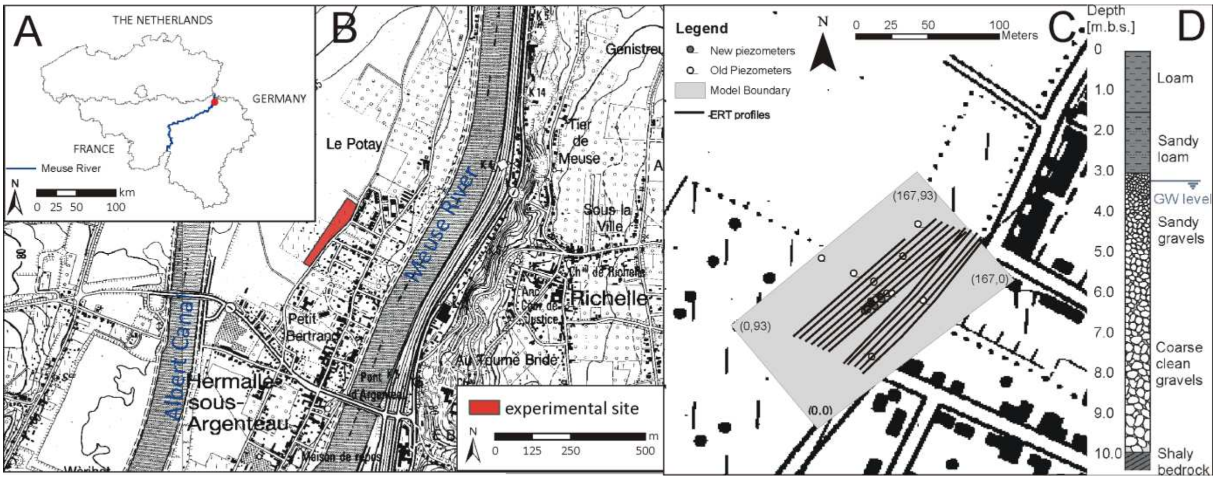

Figure 3: Location of the field site in the Meuse River alluvial aquifer (A and B), of the boreholes on the site together with the position of ERT profiles and model boundaries (C) and typical geological logs on the site (D).

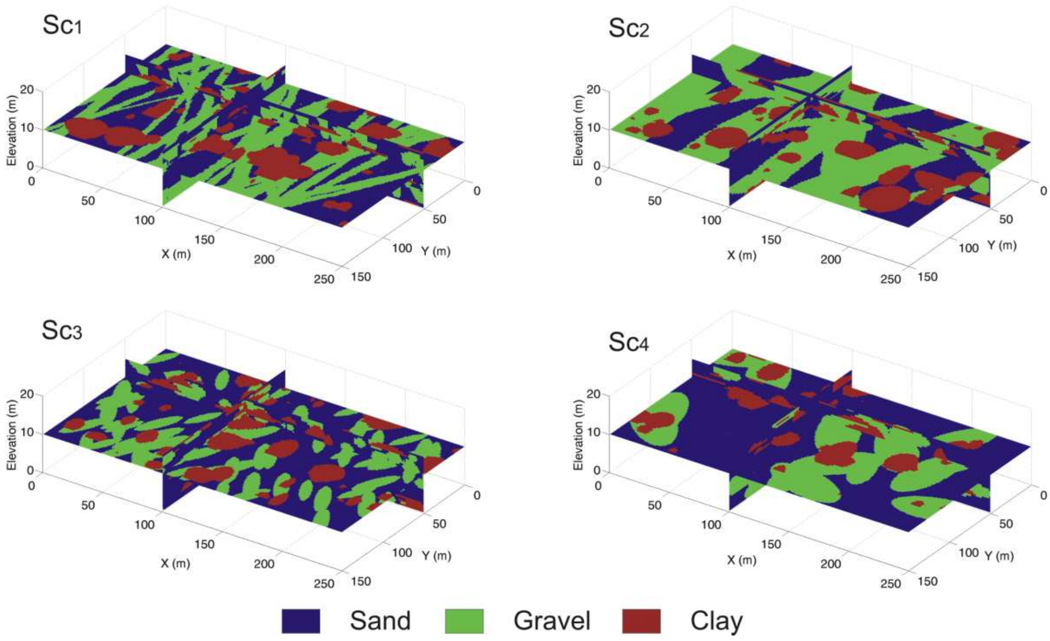

Figure 4: The four training image scenarios considered at Hermalle-sous-Argenteau. Sc_1 = small channels, Sc_2= medium channels, Sc_3 = small bars, Sc_4 = big bars.

