

Probabilistic Inversion with Graph Cuts: Application to the Boise Hydrogeophysical Research Site

Guillaume Pirot¹, Niklas Linde¹, Grégoire Mariethoz², John Bradford³

¹Applied and Environmental Geophysics Group, Institute of Earth Sciences, University of Lausanne, Switzerland

²Institute of Earth Surface Dynamics, University of Lausanne, Switzerland

³Department of Geosciences, Boise State University, USA

Key Points:

- First field demonstration of probabilistic inversion based on graph cuts
- Adaptations are brought to the original algorithm to enhance geological realism
- Approach makes multiple-point statistics inversion feasible for field situations

Corresponding author: Guillaume Pirot, guillaume.pirot@unil.ch

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/2016WR019347

Abstract

Inversion methods that build on multiple-point statistics tools offer the possibility to obtain model realizations that are not only in agreement with field data, but also with conceptual geological models that are represented by training images. A recent inversion approach based on patch-based geostatistical resimulation using graph cuts outperforms state-of-the-art multiple point statistics methods when applied to synthetic inversion examples featuring continuous and discontinuous property fields. Applications of multiple-point statistics tools to field data are challenging due to inevitable discrepancies between actual subsurface structure and the assumptions made in deriving the training image. We introduce several amendments to the original graph cut inversion algorithm and present a first-ever field application by addressing porosity estimation at the Boise Hydrogeophysical Research Site, Boise, Idaho. We consider both a classical multi-Gaussian and an outcrop-based prior model (training image) that are in agreement with available porosity data. When conditioning to available crosshole ground-penetrating radar data using Markov chain Monte Carlo, we find that the posterior realizations honor overall both the characteristics of the prior models and the geophysical data. The porosity field is inverted jointly with the measurement error and the petrophysical parameters that link dielectric permittivity to porosity. Even though the multi-Gaussian prior model leads to posterior realizations with higher likelihoods, the outcrop-based prior model shows better convergence. In addition, it offers geologically more realistic posterior realizations and it better preserves the full porosity range of the prior.

1 Introduction

Since the advent of inverse modeling in hydrogeology [Kitanidis and Vomvoris, 1983; Carrera and Neuman, 1986], many geostatistical approaches have been developed and compared [Zimmerman *et al.*, 1998; Hendricks Franssen *et al.*, 2009]. Among them, the quasi-linear inversion method [Kitanidis, 1995], pilot points methodologies [Certes and de Marsily, 1991; RamaRao *et al.*, 1995], ensemble Kalman filters [Houtekamer and Mitchell, 1998; Hendricks Franssen and Kinzelbach, 2008] and the gradual deformation method [Hu, 2000], rely on the use of multi-Gaussian random fields. Such representations of geological properties are not always the most appropriate when considering flow and transport applications [Gómez-Hernández and Wen, 1998; Journel and Zhang, 2006]. Normal score transformation ensures Gaussianity [Zhou *et al.*, 2011], but not multi-Gaussianity. More

generally, multi-Gaussian fields involve latent assumptions of minimal connectivity of the extreme values, leading to systematic underestimations in contaminant transport simulations [Zinn and Harvey, 2003; Schlüter and Vogel, 2011]. When model parameters describe subsurface property fields with complex connectivity patterns, then multiple-point statistics [MPS, Guardiano and Srivastava, 1993] usually allows for a better characterization of structures than those based on first and second order moments only [e.g. histogram and semi-variogram; Journel and Zhang, 2006; Hu and Chugunova, 2008].

Alternative inversion approaches such as the probability perturbation method [Caers and Hoffman, 2006] or iterative spatial re-sampling [Alcolea and Renard, 2010; Mariethoz et al., 2010a] allow for considering more realistic geological conceptual models within the inversion framework. These methods are compatible with the use of training images (TI), which describe the MPS of a conceptual random field and allow a realistic representation of geological properties (e.g. connectivity, contrasts). Hermans et al. [2015] propose an MPS inversion approach applied to real data, following what they refer to as a Popper-Bayes top-down approach. It starts by reducing the prior conceptual model uncertainty by using soft data conditioning (e.g. summary statistics based on electrical resistivity tomography) and continues with a stochastic search, based on applying the probability perturbation method to invert for the hydrogeological properties while honoring hard data conditioning (e.g. porosity or hydraulic conductivity measured in boreholes). Yet, repeated geostatistical re-simulation make these types of algorithms very costly. In their review, Linde et al. [2015a] highlight that there are very few field-applications of MPS-based inversion that impose strong geological realism. Another possibility is to simulate MPS realizations conditionally to soft data [e.g., a tomogram derived from geophysical data; Lochbühler et al., 2013; Linde et al., 2015b; Straubhaar et al., 2016], but then there is no guarantee that the conditional realizations strictly honor the underlying data (e.g., the geophysical data used to construct the tomogram).

Since their conceptual introduction [Guardiano and Srivastava, 1993] and their first practical implementation [Strebelle, 2002], MPS techniques have known many developments [Renard and Mariethoz, 2014]. After first being limited to the simulation of categorical variables, they were extended to the simulation of continuous variables [Mariethoz et al., 2010a]. Nonstationarity in the TI or in the simulation grid can be overcome by using auxiliary variables [Chugunova and Hu, 2008; de Vries et al., 2009; Piro et al., 2014]. Working with transformed variables allows preserving the continuity of complex

structures such as meandering channels [Mariethoz *et al.*, 2014]. More recently, inspired from texture synthesis algorithm [Mariethoz and Lefebvre, 2014], MPS algorithms have developed from pixel-wise sequential simulations toward patch-based approaches [Rezaee *et al.*, 2013; Mahmud *et al.*, 2014], which greatly decrease the computing costs. An important aspect related to MPS algorithms, is that the prior model is not given by the TI, but by the ensemble of MPS realizations that can be produced by a combination of a specific MPS algorithm with a set of parameters and TIs.

Modern Markov chain Monte Carlo (MCMC) algorithms, such as the popular DREAM suite of algorithms [e.g., Vrugt *et al.*, 2008; Laloy and Vrugt, 2012; Vrugt, 2016] have opened the door to fully probabilistic hydrogeological inversion in comparatively high parameter dimensions (100's of model parameters). These algorithms are very powerful when the prior probability density function (pdf) is described by uncorrelated model parameters. By using appropriate model parameterizations (e.g., wavelets, discrete cosine transforms, or Fourier transforms), it is possible to define an uncorrelated prior parameter field in a transformed domain that yields spatially correlated hydrogeological property fields in the Cartesian domain [e.g., multi-Gaussian; Laloy *et al.*, 2015]. However, such algorithms rely fundamentally on a mathematically defined prior pdf. In practice, it is often impossible to define explicit prior pdfs that enable geologically realistic subsurface realizations. In such cases, it is necessary to choose between mathematically elegant prior pdfs that are a simplified and sometimes sketchy representation of the subsurface geology, or to define the prior implicitly by relying on TIs and MPS [e.g., Strebelle, 2002; Hu and Chugunova, 2008; Mariethoz and Caers, 2014]. Presently, the extended Metropolis algorithm [Mosegaard and Tarantola, 1995] is the only published approach to probabilistic inversion with TIs [e.g., Hansen *et al.*, 2012; Mariethoz *et al.*, 2010a; Ruggeri *et al.*, 2015]. This is not to say that this is the most efficient algorithm possible, but it represents the present state-of-the-art. Possible strategies to define and sample prior pdfs in hydrogeological and geophysical inversion were recently reviewed by Linde *et al.* [2015a].

Zahner *et al.* [2016] combined the computational advantages of patch-based MPS algorithms with the extended Metropolis algorithm to enable sampling the posterior pdf in the context of MPS. The MPS model proposal mechanism consists in optimally merging a part of the current model realization with a piece of the TI. Discrepancies occurring at the boundaries between the two complementary parts are minimized using the graph cut algorithm developed by Boykov *et al.* [2001]. The algorithm proposed by Zahner *et al.* [2016]

performs very well on synthetic test cases with a speedup of about 40 compared to pixel-based re-sampling alternatives. However, high performance on synthetic test cases does not necessary imply practical applicability for real cases. First, the TI describing the conceptual random field cannot match reality to perfection. Second, physical simulation (forward) models are simplifications of physical processes. The main challenges are thus 1) to propose a conceptual random field that is compatible with the data to match and with auxiliary field observations and 2) to parameterize the inverse problem such that the MCMC algorithm converges within a reasonable computing time.

In this paper, we propose a workflow that addresses the challenges associated with graph cut inversion for field applications and illustrate it with an application to the Boise Hydrogeophysical Research Site (BHRS), Idaho, USA. In the presented application, the porosity field, the petrophysical parameters and the standard deviation of the data measurement errors are inverted jointly. Section 2 recalls the graph cut algorithm by *Zahner et al.* [2016]. Section 3 presents the BHRS field data. It starts with a description of how geological properties are captured in the prior and is followed by a presentation of the data to match and the underlying physical processes. Then, Section 4 develops the proposed workflow to build a compatible prior in three steps: first, by designing two TIs, we suggest two alternative conceptual models; second, we introduce some modifications to the graph cut inversion algorithm that are needed to avoid inversion artifacts; third, we propose a method to calibrate the algorithmic parameters prior to the inversion. Section 5 displays the inversion results and a convergence analysis. Section 6 discusses our findings and highlights the methods potential within the field of hydrogeology.

2 Probabilistic inversion with graph cuts

2.1 General formulation

We employ the newly developed Bayesian inversion method by *Zahner et al.* [2016]. The model parameters \mathbf{m} are characterized by a prior pdf $p(\mathbf{m})$. The observed data, denoted by \mathbf{d} , are compared to simulated forward responses $g(\mathbf{m})$. The posterior pdf $p(\mathbf{m}|\mathbf{d})$ of the model parameters conditional on the data is inferred from Bayes theorem [*Tarantola and Valette*, 1982]:

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m}) p(\mathbf{m})}{p(\mathbf{d})}, \quad (1)$$

where $p(\mathbf{d}|\mathbf{m})$ denotes the likelihood $L(\mathbf{m})$ and $p(\mathbf{d})$ is the evidence. In the method of *Zahner et al.* [2016], the evidence is an unknown constant and only the two terms in the numerator of Equation 1 are considered and needed to solve the parameter inference problem.