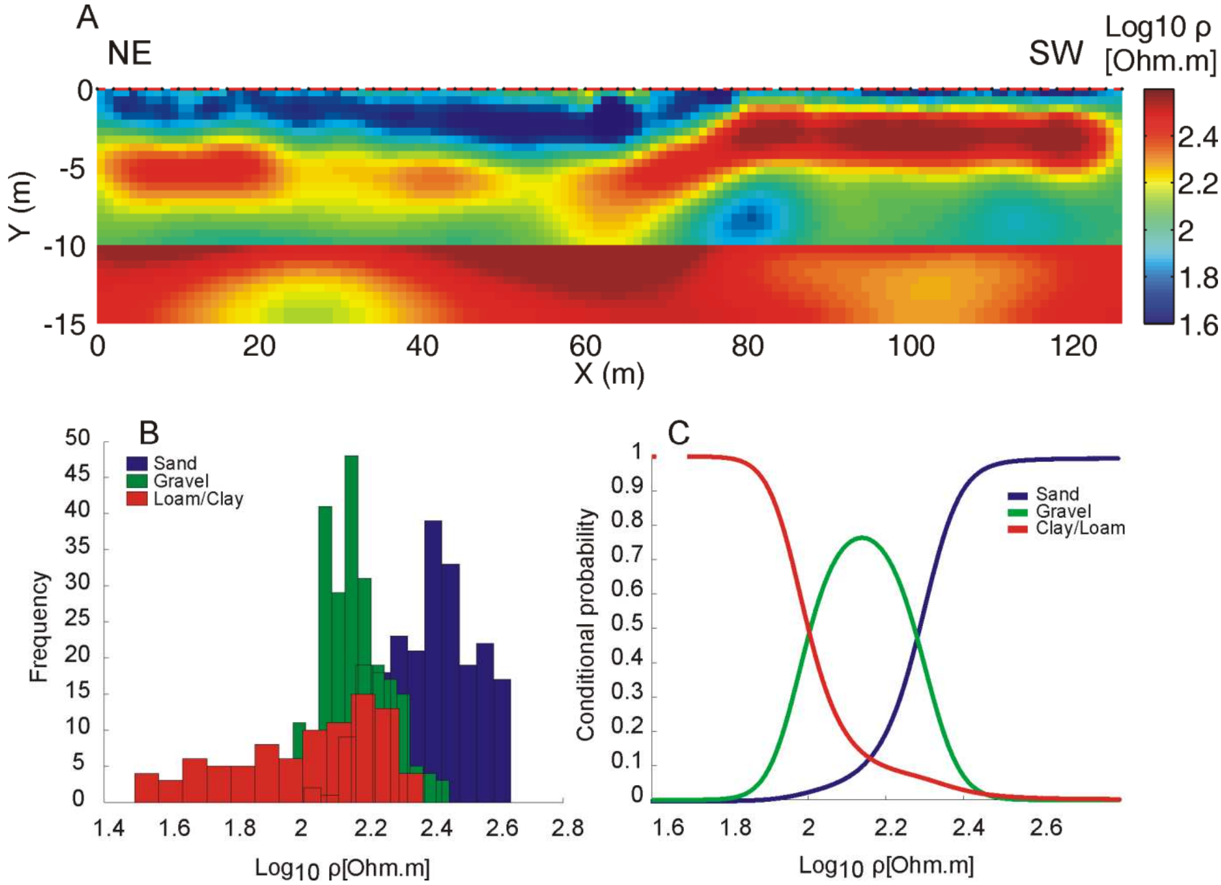

Figure 5: (A) Inversion results of one typical profile collected on the site. The global structures are perpendicular to the profiles. The histograms of resistivity (B) show that the sand facies is the most resistive and that low resistivity values correspond to clay. The histograms are used to compute the conditional probability (C) of facies given the resistivity.