MCMC methods sample model realizations proportionally to their posterior probability by performing a random walk through the parameter space [e.g. *Sambridge and Mosegaard*, 2002; *Robert and Casella*, 2013]. At each step of the walk, the current model realization \mathbf{m}_{cur} is slightly perturbed by a model proposal step that produces a model proposal \mathbf{m}_{prop} . Then, either the current model is kept or the model proposal is accepted. In the Metropolis algorithm [*Metropolis and Ulam*, 1949], a symmetric proposal distribution is used and the probability to accept the proposed model is:

$$P_{\text{acc}} = \min \left(1, \frac{p(\mathbf{d}|\mathbf{m}_{\text{prop}}) p(\mathbf{m}_{\text{prop}})}{p(\mathbf{d}|\mathbf{m}_{\text{cur}}) p(\mathbf{m}_{\text{cur}})} \right). \quad (2)$$

If the model proposals are drawn according to the prior pdf, which is the case for the algorithm by *Zahner et al.* [2016], then the algorithm is an extended Metropolis algorithm [*Mosegaard and Tarantola*, 1995], and the acceptance probability of the model proposal can be calculated without considering the prior probability:

$$P_{\text{acc}} = \min \left(1, \frac{p(\mathbf{d}|\mathbf{m}_{\text{prop}})}{p(\mathbf{d}|\mathbf{m}_{\text{cur}})} \right) = \min \left(1, \frac{L(\mathbf{m}_{\text{prop}})}{L(\mathbf{m}_{\text{cur}})} \right). \quad (3)$$

The initial phase of the MCMC run corresponds to the burn-in phase. The end of the burn-in is often defined in relation to when the Markov chains start to oscillate around a constant likelihood or posterior probability. After the burn-in phase, the MCMC algorithm explores and samples the posterior pdf. MCMC algorithms are often tuned to ensure an appropriate acceptance rate α , defined as the fraction of accepted model proposals. A classical target value in high dimensions is often 23% [*Roberts et al.*, 2001; *Xu et al.*, 2006; *Hansen et al.*, 2012], while acceptance rates as low as 3% have allowed for good posterior inference [*Efendiev et al.*, 2005].

In MPS-based MCMC algorithms, the prior is sampled by running the algorithm and accepting all model proposals. The prior realizations thus obtained present very similar MPS characteristics to those of the TI used by the algorithm. Nevertheless, we emphasize that the actual prior is not the TI, but the results of the MPS based MCMC algorithm in the absence of data for a given set of algorithmic parameters and TI used. The quality and geological realism of the prior is best appreciated using the movie strategy proposed by

Tarantola [2005]. The underlying idea being that emphasis should be placed on obtaining prior realizations that appear geologically reasonable in the eye of experts with prior knowledge about the site characteristics.

2.2 Graph cut inversion

The proposal step mechanism by *Zahner et al.* [2016] (Figure 1) consists in merging two images by an optimal cut, such that the complementary assembled parts do not create artifacts where they overlap, for example, discontinuities that are not present in the TI. The graph cut model proposal mechanism is composed of three steps. Starting with the current spatial model \mathbf{m}_{cur} and a random patch \mathbf{m}_{TI} from the TI that has the same size as \mathbf{m}_{cur} (Figure 1A), a cost image δ is computed as the absolute value of the difference between the two images (spatial models): $\delta = |\mathbf{m}_{\text{cur}} - \mathbf{m}_{\text{TI}}|$ (Figure 1 panel ①). Regions where

Figure 1: Schematic representation of the graph cut model proposal algorithm proposed by *Zahner et al.* [2016]. A) A small fraction of the TI around the random patch \mathbf{m}_{TI} from the TI, B) current state of the model parameters \mathbf{m}_{cur} , C) cost image $\delta = |\mathbf{m}_{\text{cur}} - \mathbf{m}_{\text{TI}}|$, D) Randomly chosen terminals and the resulting cut, E) resulting model proposal \mathbf{m}_{prop} .

the cost δ is greater than the mean of the cost image $\bar{\delta}$ are defined as possible terminals; two of them that have a similar size, the source \mathbf{s} (blue) and the sink \mathbf{t} (green), are randomly selected (Figure 1 panel ②). Finally, the trajectory (Figure 1C) of an optimal cut separating \mathbf{s} and \mathbf{t} is computed using the highly efficient min-cut/max-flow algorithm developed by *Boykov et al.* [2001]. This cut defines the size and shape of the patch $\mathbf{m}_{\text{patch}}$ that is copied from \mathbf{m}_{TI} and pasted onto \mathbf{m}_{cur} (Figure 1C) panel ③) to form \mathbf{m}_{prop} (Figure 1D)).

The likelihood is computed by comparing the simulated forward response $g(\mathbf{m})$ with the field data \mathbf{d} . We assume that the measurement data errors, denoted by ϵ follow an exponential distribution, which is more robust with respect to possible outliers than a Gaussian distribution [*Claerbout and Muir*, 1973]. We further assume that they are independent, have a zero mean and a constant mean deviation σ_{ϵ} . Under these assumptions the likelihood function L is given by [*Tarantola*, 2005]:

$$L(\mathbf{m}) = \frac{1}{2^N \sigma_{\epsilon}^N} \times \exp\left(-\sum_{i=1}^N \left| \frac{g_i(\mathbf{m}) - d_i}{\sigma_{\epsilon}} \right|\right), \quad (4)$$

where N denotes the number of observations, and subscript i indicates the i^{th} observation.

The algorithm proposed by *Zahner et al.* [2016] is described in Algorithm 1. The model parameters \mathbf{m} are initialized by taking a random patch from the TI. This patch has the same size as the simulation domain (this is one reason why the TI must be several times larger than the simulation). The likelihood is computed for the initial state. Then, the algorithm enters a loop, where at each iteration, a new model proposal is made using the graph cut model proposal mechanism. The likelihood of the model proposal is computed (Equation 4) and the model is accepted or rejected according to the acceptance probability (Equation 2). For more details we refer the reader to *Zahner et al.* [2016].

Algorithm 1 Extended Metropolis algorithm proposed by *Zahner et al.* [2016] for inversion with graph cuts

- 1: Initialize the model parameters by drawing \mathbf{m}_1 randomly in the TI
 - 2: Compute the likelihood $L(\mathbf{m}_1)$ for the initial state (Equation 4)
 - 3: Set $k = 1$
 - 4: **loop**
 - 5: Set $\mathbf{m}_{\text{cur}} = \mathbf{m}_k$
 - 6: Propose $\mathbf{m}_{\text{prop}} = \mathbf{m}_{\text{cur}} + \text{MPS perturbation based on graph cuts}$
 - 7: Compute the likelihood $L(\mathbf{m}_{\text{prop}})$ for the proposed state (Equation 4)
 - 8: Accept \mathbf{m}_{prop} with probability P_{acc} (Equation 3)
 - 9: **if** \mathbf{m}_{prop} is accepted **then**
 - 10: $\mathbf{m}_{k+1} = \mathbf{m}_{\text{prop}}$
 - 11: **else**
 - 12: $\mathbf{m}_{k+1} = \mathbf{m}_{\text{cur}}$
 - 13: **end if**
 - 14: Set $k = k + 1$
 - 15: Stop when the maximum number of iterations is reached
 - 16: **end loop**
-

Zahner et al. [2016] demonstrated the efficiency and stability of this algorithm on various synthetic test cases: continuous multi-Gaussian fields, discrete channels and lenses. Applying this method to a real case requires two kinds of data: first, interpreted field observations to build one or more TIs and second, indirect data such as geophysical or hy-

hydrogeological measurements. In addition, an appropriate forward model is required to compute the simulated data for a given model proposal.

3 The Boise Hydrogeophysical Research Site and geophysical modeling

The BHRS is located about 15 km southeast of downtown Boise, Idaho, USA, on a gravel bar adjacent to the Boise river (Figure 2A). This braided river aquifer consists of late Quaternary fluvial deposits, dominated by coarse cobble-and-sand over a clay layer situated at about 20 m depth. Over the years, this site has developed into a field laboratory, that is mostly used to test and improve methods to characterize aquifer properties [Barrash and Clemo, 2002] by refining various hydrogeological or geophysical characterization methods such as tracer tests [Hu *et al.*, 2009; Thoma *et al.*, 2014], slug tests and hydraulic tomography [Cardiff *et al.*, 2013a; Hochstetler *et al.*, 2015], oscillatory pumping tests [Cardiff *et al.*, 2013b; Rabinovich *et al.*, 2015], seismic [Moret *et al.*, 2006] and ground penetrating radar (GPR) [Tronicke *et al.*, 2004; Bradford *et al.*, 2009; Dafflon and Barrash, 2012]. The comprehensive testing that has been carried out at the BHRS makes it very suitable for testing new inversion methodologies. [Fienen *et al.*, 2008; Dafflon *et al.*, 2009].

3.1 Geostatistical analysis of porosity and outcrop data

Barrash and Clemo [2002] performed a three-level hierarchical geostatistical analysis of porosity at the BHRS that was based on 4699 measurements. The first level of analysis considers all porosity values as a global dataset. At the second level of analysis, a combination of borehole porosity logs with crosshole GPR reflection profiles are used to classify the subsurface deposits into five units. Units 1 to 4 are dominated by cobbles and contain some small sand lenses and channels, but their porosity distributions are different, as confirmed by various statistical tests realized by Barrash and Clemo [2002]. Unit 5 is a single unit comprised of a sand-filled erosional channel. In the third level of analysis, unit 4 was decomposed into three subunits according to complementary geological information from an outcrop, as a way to understand the porosity of unit 4 as a sum of three multi-Gaussian porosity distributions. From a sedimentological point of view, units 1 to 4 are successive layers, deposited on top of each other while unit 5 manifests itself as inclusions within other units. The results of this detailed geostatistical analysis on level I are described in Table 1.

Table 1: Global geostatistical analysis of porosity at the BHRS by *Barrash and Clemo* [2002]

Level	%	Porosity mean	Porosity variance	direction	Variogram model	parameters nugget	range, m	sill
I	100.0	0.222	0.00429	vertical	exponential	-	5.13	0.00481
				horizontal	periodic	-	4.6	0.00167
				horizontal	periodic	-	46	0.00226

Although the analysis by *Barrash and Clemo* [2002] describes the detailed two-point statistics of the porosity field, the characterization of geological structures can be enhanced by adapting an MPS framework that integrates analog data such as the outcrop located 2.5 km upstream of the BHRS [*Hinz and Bradford*, 2010]. The $23.7 \text{ m} \times 12.4 \text{ m}$ outcrop (Figure 2D) is expected to be representative of the BHRS. The photograph of the outcrop allows capturing two-dimensional characteristic length scales and connectivity patterns of the sand units among the cobble-dominated units at a high resolution (8.3 mm). It is used herein to build a TI of the porosity after appropriate image processing (e.g., adapting the resolution of the picture, correcting light effects, accounting for the porosity distribution) as detailed in Section 4.

Figure 2: Presentation of the BHRS dataset: A) site location, B) porosity distribution at the site, C) high resolution photograph of the outcrop located 2.5 km upstream from the BHRS, D) velocity derived from deterministic smoothness-constrained inversion of crosshole first-arrival GPR travel time data between wells A1 and B2.

3.2 Cross-hole GPR data and forward geophysical simulator

Here we consider the crosshole first-arrival GPR travel time data presented by *Irving et al.* [2007]. The measurements were performed between two wells, labeled A1 and B2, separated by 4 m over a depth range of about 15 m within the saturated zone. The emitter and receiver positions were varied every 0.2 m along the vertical direction. The specified center frequency for the antennas during the survey was 250 MHz. The travel time data considered were limited to ray paths with angles lower than 30° to avoid problems associated with high-angle ray paths [*Peterson*, 2001].

We performed a deterministic smoothness-constrained inversion from the entire data set of first-arrival travel-times. The inversion includes receiver-position-dependent travel-time corrections as well as angle-dependent travel-time corrections. The velocity field resulting from this deterministic inversion is displayed in Figure 2D) for a model discretization of $0.2 \text{ m} \times 0.2 \text{ m}$.