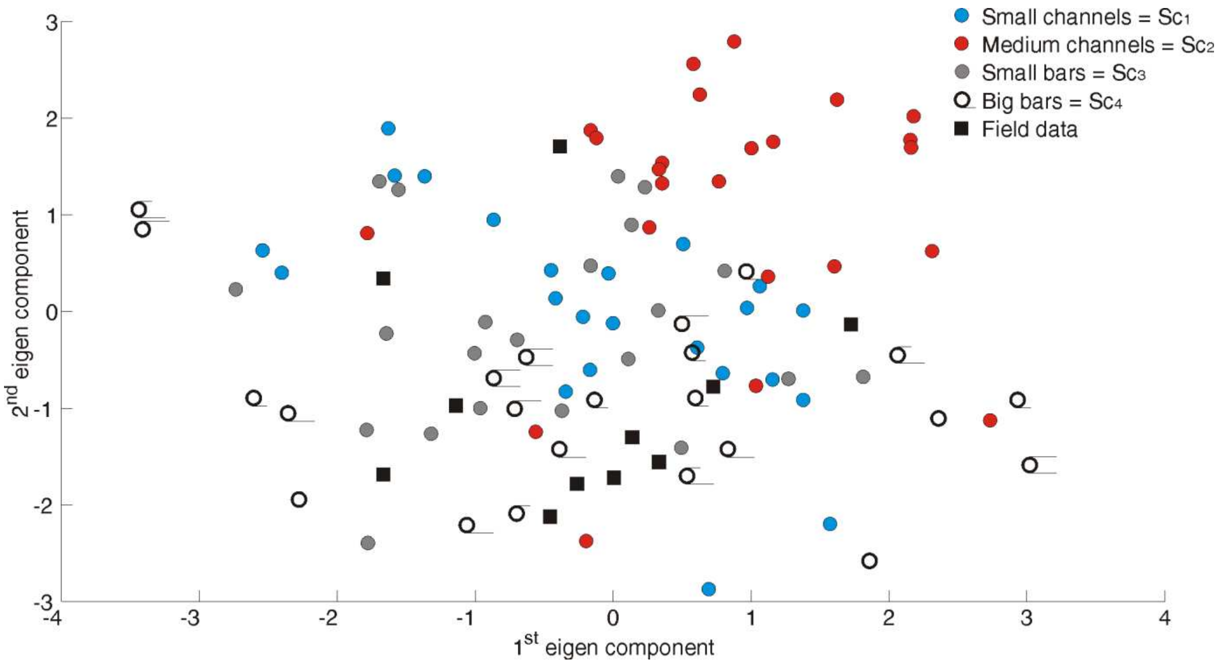

Figure 6: On the 2D MDS map, the field models fall inside the distribution of simulated models, showing that all training images are visually consistent with geophysical data.