As the outcrop available to build the TI is only 23.7 m wide by 12.4 m deep, we limit the model domain to a width of 4 m and a depth range of 8 m (between 10 m and 18 m as referred in Figure 2D)). The associated GPR data set is composed of $N = 743$ first-arrival travel times. In addition, as inversion results are sensitive to the density of the measured ray paths, the plotted posterior realizations are restrained to a depth range between 11 m and 17 m, which we refer to as the zone of interest.

The forward model g is a composition of two functions $g = f \circ s$, producing data simulations $g(\mathbf{m})$. First, a slowness field s is computed from the model parameters \mathbf{m} . Second, a geophysical forward solver f is used to transform the experimental design and the slowness field into travel times. The slowness field s is computed in two steps. Step 1 converts the model parameters \mathbf{m} , composed of a porosity field Φ , a cementation factor m , the rock matrix dielectric constant κ_s and the water dielectric constant κ_w , into an effective dielectric constant κ_{eff} using the relation by *Pride* [1994]:

$$\kappa_{eff} = \Phi^m (\kappa_w + (\Phi^{-m} - 1)\kappa_s) . \quad (5)$$

Step 2 computes the slowness field s with Equation 6:

$$s = \frac{\sqrt{\kappa_{eff}}}{c}, \quad (6)$$

where c is the speed of light in a vacuum. The corresponding first-arrival travel times are computed with f [Podvin and Lecomte, 1991], over the depth between 10 m and 18 m, by solving the Eikonal equation with emitter-receiver positions following the same experimental design as for the field data acquisition. The grid-resolution is set to 0.1 m. This leads to a 39×79 grid for the posterior realizations. The zone of interest is limited to a 39×60 grid.

4 Building a compatible prior scenario

In this section, we address the issue of defining a TI and ranges of petrophysical parameters, and setting algorithmic parameters such that 1) the Markov chain converges to

a data misfit that is in agreement with the expected noise characteristics and 2) the type of structures found in the TI are largely preserved in the sampled posterior realizations.

Ideally, the prior contains only what is truly known (nothing more, nothing less) about a site prior to considering the available data. In actual subsurface investigations, this is often very difficult in high parameter dimensions and the only viable approach (in our view) is to define alternative prior scenarios. In the present work, our focus is to present one of the first field-based MPS inversions and to contrast the results obtained by two fundamentally different conceptions of the subsurface structure (i.e., very different priors). The prior scenario is qualified as compatible when the spatial structure characteristics of the prior are propagated into the posterior and when the field observations and data are honored.

4.1 Designing porosity TIs of the BHRS

When considering geological properties to invert for, the prior and especially the TI (Figure 1A) should be consistent with the geological knowledge of the site under consideration, as well as with the data to match. The TI can be derived from outcrop photography (Figure 2C) or remote sensing data [Bayer *et al.*, 2015; Dickson *et al.*, 2015], drawn by a geologist, issued from object-based models [Deutsch and Tran, 2002], process-based models [Pyrz *et al.*, 2009], physics-based models [Davy and Lague, 2009] or pseudo-genetic models [Pirot *et al.*, 2015a]. The TI should ideally have a much larger size than the model domain. In case the TI is too small, various techniques exist to enrich it [Comunian *et al.*, 2012; Rezaee *et al.*, 2015].

The first step in TI model building consists in defining one or several alternative geological conceptual models. Here, we introduce two porosity TIs that are in accordance with available porosity data at the BHRS. The first TI relies exclusively on the analysis by Barrash and Clemo [2002] and depicts the porosity field as a multi-Gaussian random field. The second TI exploits the outcrop information, as well as the first and second moments of the porosity distribution at the BHRS. We expect that the posterior samples obtained from this model will look geologically more realistic than those obtained by the first TI. In what follows, the TIs have the same grid-resolution as the modeled domain: $0.1 \text{ m} \times 0.1 \text{ m}$.

4.1.1 Multi-Gaussian TI

From the detailed geostatistical analysis of the porosity field at the BHRS [Barrash and Clemo, 2002], we use the first level variogram model parameters (Table 1) to generate a $250 \text{ m} \times 250 \text{ m}$ multi-Gaussian simulation that defines the standard score spatial structure of the porosity field, denoted by S . S is characterized by a zero mean $\bar{S} = 0$ and a standard deviation $\sigma_S = 1$. The porosity TI is then obtained by applying a lognormal transformation $\Phi_{MGS} = \exp(S \times 0.22361 - 1.579)$, which corresponds to the lognormal distribution of the porosity that is depicted in Figure 2B. A fraction (711 px \times 372 px) of the resulting multi-Gaussian TI is displayed in Figure 3A.

Figure 3: Porosity TIs, A) 71.1 m \times 37.2 m multi-Gaussian structures extracted from a 250 m \times 250 m TI, B) binary sand lenses derived from the 23.7 m \times 12.4 m outcrop and C) corresponding porosity field. The black rectangle represents the size of the model domain.

4.1.2 From outcrop photograph to TI

The main idea behind interpreting an outcrop picture (Figure 2C) as a porosity TI is based on the supposed correlation between the gray-scale values of the photograph and the grain size [Hinz and Bradford, 2010]. Indeed, finer grains that form sand lenses appear whiter than cobbles that form the matrix. Then, the identified sand lenses can be assigned the porosity mean and standard deviation computed by Barrash and Clemo [2002] for the sand unit (5), and the matrix porosity mean and standard deviation are adjusted accordingly such that the global porosity mean and standard deviation match those computed by Barrash and Clemo [2002]. However, one should be aware that the relationship between the gray-scale and the porosity might not be linear. In addition, the non-uniform brightness over the picture needs to be corrected and the resolution of the picture has to be adapted to fit the resolution of the simulation grid.

To account for these considerations, we apply the following processing steps to build a porosity TI (Figure 3 C) from the high-resolution outcrop picture:

1. filter out the effect of non-uniform exposure by subtracting the moving average of the original gray-scale picture over a $1.66 \text{ m} \times 1.66 \text{ m}$ window which corresponds to a $200 \text{ px} \times 200 \text{ px}$ window at the original resolution,
2. upscale the filtered image from a 0.0083 m resolution to a 0.1 m resolution by arithmetic averaging,
3. manually classify the filtered upscaled porosity image into sand and matrix facies (Figure 3 B)),
4. compute the matrix porosity mean $\bar{\phi}_M = \frac{\bar{\phi}_{S+M} - p_S \times \bar{\phi}_S}{1 - p_S}$ and standard deviation $\sigma_M = \frac{\sigma_{\phi_{S+M}}^2 + \bar{\phi}_{S+M}^2 - [p_S \times \bar{\phi}_S^2 + (1 - p_S) \times \bar{\phi}_M^2 + p_S \times \sigma_{\phi_S}^2]}{1 - p_S}$ given the sand proportion p_S , porosity mean $\bar{\phi}_S$ and standard deviation σ_{ϕ_S} and the overall porosity mean $\bar{\phi}_{S+M}$ and standard deviation $\sigma_{\phi_{S+M}}$,
5. transform the gray-scale image into a porosity TI $\Phi_{outcrop}$ (Figure 3 C) by rescaling linearly the gray-scale values within each lithological facies with the corresponding porosity mean and standard deviation.

4.2 Adaptations of the graph cut inversion algorithm

When applying the graph cut inversion algorithm proposed by *Zahner et al.* [2016] to real data, we found that four refinements were needed to preserve the structures defined in the TIs along the sampling process.

1) The first adaptation is the introduction of a cutting threshold δ_{th} . In the original algorithm, the terminals (source and sink) are defined as regions where the differences δ are greater than $\delta_{th} = \bar{\delta}$, the mean of the differences δ . The choice of δ_{th} influences the definition of the terminals, and hence also the cut and the model proposal \mathbf{m}_{prop} . Too high values of δ_{th} can lead to poor cuts through regions where there is a large overlap error. In such cases, after many iterations, the propagation of bad cuts can produce posterior realizations of poor quality, with visible sharp artifacts and contrasts as illustrated in Figure 4, which break the coherence of the structures defined in the TI. In what follows, we define $\delta_{th} \in [0, 100]$ as a percentage of $\max(\delta)$.

2) The second adaptation consists in accounting for the anisotropy $r_{\frac{x}{z}}$ that is present in the TI in the cutting cost function. This is done to prevent cutting through high δ values. *Boykov et al.* [2001] and *Zahner et al.* [2016] consider an isotropic ($r_{\frac{x}{z}} = 1$) cutting cost that is only related to the length of the cut and to the cost image δ . We propose to

weight the vertical cutting cost c_i^z by the anisotropy ratio compared to the horizontal cutting cost c_i^x : $c_i^z = r_{\frac{x}{z}} \times c_i^x$, where c_i^x is defined between cells j and k as $c_i^x = \delta_j + \delta_k$ [for more details, see *Zahner et al.*, 2016]. The effect of the anisotropy ratio $r_{\frac{x}{z}}$ on posterior realizations is illustrated in Figure 4, jointly with the effect of δ_{th} . For $r_{\frac{x}{z}} = 1$, artifacts in the posterior realizations appears for $\delta_{th} \geq 21\%$, while with $r_{\frac{x}{z}} = 3.4$ they start to appear at $\delta_{th} \geq 30\%$. It shows that the use of this parameter reduces sharp cuts and broadens the range of δ_{th} for which the posterior realizations are visually close to the TI.

Figure 4: Influence of the cutting threshold δ_{th} and the anisotropy ratio $r_{\frac{x}{z}}$ on posterior realizations after 10^5 iterations of the MCMC chain.

3) The third adaptation is related to issues caused by overlapping property distributions in the TI and does not concern the multi-Gaussian TI. In the outcrop-derived TI, the porosity distribution of the sand lenses and the cobble dominated matrix are overlapping. This implies that the GPR data might be explained by the cobble dominated matrix only, and it is possible to obtain posterior realizations without any sand lenses. To better understand why this happens, let us consider the equivalent velocity that corresponds to the velocity of an homogeneous medium for which the travel time residuals are centered on zero. This equivalent velocity is centered on the velocity distribution of the matrix rather than on the velocity distribution of the sand lenses. Consequently, sand lenses are likely to disappear in the sampling of the posterior. As we aim not to fit our observations with random noise but rather with realistic geological structures, we introduce a hard constrain in the proposal mechanism to ensure that the sand facies proportion must be above 2% in the zone of interest. In practice, this implies that the proposal step (line 6 of Algorithm 1) is repeated until this condition is satisfied. Of course, this requires the ability to differentiate facies in the TI.

4) The last adaptation is related to the inversion of petrophysical parameters and the mean deviation of the observational noise in addition to the porosity field. For a sandy gravel aquifer, we assume a uniform prior for $m \in [1.3; 1.8]$ [*Lesmes and Friedman*, 2005] and $\kappa_s \in [3; 6]$ and a known dielectric constant for water $\kappa_w = 81$ [*Daniels*, 1996]. The prior of the mean deviation of the observational noise is considered uniform with

$\sigma_\epsilon \in [0.2; 0.8]$ ns [Hansen et al., 2014]. To invert for m , κ_s and σ_ϵ in addition to the porosity field (4 variables in total), the proposal step is enriched as follows. The number of variables to update is randomly drawn. The probability used to update 1, 2, 3 or 4 variables is respectively 0.66, 0.19, 0.10 or 0.05. The variables to update are randomly drawn according to their update probability. The probability to update the porosity field is set to 0.7, while the remaining variables have a probability of 0.1 each. When m , κ_s or σ_ϵ are updated, the proposed value is drawn from a Gaussian proposal distribution that is centered on the current value and with a standard deviation of respectively 1/50, 1/50 and 1/30 of their prior width. Folding is used to stay within prior boundaries.