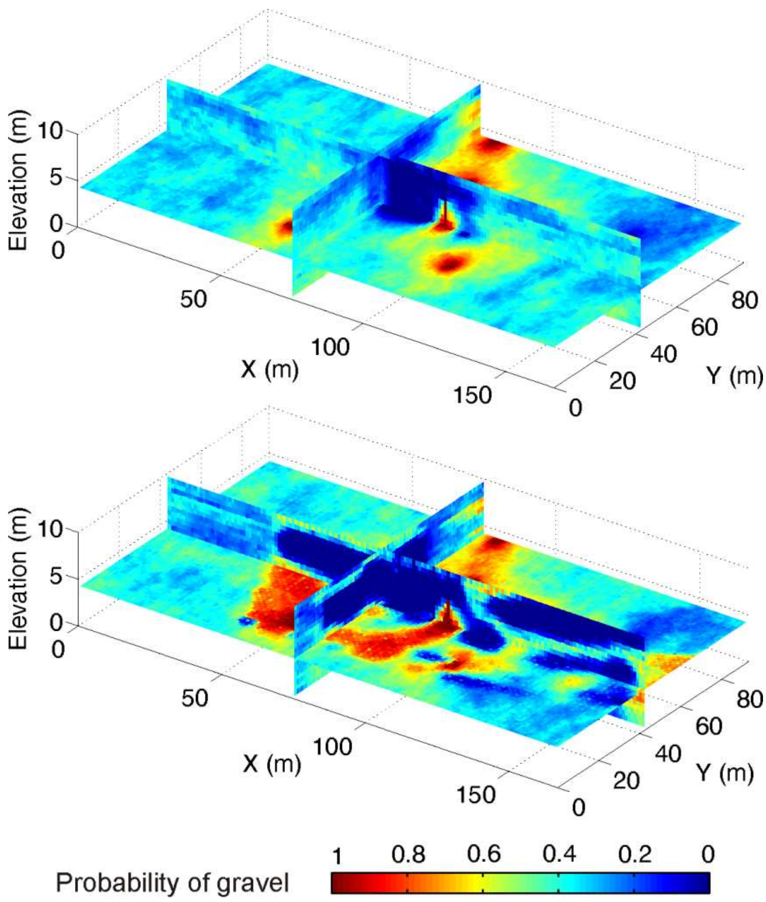

Figure 7: The updated prior uncertainty was computed using MPS simulations with borehole data (hard data only) and the probabilities deduced from the MDS map (top). Then, we introduce ERT data as soft data to deduce the updated and constrained prior uncertainty (bottom). The latter is used for matching hydraulic heads with PPM.

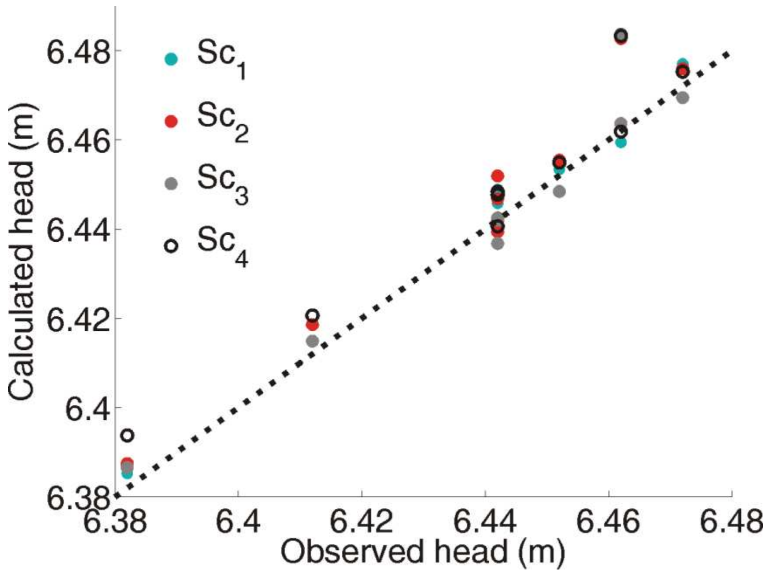

Figure 8: The data misfit is similar for all the scenarios showing that on the hydrogeological point of view, all scenarios can explain the observed hydraulic heads.