4.3 Pre-calibration of algorithmic settings

We assess the anisotropy ratio $r_{\frac{x}{z}}$ directly from the TI and calibrate the cutting threshold δ_{th} by running preliminary MCMC chains and performing a sensitivity analysis. Our aim is to enhance convergence by targeting a suitable log-likelihood, acceptance rate and a mean update size of the porosity field around $20 \pm 10\%$ of the model size, while preserving geological structures as defined in the TI.

The anisotropy ratio is simply computed after the TI characteristic length scales. For the multi-Gaussian TI, the anisotropy ratio is computed from the characteristic length scales provided by Barrash and Clempson [2002]: $r_{\frac{x}{z}}^{MGS} = \frac{5.8}{1.7} = 3.4$. Regarding the outcrop-derived TI, it is computed as the ratio of the mean length of sand lenses over their mean thickness $r_{\frac{x}{z}}^{outcrop} = 9.2$.

The cutting threshold δ_{th} influences the quality of the MPS model proposal, as explained in Section 4.2. Over several iterations, it might lead to deteriorated structures that are markedly different than those in the TI. As δ strongly depends on the structures and contrasts present in the TI, δ_{th} can not be universally defined as a fixed proportion of the maximum value of δ . Therefore, a sensitivity analysis of δ_{th} on the resulting mean log-likelihood, mean update size and acceptance rate is performed (Figure A.1). It results in $\delta_{th} = 11\%$ for the multi-Gaussian TI and of $\delta_{th} = 26\%$ for the outcrop-derived TI. More details about the suggested calibration approach are given in Appendix A: .

5 Results

After implementing our proposed modifications to the algorithm by *Zahner et al.* [2016], the graph cut algorithm was run for $n_{chains} = 20$ chains of length $2n = 2 \cdot 10^6$ iterations for both TIs defined in Section 4.1. Each chain run required 16 hours of single CPU time on one processor. In what follows, the posterior distribution shown is limited to the second half of each chain. The convergence of the chains are quantified using the Gelman-Rubin statistic R [*Gelman and Rubin*, 1992].

For both TIs, the mean of the 10^6 posterior realizations over the second half of each chain and over all chains are in agreement with the deterministic inversion results (Figure 5). Two zones of higher porosities can be identified around depths of $z = 12.7$ m (though

Figure 5: A), B) and C): porosity inversion results for A) single deterministic inversion with $m = 1.6$ and $\kappa_s = 5$, B) mean of the 10^6 posterior realizations sampled by graph cut inversion with the multi-Gaussian TI and C) with the outcrop-derived TI; to facilitate the comparison of inversion results presenting different porosity ranges, the color scale has been stretched to its higher values; D) and E): standard deviation over the 20 chains for D) the multi-Gaussian TI and E) the outcrop-derived TI.

less obvious for the posterior mean obtained using the outcrop-derived TI) and $z = 15.3$ m. These results are coherent with those obtained by *Dafflon and Barrash* [2012]. The mean of the posterior in the multi-Gaussian case is smooth and porosity values are comprised between 0.19 and 0.30, while in the outcrop-derived TI case, porosity values are comprised between 0.18 and 0.35. The standard deviation of the posterior porosity field derived from the multi-Gaussian TI (Figure 5D) is rather homogeneous. On the contrary, the standard deviation of the posterior porosity field obtained with the outcrop-derived TI (Figure 5E) is heterogeneous. There is less variability in the gravel matrix and more at the possible location of sand lenses and on the border of the domain. The likelihood ratio Λ of the scenario based on a multi-Gaussian field TI (MGS) over the scenario based on a outcrop-derived TI (OUT) is computed as $\Lambda = \frac{\sum_k \sum_i L_{MGS}^{k,i}}{\sum_k \sum_i L_{OUT}^{k,i}} = 4 \cdot 10^{112}$ where $1 \leq k \leq 20$ denotes the chain, $10^6 + 1 \leq i \leq 2 \cdot 10^6$ denotes the posterior realization in the related chain and L denotes the likelihood.

Detailed results and convergence analysis are shown in Figure 6 for the posterior sampling based on the multi-Gaussian TI. The posterior realizations are bimodal and re-

Figure 6: Posterior realizations for two chains and convergence analysis for the multi-Gaussian TI:

A) to D) posterior realizations for chain 1, E) posterior mean for chain 1 and F) posterior standard deviation for chain 1; G) to J) posterior realizations for chain 2, K) posterior mean for chain 2 and L) posterior standard deviation for chain 2; M) to P): Gelman-Rubin statistics for various chain lengths (M): $2n = 2.5 \cdot 10^5$, N): $2n = 5 \cdot 10^5$, O): $2n = 10^6$; P): $2n = 2 \cdot 10^6$; Q): Gelman-Rubin statistics for the petrophysical parameters as a function of the chain length; R): *WMAE* for the second half of the 20 chains; S): posterior distribution of m and initial value for each chain; T): posterior distribution of κ_s and initial value for each chain; U): posterior distribution of σ_ϵ and initial value for each chain.

veal similar locations for zones of high porosities, while variations between realizations are present at smaller scales. One mode, illustrated by chain 1, is characterized by $m \approx 1.8$, $\kappa_s \approx 3$ and a porosity field comprised in $[0.25; 0.38]$. The second mode, illustrated by chain 2, is characterized by $m \approx 1.3$, $\kappa_s \approx 6$ and a porosity field comprised in $[0.11; 0.22]$. These posterior distribution modes can be compared with the $[1.3; 1.8]$ uniform prior for m , with the $[3; 6]$ uniform prior for κ_s and with the initial porosity range $[0.05; 0.45]$ defined in the TI. Within each chain, the variability is rather homogeneous. The weighted mean absolute error (*WMAE*, Equation 7) is computed as

$$WMAE = \frac{1}{N} \sum_{i=1}^N \frac{|\epsilon_i|}{\sigma_\epsilon}, \quad (7)$$

where i denotes an observation and σ_ϵ the measurement error, stabilizes around 1 after 10^5 iterations already and independantly from the mode. MCMC chains convergence is commonly accepted when the Gelman-Rubin statistic R is below 1.2. As $R \gg 1.2$, there is no convergence of the porosity fields, of m and κ_s for chains of length $2n = 2 \cdot 10^6$. The Gelman-Rubin statistic R even increases strongly from 10^6 to $2 \cdot 10^6$ steps, due to independent convergence of the 20 chains to a local mode. There is convergence only for the data measurement error standard deviation σ_ϵ whose posterior distribution looks Gaussian and is centered around 0.3 ns. The acceptance rate is 7%.

Detailed results and convergence analysis are shown in Figure 7 for the posterior sampling based on the outcrop-derived TI. Sand lenses are present at slightly different lo-

Figure 7: Posterior realizations for two chains and convergence analysis for the outcrop-derived TI: A) to D) posterior realizations for chain 1, E) posterior mean for chain 1 and F) posterior standard deviation for chain 1; G) to J) posterior realizations for chain 2, K) posterior mean for chain 2 and L) posterior standard deviation for chain 2; M) to P): Gelman-Rubin statistics for various chain lengths (M): $2n = 2.5 \cdot 10^5$, N): $2n = 5 \cdot 10^5$, O): $2n = 10^6$; P): $2n = 2 \cdot 10^6$); Q): Gelman-Rubin statistics for the petrophysical parameters as a function of the chain length; R): $WMAE$ for the second half of the 20 chains; S): posterior distribution of m and initial value for each chain; T): posterior distribution of κ_s and initial value for each chain; U): posterior distribution of σ_ϵ and initial value for each chain.

cations and with different shapes when we compare realizations of different chains. The standard deviation of the porosity is higher across the sand lenses and close to the boreholes. The range of the porosity $[0.05; 0.5]$ defined in the TI (Figure 3C) is rather well preserved in the posterior $[0.15; 0.49]$. The $WMAE$ stabilizes around 1 after 10^5 iterations. The Gelman-Rubin statistic R decreases for the porosity field realizations as well as for the petrophysical parameters but convergence is not yet achieved as $R > 1.2$. However, the posterior distribution of m converges toward a lognormal distribution centered around 1.3, the posterior distribution of κ_s converges toward a lognormal distribution centered around 3.25 and the posterior distribution of σ_ϵ converges toward a lognormal distribution centered around 0.5 ns. The acceptance rate is 7%.

6 Discussion

After appropriate modifications, we find that the first field-based demonstration of the graph cut inversion introduced by Zahner *et al.* [2016] is overall a success. First, the method is able to produce posterior structures that are in agreement with the TI and that fit the observational data, jointly with the inversion of the petrophysical parameters and the data measurement error standard deviation. Second, the main features of the porosity field are properly retrieved and consistent 1) with inversion results obtained with other methods such as smoothness-constrained deterministic inversion [Irving *et al.*, 2007] or

based on simulated annealing [Dafflon and Barrash, 2012] and 2) with porosity logs [Dafflon and Barrash, 2012]. We argue that this is a significant step forward as there are very few successful field applications of MPS-based inversions [Linde *et al.*, 2015a]. This inversion method is valuable when the prior cannot be stated in an explicit mathematical form such as multi-Gaussian random fields (for these cases, there exists much faster alternatives) but can be generated from a TI. Obviously, the global computational time of the algorithm also relies strongly on the efficiency of the forward model calculations.

The important algorithmic speed-up [by a factor 40 according to Zahner *et al.*, 2016] with respect to previously presented alternatives [e.g. iterative spatial re-sampling, Mariethoz *et al.*, 2010a] is balanced by a slow or limited convergence when applying the method to field data. Indeed when using the multi-Gaussian TI, the joint inversion for the porosity field with measurement error and petrophysical parameters does not converge because individual chains are exploring individual posterior modes. A computation of the Gelman Rubin statistic on the velocity field, which results from the combination of the porosity field with the petrophysical parameters, shows convergence for chains of length $2 \cdot 10^6$. This bimodal posterior is the combined result of unimodal chains. A thinner prior range would allow removing the less common modes that correspond to a rather unlikely combination of $m = 1.8$ and $\kappa_s = 3$. However, these results obtained with a wide prior illustrate the practical challenges of high-dimensional MCMC for real data sets.

When considering a pre-determined measurement error ($\sigma_\epsilon = 0.5$ ns) and petrophysical parameters ($m = 1.48$ and $\kappa_s = 6$), the graph cut inversion using the multi-Gaussian TI converges quickly, already after 10^5 MCMC steps (see Figure 8). The intrinsic vari-

Figure 8: Posterior realizations over two chains of length 10^6 and convergence analysis over 20 chains of length 10^6 with $m = 1.48$, $\kappa_s = 6$ and $\sigma_\epsilon = 0.5$ ns for the multi-Gaussian training image: A) to D) posterior realizations for chain 1, E) posterior mean for chain 1 and F) posterior standard deviation for chain 1; G) to J) posterior realizations for chain 2, K) posterior mean for chain 2 and L) posterior standard deviation for chain 2; M) to O): Gelman-Rubin statistics for various chain lengths (M): $2n = 62500$, N): $2n = 125000$, O): $2n = 250000$); P): *WMAE* for the second half of the 20 chains.

ability of the posterior realizations in a chain is similar to the variability of the posterior realizations between different chains, which leads to Gelman-Rubin statistics below 1.2.

The posterior realizations fit well the measured geophysical data as the *WMAE* stabilizes around 0.76. This example illustrates how well-determined petrophysical parameters can facilitate convergence.

In the case of the outcrop-derived TI, when inverting for the porosity field, the measurement error and the petrophysical parameters, the Gelman-Rubin statistic does not indicate convergence of the algorithm after $2 \cdot 10^6$ iterations, but the convergence is better than in the multi-Gaussian case. The posterior of the measurement error and petrophysical parameters present a unique posterior mode. The data measurement error standard deviation posterior is higher than in the multi-Gaussian case because maintaining a much higher porosity contrast throughout the posterior is much more constraining. Nevertheless, the inversion results present realistic spatial structures of the porosity field, in conformity with the TI. Moreover, the porosity range derived from the analysis performed by *Bar-rash and Clemo* [2002] is better preserved in the posterior than for the multi-Gaussian TI. In addition, the relatively small size of the TI appears sufficient to adequately fit the data and provide a posterior in agreement with the prior structures. These advantages make the inversion results for the outcrop-based TI more efficient and geologically more realistic.

The slow convergence in the case of the outcrop-derived TI is not due to the small size of the TI. Using larger TIs built from truncated ellipsoids and presenting similar lense structures as in the outcrop did not improve the inversion or the convergence results. The slow convergence manifests itself in that the intrinsic variability of the chain is smaller than the variability between the chains. The cause of this is related to the sharp and strong contrasts of the sand and matrix properties that result in a complex likelihood surface that is difficult to sample [*Laloy et al.*, 2016]. Indeed, once the posterior has been located (i.e., the data are adequately fitted), it is highly unlikely that large geometrical rearrangements of the proposed model will be accepted. This is also the underlying reason for the bi-modal posterior of the multi-Gaussian case.

From a more pragmatic perspective, the slow convergence when using the outcrop-derived TI might not be a real issue: from a visual perspective, the sand lenses are approximately located at the same place. As the *WMAE* quickly stabilizes after 10^5 iterations, a more practically-oriented non-formal approach to sample a larger range of poste-

prior realizations might be to run many chains and stop the chains as soon as the *WMAE* gets below a user-defined threshold. Another approach to increase the intrinsic variability of the chains and to improve the convergence would be to continue the sampling of the posterior with a classical iterative geostatistical resampling technique [Mariethoz *et al.*, 2010a; Hansen *et al.*, 2012] once the *WMAE* falls below some threshold; the large scale structures would be put in place by the graphcut algorithm, and smaller-scale features would be updated iteratively using a geostatistical simulation method such as direct sampling [Mariethoz *et al.*, 2010b]. Another possibility to improve convergence in both cases could be to reinforce the possibility to jump from one mode to another one by allowing the chains to communicate [Vrugt *et al.*, 2009].

More generally, when using MPS-based methods, the choice of the TI strongly influences the results of the algorithm. Ideally, the modeler should generate multiple prior scenarii that cover the prior knowledge uncertainty such as different geological features or parameter distributions [Piot *et al.*, 2015b]. The scarcity of field data can be compensated by the use of analog data to define elements of the prior. For the proposed algorithm, the prior is an ensemble of realizations that are generated by the algorithm when setting the likelihood to unity: it depends on the choice of both the TI and the graph cut parameters (cutting threshold and anisotropy ratio). Here two prior model scenarii were proposed.

A very rough approach to model (scenario) comparison is to compute likelihood ratios. In the presented field application, we obtain an extremely high ratio of $\Lambda = 4 \cdot 10^{112}$ in favor of the prior scenario based on a multi-Gaussian derived TI. However, it does not consider the realistic aspect of spatial geological structures and the risk of fitting observation noise or model simulation errors. Theoretically solid model comparison uses the Bayes factor [Kass and Raftery, 1995], but its computation, by Monte Carlo estimation is very expensive. Schöniger *et al.* [2014] show that results obtained by popular information criteria that account for model complexity (e.g., Bayesian information criterion) are often unreliable though computationally fast. Unfortunately, theoretically-consistent brute force Monte Carlo requires a very large number of samples and it often fails in high dimensional problems [Brunetti and Linde, 2016]. Other, more efficient methods for evidence estimation are currently not applicable for MPS problems. Evidence estimation without an explicit prior model is a research path worth exploring.

In the presented application, forward modeling errors are not considered. Such errors might be due to grid resolution, numerical errors, the assumption of homogeneous petrophysical parameters or the approximation made in deriving the physical model. In the presented application, the first arrival travel times are simulated using a high-frequency approximation. By considering full waveform modeling results as exempt of modeling error, *Hansen et al.* [2014] showed that high-frequency approximations present a negative bias up to 3 times the magnitude of measurement errors in highly contrasted binary medium. Using full waveform modeling in the inversion algorithm would reduce the modeling error but would strongly increase the computation costs. Presently, it is not quite clear which assumption could be made about modeling errors to estimate and how to best integrate them in the inversion procedure. It offers perspectives for future research.

While this study focuses on 2-D inversion, we believe there is a real potential for extending it to 3D real-case hydrogeological applications. Recently, the graph cut technique has been adapted successfully to produce 3-D geostatistical realizations from 3-D TI while honoring point data conditioning [*Li et al.*, 2016]. The next step is clearly to consider 3-D inversion. Given the computational efficiency of the technique (with respect to other MPS techniques), the expansion to 3-D aquifer characterization data [*Doetsch et al.*, 2010; *Cardiff et al.*, 2013a], the availability of fast forward models [*Podvin and Lecomte*, 1991; *Bakhos et al.*, 2014], the available knowledge and tools to build geologically realistic TIs [*Comunian et al.*, 2012; *Rezaee et al.*, 2015] suggest that the extension of the proposed method to 3-D applications is possible, albeit challenging.

7 Conclusions

The first field application of probabilistic inversion with model proposals based on graph cuts is rather successful, in terms of preservation of geological realism when using an outcrop-derived TI and in terms of data fit and acceptance rate for two types of conceptual models. Several small adaptations of the original graph cut inversion algorithm were needed to ensure high quality MPS cuts, to preserve various classes of geological objects with overlapping properties in the posterior and to improve the posterior realizations. The algorithm works well even in the presence of a small TI. Strong property contrasts or a too wide prior reinforce local exploration by the chains, which slows the convergence of the algorithm. A future venue for future research is to develop a hybrid method combining this approach with slower methods based on iterative geostatistical re-

sampling. Extensions to 3-D applications, full waveform modeling, hydrogeological data and joint inversion will be investigated in the future.

Acknowledgments

The authors would like to thank James Irving for having provided the GPR data set, the reviewers, the editor Xavier Sanchez-Vila and the associate editor for their very constructive remarks. The graph cut algorithm is available at the website of the second author: <http://wp.unil.ch/linde-hydrogeophysics>.

References

- Alcolea, A., and P. Renard (2010), Blocking Moving Window algorithm: Conditioning multiple-point simulations to hydrogeological data, *Water Resources Research*, 46(8), W08511.
- Bakhos, T., M. Cardiff, W. Barrash, and P. K. Kitanidis (2014), Data processing for oscillatory pumping tests, *Journal of Hydrology*, 511, 310–319.
- Barrash, W., and T. Clemo (2002), Hierarchical geostatistics and multifacies systems: Boise Hydrogeophysical Research Site, Boise, Idaho, *Water Resources Research*, 38(10), 1–18.
- Bayer, P., A. Comunian, D. Höyng, and G. Mariethoz (2015), High resolution multi-facies realizations of sedimentary reservoir and aquifer analogs, *Scientific data*, 2, 150033.
- Boykov, Y., O. Veksler, and R. Zabih (2001), Fast approximate energy minimization via graph cuts, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11), 1222–1239.
- Bradford, J. H., W. P. Clement, and W. Barrash (2009), Estimating porosity with ground-penetrating radar reflection tomography: A controlled 3-D experiment at the Boise Hydrogeophysical Research Site, *Water Resources Research*, 45(4), W00D26.
- Brunetti, C., and N. Linde (2016), Geophysics-based model selection of conceptual subsurface models: Application to the South Oyster Bacterial Transport Site, Virginia, USA, in *10th International Geostatistical Congress, Valencia, Spain*.
- Caers, J., and T. Hoffman (2006), The probability perturbation method: a new look at Bayesian inverse modeling, *Mathematical Geology*, 38(1), 81–100.
- Cardiff, M., W. Barrash, and P. K. Kitanidis (2013a), Hydraulic conductivity imaging from 3-D transient hydraulic tomography at several pumping/observation densities, *Water*

- Resources Research*, 49(11), 7311–7326.
- Cardiff, M., T. Bakhos, P. Kitanidis, and W. Barrash (2013b), Aquifer heterogeneity characterization with oscillatory pumping: Sensitivity analysis and imaging potential, *Water Resources Research*, 49(9), 5395–5410.
- Carrera, J., and S. P. Neuman (1986), Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data, *Water Resources Research*, 22(2), 228–242.
- Certes, C., and G. de Marsily (1991), Application of the pilot point method to the identification of aquifer transmissivities, *Advances in Water Resources*, 14(5), 284–300.
- Chugunova, T., and L. Hu (2008), Multiple-point simulations constrained by continuous auxiliary data, *Mathematical Geosciences*, 40(2), 133–146.
- Claerbout, J. F., and F. Muir (1973), Robust modeling with erratic data, *Geophysics*, 38(5), 826–844.
- Comunian, A., P. Renard, and J. Straubhaar (2012), 3D multiple-point statistics simulation using 2D training images, *Computers & Geosciences*, 40, 49–65.
- Dafflon, B., and W. Barrash (2012), Three-dimensional stochastic estimation of porosity distribution: Benefits of using ground-penetrating radar velocity tomograms in simulated-annealing-based or Bayesian sequential simulation approaches, *Water Resources Research*, 48(5), W05553.
- Dafflon, B., J. Irving, and K. Holliger (2009), Simulated-annealing-based conditional simulation for the local-scale characterization of heterogeneous aquifers, *Journal of Applied Geophysics*, 68(1), 60–70.
- Daniels, D. J. (1996), Surface-penetrating radar, *Electronics & Communication Engineering Journal*, 8(4), 165–182.
- Davy, P., and D. Lague (2009), Fluvial erosion/transport equation of landscape evolution models revisited, *Journal of Geophysical Research*, 114(F3), F03007.
- de Vries, L. M., J. Carrera, O. Falivene, O. Gratacós, and L. Slooten (2009), Application of multiple point geostatistics to non-stationary images, *Mathematical Geosciences*, 41(1), 29–42.
- Deutsch, C., and T. Tran (2002), FLUVSIM: a program for object-based stochastic modeling of fluvial depositional systems, *Computers & Geosciences*, 28(4), 525–535.
- Dickson, N. E., J.-C. Comte, P. Renard, J. A. Straubhaar, J. M. McKinley, and U. Ofterdinger (2015), Integrating aerial geophysical data in multiple-point statistics simulations

- p>to assist groundwater flow models,
- Hydrogeology Journal*
- , 23(5), 883–900.
- 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.
- Efendiev, Y., A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallick (2005), An efficient two-stage Markov chain Monte Carlo method for dynamic data integration, *Water Resources Research*, 41(12), W12423.
- Fienen, M. N., T. Clemo, and P. K. Kitanidis (2008), An interactive Bayesian geostatistical inverse protocol for hydraulic tomography, *Water Resources Research*, 44(12), W00B01.
- Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, 7(4), 457–472.
- Gómez-Hernández, J., and X. Wen (1998), To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology, *Advances in Water Resources*, 21(1), 47–61.
- Guardiano, F., and R. Srivastava (1993), Multivariate geostatistics: beyond bivariate moments, in *Geostatistics Troia 1992*, pp. 133–144, Springer Netherlands.
- Hansen, T. M., K. S. Cordua, and K. Mosegaard (2012), Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling, *Computational Geosciences*, 16(3), 593–611.
- Hansen, T. M., K. S. Cordua, B. H. Jacobsen, and K. Mosegaard (2014), Accounting for imperfect forward modeling in geophysical inverse problems—Exemplified for cross-hole tomography, *Geophysics*, 79(3), H1–H21.
- Hendricks Franssen, H., and W. Kinzelbach (2008), Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resources Research*, 44(9), W09408.
- Hendricks Franssen, H., A. Alcolea, M. Riva, M. Bakr, N. Van der Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments, *Advances in Water Resources*, 32(6), 851–872.
- Hermans, T., F. Nguyen, and J. Caers (2015), Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: Workflow and case study, *Water Resources Research*, 51(7), 5332–5352.