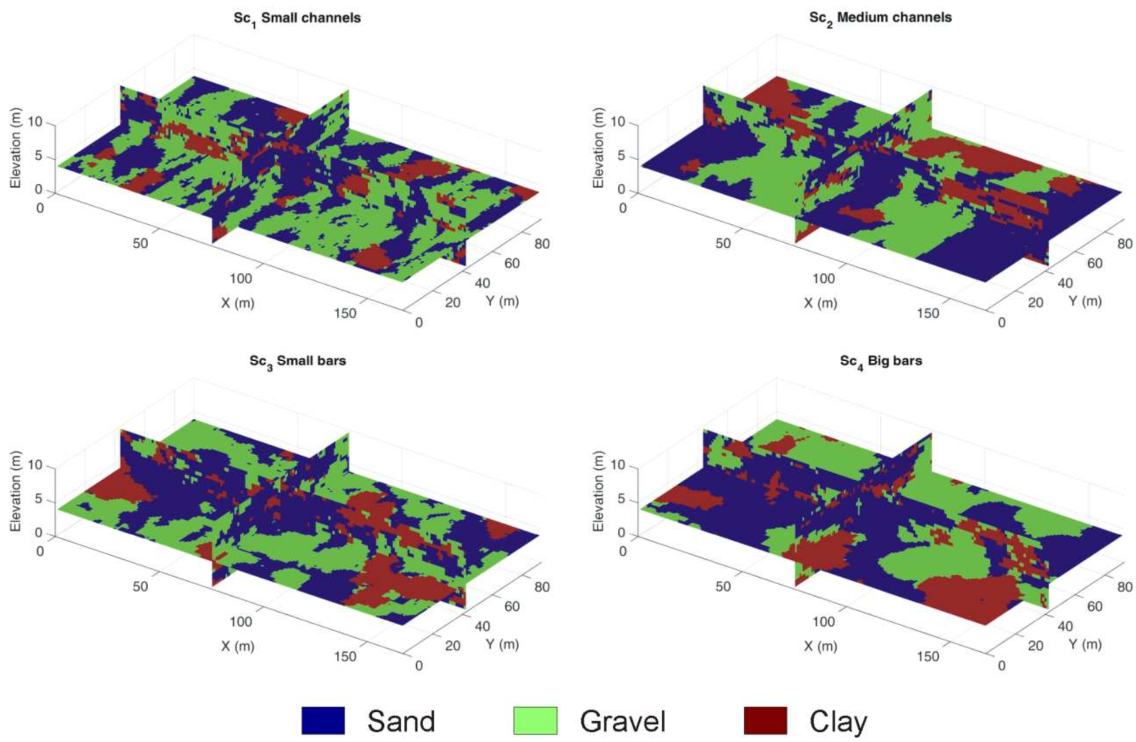

Figure 9: Individual realizations display various geometrical characteristics corresponding to their training image-based geological scenario.