- Hinz, E. A., and J. H. Bradford (2010), Ground-penetrating-radar reflection attenuation tomography with an adaptive mesh, *Geophysics*, 75(4), WA251–WA261.
- Hochstetler, D. L., W. Barrash, C. Leven, M. Cardiff, F. Chidichimo, and P. K. Kitanidis (2015), Hydraulic Tomography: Continuity and Discontinuity of High-K and Low-K Zones, *Groundwater*, 54(2), 171–185.
- Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Monthly Weather Review*, 126(3), 796–811.
- Hu, B. X., M. M. Meerschaert, W. Barrash, D. W. Hyndman, C. He, X. Li, and L. Guo (2009), Examining the influence of heterogeneous porosity fields on conservative solute transport, *Journal of Contaminant Hydrology*, 108(3), 77–88.
- Hu, L., and T. Chugunova (2008), Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review, *Water Resources Research*, 44(11), W11,413.
- Hu, L. Y. (2000), Gradual deformation and iterative calibration of Gaussian-related stochastic models, *Mathematical Geology*, 32(1), 87–108.
- Irving, J. D., M. D. Knoll, and R. J. Knight (2007), Improving crosshole radar velocity tomograms: A new approach to incorporating high-angle traveltimes data, *Geophysics*, 72(4), J31–J41.
- Journel, A., and T. Zhang (2006), The necessity of a multiple-point prior model, *Mathematical Geology*, 38(5), 591–610.
- Kass, R. E., and A. E. Raftery (1995), Bayes factors, *Journal of the American Statistical Association*, 90(430), 773–795.
- Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resources Research*, 31(10), 2411–2419.
- Kitanidis, P. K., and E. G. Vomvoris (1983), A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resources Research*, 19(3), 677–690.
- Laloy, E., and J. A. Vrugt (2012), High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing, *Water Resources Research*, 48(1), W01526.
- Laloy, E., N. Linde, D. Jacques, and J. A. Vrugt (2015), Probabilistic inference of multi-Gaussian fields from indirect hydrological data using circulant embedding and dimensionality reduction, *Water Resources Research*, 51(6), 4224–4243.

- Laloy, E., N. Linde, D. Jacques, and G. Mariethoz (2016), Merging parallel tempering with sequential geostatistical resampling for improved posterior exploration of high-dimensional subsurface categorical fields, *Advances in Water Resources*, 90, 57–69.
- Lesmes, D. P., and S. P. Friedman (2005), Relationships between the Electrical and Hydrogeological Properties of Rocks and Soils, in *Hydrogeophysics*, edited by Y. Rubin and S. S. Hubbard, pp. 87–128, Springer.
- Li, X., G. Mariethoz, D. Lu, and N. Linde (2016), Patch-based iterative conditional geostatistical simulation using graph cuts, *Water Resources Research*, 52(8), 6297–6320.
- Linde, N., P. Renard, T. Mukerji, and J. Caers (2015a), Geological realism in hydrogeological and geophysical inverse modeling: A review, *Advances in Water Resources*, 86, 86–101.
- Linde, N., T. Lochbühler, M. Dogan, and R. L. Van Dam (2015b), Tomogram-based comparison of geostatistical models: Application to the Macrodispersion Experiment (MADE) site, *Journal of Hydrology*, 531, 543–556.
- Lochbühler, T., G. Pirot, J. Straubhaar, and N. Linde (2013), Conditioning of multiple-point statistics facies simulations to tomographic images, *Mathematical Geosciences*, 46(5), 625–645.
- Mahmud, K., G. Mariethoz, J. Caers, P. Tahmasebi, and A. Baker (2014), Simulation of Earth textures by conditional image quilting, *Water Resources Research*, 50(4), 3088–3107.
- Mariethoz, G., and J. Caers (2014), *Multiple-Point Geostatistics: Stochastic Modeling With Training Images*, John Wiley & Sons.
- Mariethoz, G., and S. Lefebvre (2014), Bridges between multiple-point geostatistics and texture synthesis: Review and guidelines for future research, *Computers & Geosciences*, 66, 66–80.
- Mariethoz, G., P. Renard, and J. Caers (2010a), Bayesian inverse problem and optimization with iterative spatial resampling, *Water Resources Research*, 46(11), W11530.
- Mariethoz, G., P. Renard, and J. Straubhaar (2010b), The Direct Sampling method to perform multiple-point geostatistical simulations, *Water Resources Research*, 46, W11536.
- Mariethoz, G., A. Comunian, I. Irarrazaval, and P. Renard (2014), Analog-based meandering channel simulation, *Water Resources Research*, 50, 836–854.
- Metropolis, N., and S. Ulam (1949), The Monte Carlo method, *Journal of the American Statistical Association*, 44(247), 335–341.

- Moret, G. J., M. D. Knoll, W. Barrash, and W. P. Clement (2006), Investigating the stratigraphy of an alluvial aquifer using crosswell seismic traveltimes tomography, *Geophysics*, 71(3), B63–B73.
- Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *Journal of Geophysical Research*, 100(B7), 12,431–12,447.
- Peterson, J. E., Jr (2001), Pre-inversion corrections and analysis of radar tomographic data, *Journal of Environmental & Engineering Geophysics*, 6(1), 1–18.
- Pirot, G., J. Straubhaar, and P. Renard (2014), Simulation of braided river elevation model time series with multiple-point statistics, *Geomorphology*, 214, 148–156.
- Pirot, G., J. Straubhaar, and P. Renard (2015a), A pseudo genetic model of coarse braided-river deposits, *Water Resources Research*, 51(12), 9595–9611.
- Pirot, G., P. Renard, E. Huber, J. Straubhaar, and P. Huggenberger (2015b), Influence of conceptual model uncertainty on contaminant transport forecasting in braided river aquifers, *Journal of Hydrology*, 531, 124–141.
- Podvin, P., and I. Lecomte (1991), Finite difference computation of traveltimes in very contrasted velocity models: A massively parallel approach and its associated tools, *Geophysical Journal International*, 105(1), 271–284.
- Pride, S. (1994), Governing equations for the coupled electromagnetics and acoustics of porous media, *Physical Review B*, 50(21), 15,678–15,696.
- Pyrz, M. J., J. B. Boisvert, and C. V. Deutsch (2009), ALLUVSIM: A program for event-based stochastic modeling of fluvial depositional systems, *Computers & Geosciences*, 35(8), 1671–1685.
- Rabinovich, A., W. Barrash, M. Cardiff, D. L. Hochstetler, T. Bakhos, G. Dagan, and P. K. Kitanidis (2015), Frequency dependent hydraulic properties estimated from oscillatory pumping tests in an unconfined aquifer, *Journal of Hydrology*, 531, 2–16.
- RamaRao, B. S., A. M. LaVenue, G. De Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments, *Water Resources Research*, 31(3), 475–493.
- Renard, P., and G. Mariethoz (2014), Special issue on 20 years of multiple-point statistics: part 1, *Mathematical Geosciences*, 46(2), 129.
- Rezaee, H., G. Mariethoz, M. Koneshloo, and O. Asghari (2013), Multiple-point geostatistical simulation using the bunch-pasting direct sampling method, *Computers & Geosciences*

sciences, 54, 293–308.

Rezaee, H., D. Marcotte, P. Tahmasebi, and A. Saucier (2015), Multiple-point geostatistical simulation using enriched pattern databases, *Stochastic Environmental Research and Risk Assessment*, 29(3), 893–913.

Robert, C., and G. Casella (2013), *Monte Carlo Statistical Methods*, Springer Science & Business Media.

Roberts, G. O., J. S. Rosenthal, et al. (2001), Optimal scaling for various Metropolis-Hastings algorithms, *Statistical Science*, 16(4), 351–367.

Ruggeri, P., J. Irving, and K. Holliger (2015), Systematic evaluation of sequential geostatistical resampling within MCMC for posterior sampling of near-surface geophysical inverse problems, *Geophysical Journal International*, 202(2), 961–975.

Sambridge, M., and K. Mosegaard (2002), Monte Carlo methods in geophysical inverse problems, *Reviews of Geophysics*, 40(3), 1–29.

Schlüter, S., and H.-J. Vogel (2011), On the reconstruction of structural and functional properties in random heterogeneous media, *Advances in Water Resources*, 34(2), 314–325.

Schöniger, A., T. Wöhling, L. Samaniego, and W. Nowak (2014), Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence, *Water Resources Research*, 50(12), 9484–9513.

Straubhaar, J., P. Renard, and G. Mariethoz (2016), Conditioning multiple-point statistics simulations to block data, *Spatial Statistics*, 16, 53–71.

Strebelle, S. (2002), Conditional simulation of complex geological structures using multiple-point statistics, *Mathematical Geology*, 34(1), 1–21.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, Siam, Philadelphia, USA.

Tarantola, A., and B. Valette (1982), Inverse problems= quest for information, *Geophysics*, 50(3), 150–170.

Thoma, M. J., W. Barrash, M. Cardiff, J. Bradford, and J. Mead (2014), Estimating unsaturated hydraulic functions for coarse sediment from a field-scale infiltration experiment, *Vadose Zone Journal*, 13(3), 1–17.

Tronicke, J., K. Holliger, W. Barrash, and M. D. Knoll (2004), Multivariate analysis of cross-hole georadar velocity and attenuation tomograms for aquifer zonation, *Water Resources Research*, 40(1), W01519.

- Vrugt, J. A. (2016), Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation, *Environmental Modelling & Software*, 75, 273–316.
- Vrugt, J. A., C. J. Ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resources Research*, 44(12), W00B09.
- Vrugt, J. A., C. Ter Braak, C. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3), 273–290.
- Xu, T., L. White, D. Hui, and Y. Luo (2006), Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction, *Global Biogeochemical Cycles*, 20(2), GB2007.
- Zahner, T., T. Lochbühler, G. Mariethoz, and N. Linde (2016), Image synthesis with graph cuts: A fast model proposal mechanism in probabilistic inversion, *Geophysical Journal International*, 204(2), 1179–1190.
- Zhou, H., J. J. Gómez-Hernández, H.-J. H. Franssen, and L. Li (2011), An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering, *Advances in Water Resources*, 34(7), 844–864.
- Zimmerman, D., G. d. Marsily, C. Gotway, M. Marietta, C. Axness, R. Beauheim, R. Bras, J. Carrera, G. Dagan, P. Davies, et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resources Research*, 34(6), 1373–1413.
- Zinn, B., and C. F. Harvey (2003), When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields, *Water Resources Research*, 39(3), 1051.

A: Suggested approach to calibrate the cutting threshold parameter

To define an appropriate calibration function to determine δ_{th} , we rely on three normalized indicators computed for various values of δ_{th} from MCMC chains of length 10^5 (Figure A.1). From the mean log-likelihood \overline{llh} is computed the first indicator: the normalized mean log-likelihood $\overline{llh}_{[0,1]} = \frac{\overline{llh} - \min(\overline{llh})}{\max(\overline{llh}) - \min(\overline{llh})}$ (Figure A.1 A and B). The second is the normalized update size conformity indicator usc (Figure A.1 C and D) that re-

Figure A.1: Calibration function for the cutting threshold δ_{th} based on MCMC chains of length 10^5 using the multi-Gaussian TI (A), C), E) and G)) and the outcrop-derived TI (B), D), F) and H)); dashed lines represent normalized indicators; A) and B) original and normalized mean log-likelihood as a function of δ_{th} ; C) and D) mean update size and normalized criteria related to the mean update size conformity as a function of δ_{th} ; E) and F) acceptance rate as a function of δ_{th} ; G) and H) score of the calibration criteria as a function of δ_{th} .

flects targeting a mean update size of the porosity field around $20 \pm 10\%$ of the model size. It is computed as $usc = \frac{1-|0.2-\overline{usp}|}{0.2} \times \mathbf{I}_{\overline{usp} \leq 0.3} \times \mathbf{I}_{\overline{usp} \geq 0.1}$, where \overline{usp} is the mean update size expressed as a percentage of the model size and \mathbf{I} is the indicator function.

The third is the the normalized acceptance rate $\alpha_{[0,1]}$ (Figure A.1 E and F), computed as $\alpha_{[0,1]} = \frac{\min(\alpha, 0.25) - \min(\min(\alpha, 0.25))}{\max(\min(\alpha, 0.25)) - \min(\min(\alpha, 0.25))}$, where α denotes the acceptance rate computed over the second half of the 10^5 length chains. Acceptance rate over 25% are penalized because experience showed it did not favor convergence. The calibration function (score) is the product of these three normalized indicators (Figure A.1G and H). The δ_{th} used in subsequent inversions is defined as the maximum score and the retained maximal value.

Figure 1.

Figure 2.

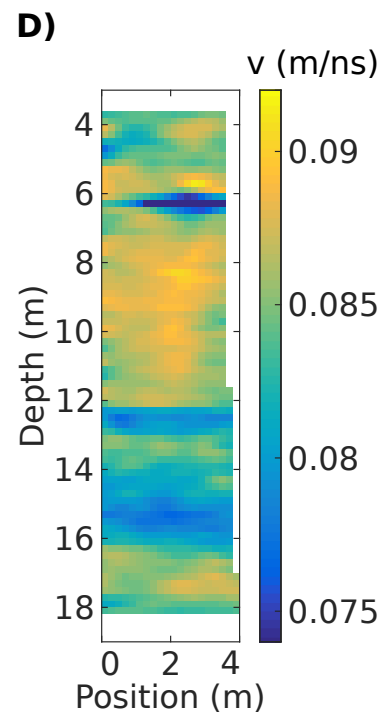
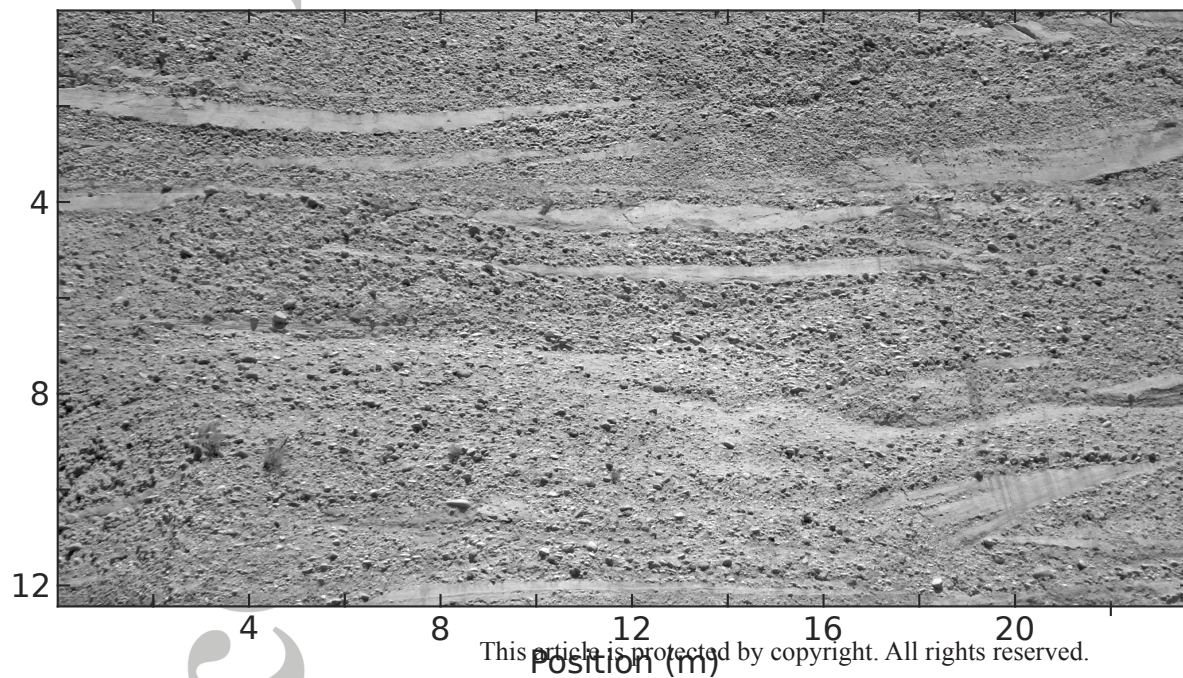
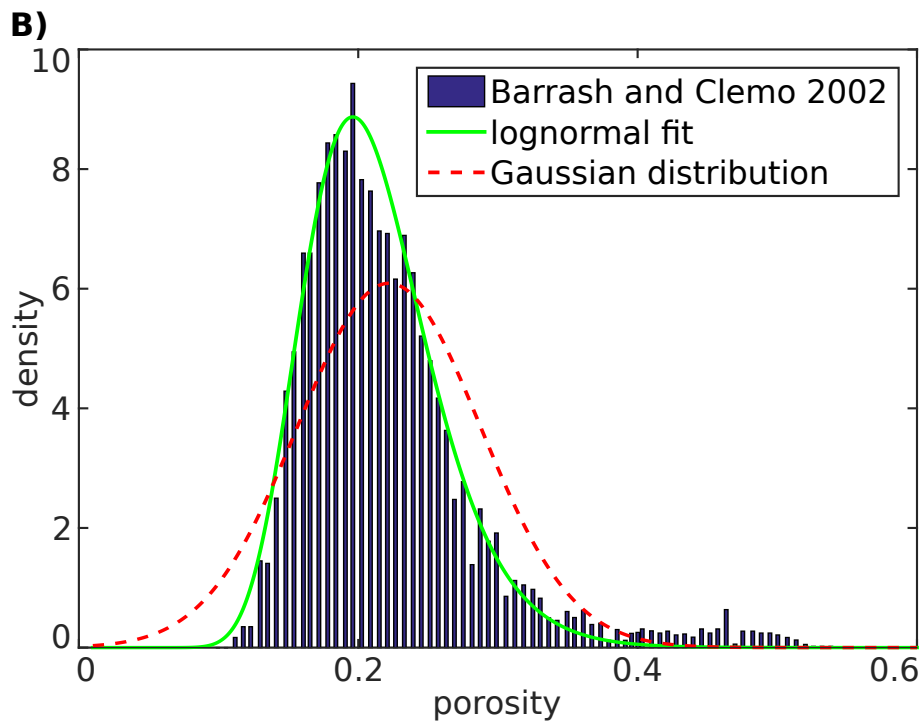
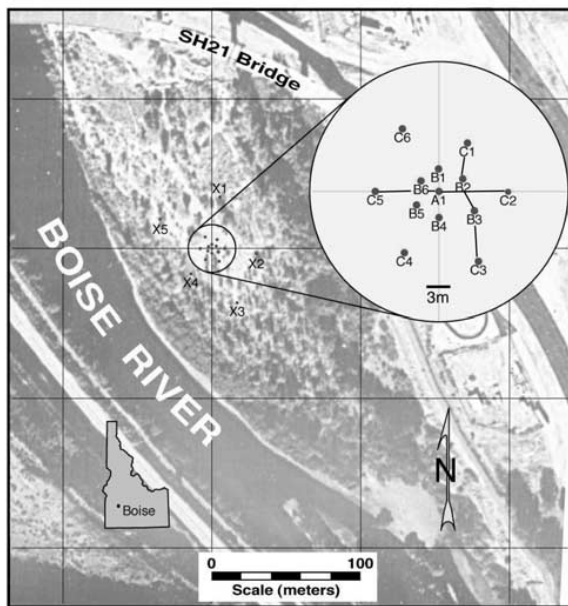


Figure 3.