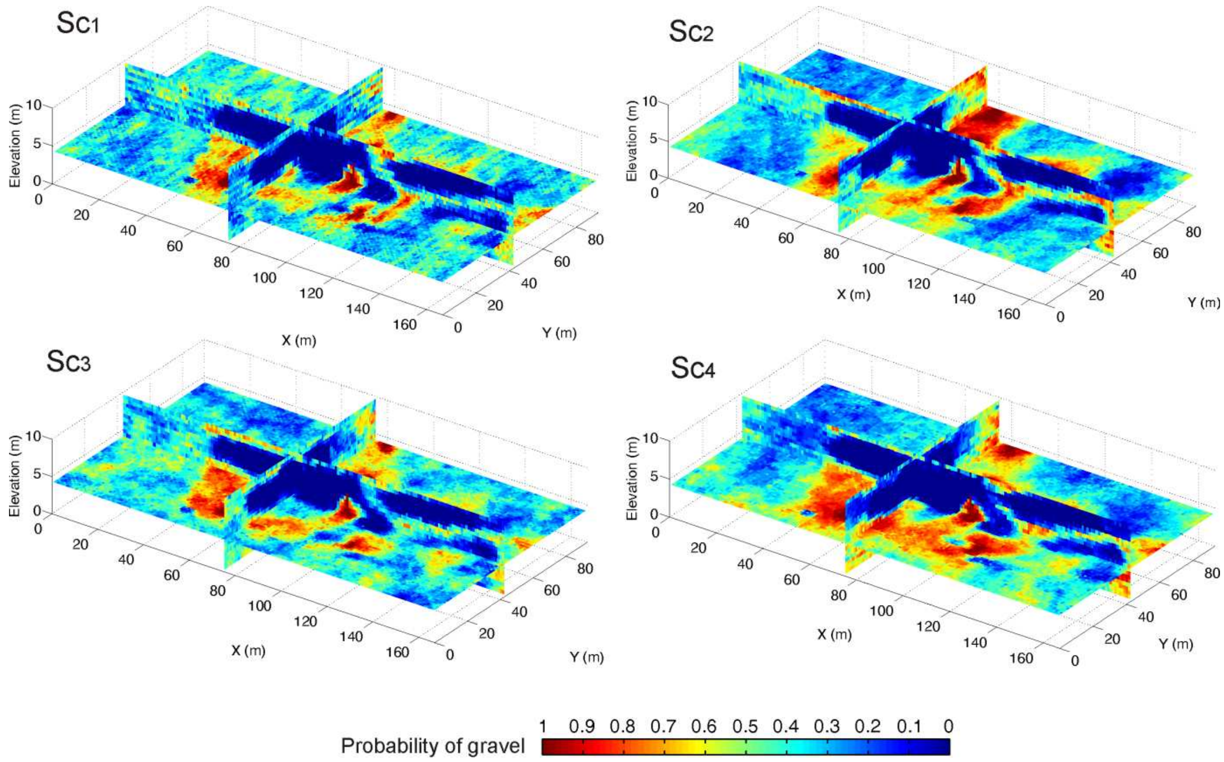

Figure 10: The pre-posterior distributions obtained for the four different training show variations due to the architecture of facies relative to each training image.

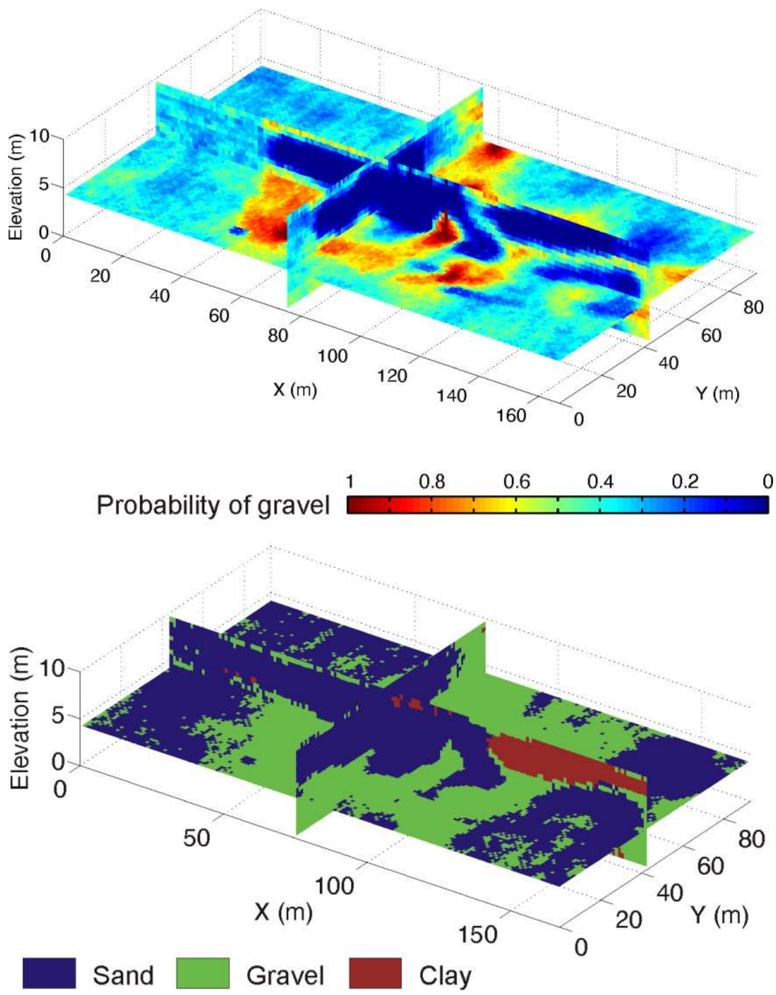

Figure 11: Posterior distribution (top) and classification (bottom) of the facies considering the use of geophysical data and the probability of training image scenarios.