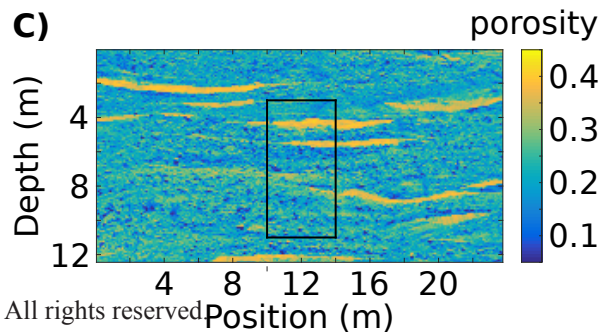
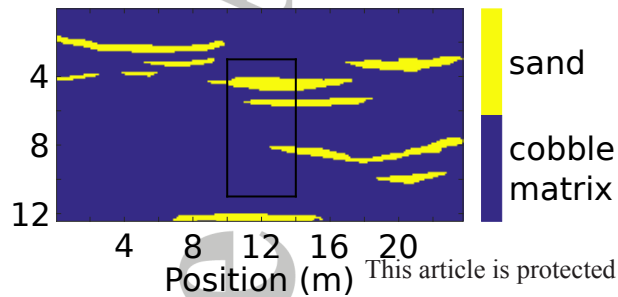
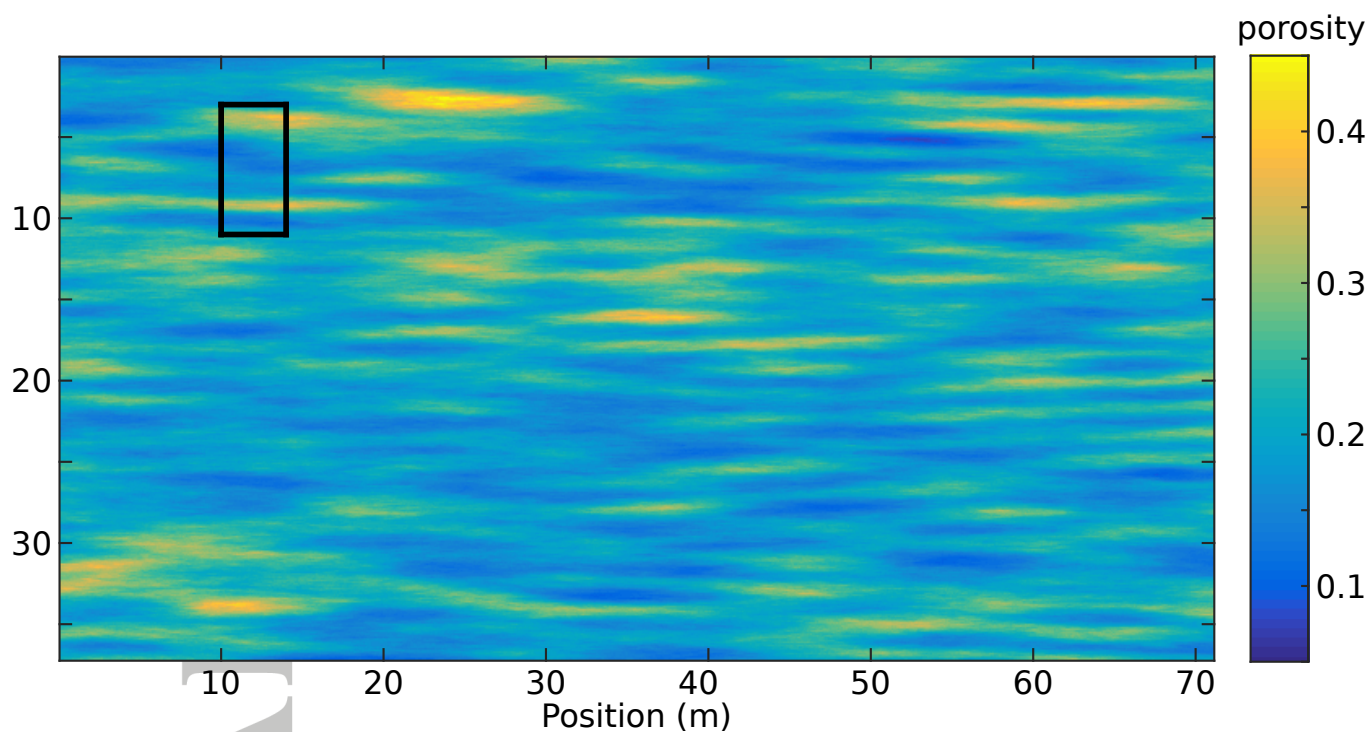


Figure 4.

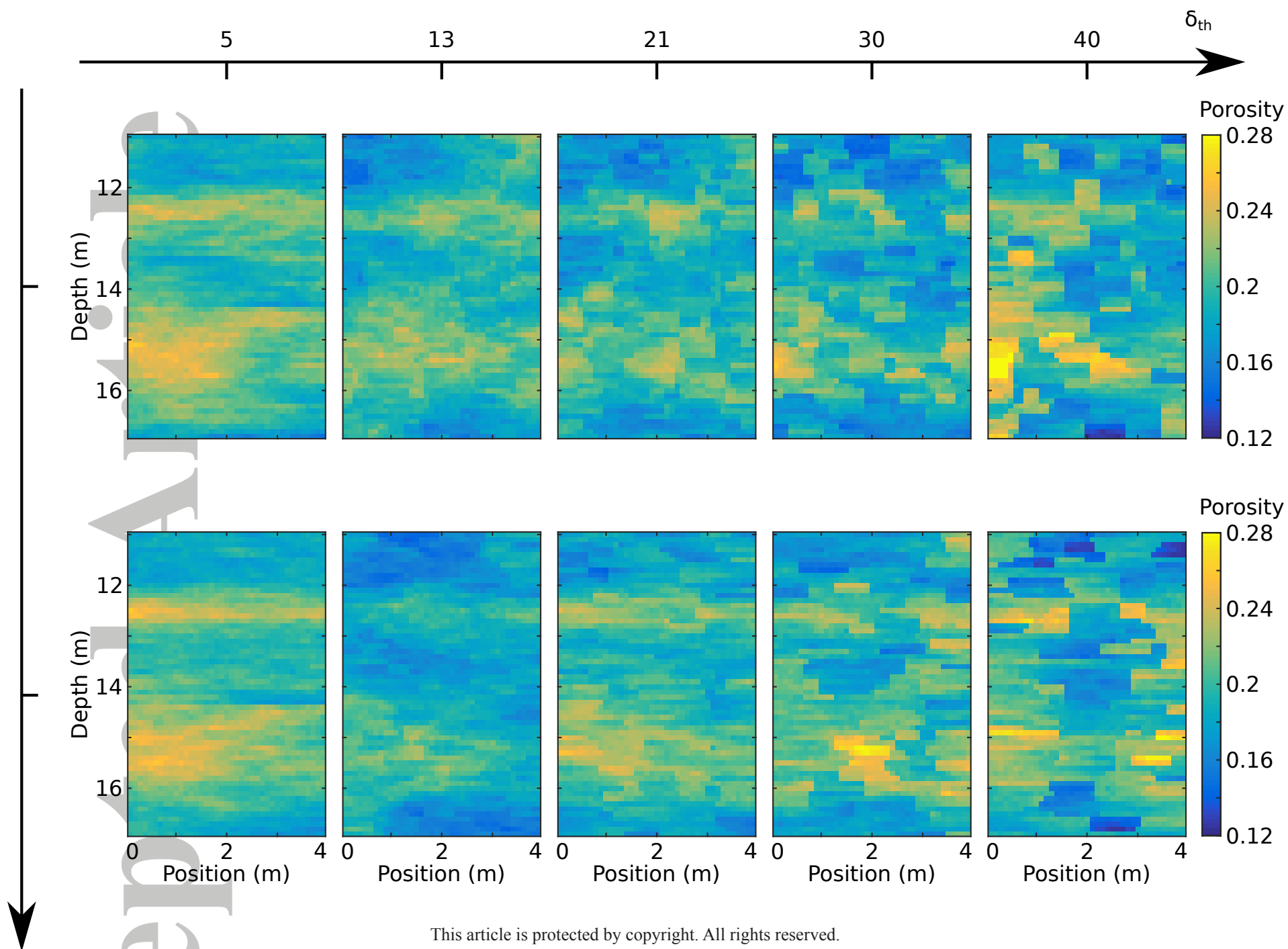
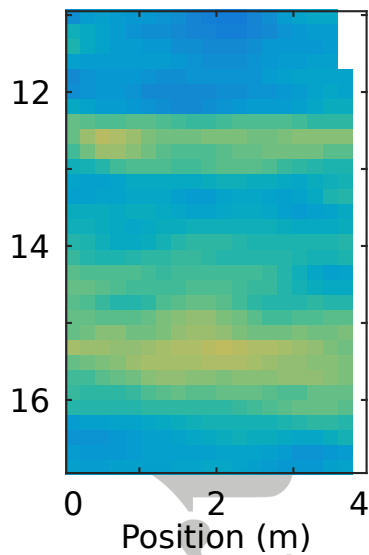
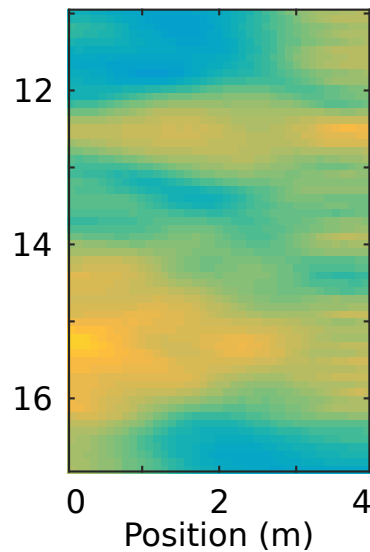
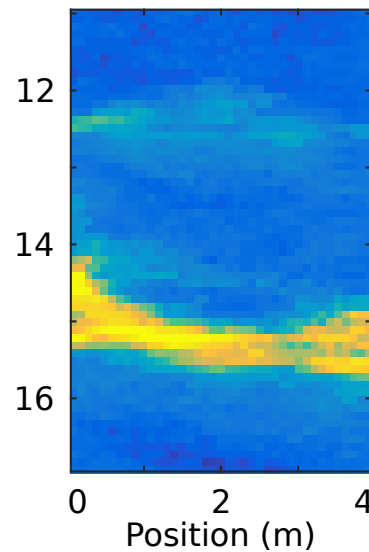


Figure 5.

A)**B)****C)**

Porosity

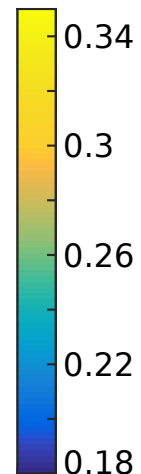
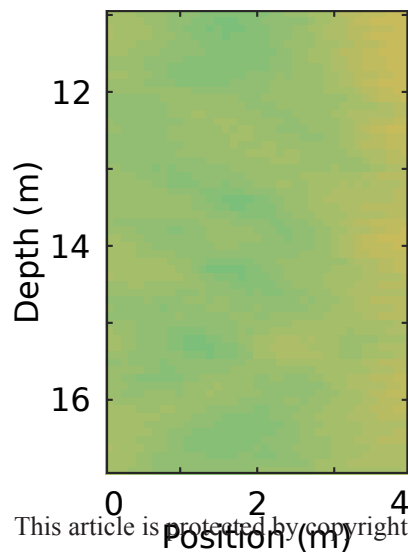
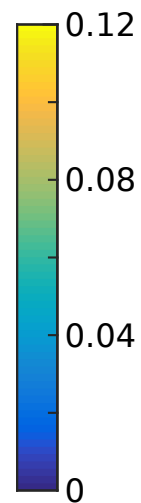
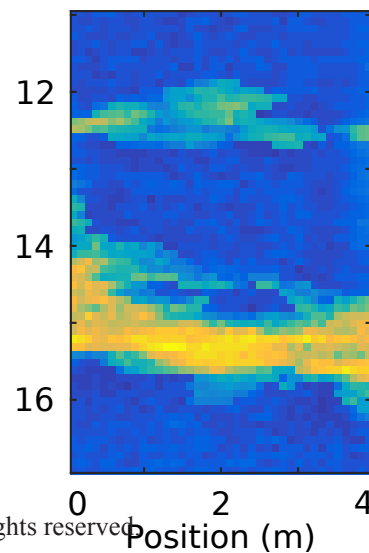
**D)****E)**

Figure 6.

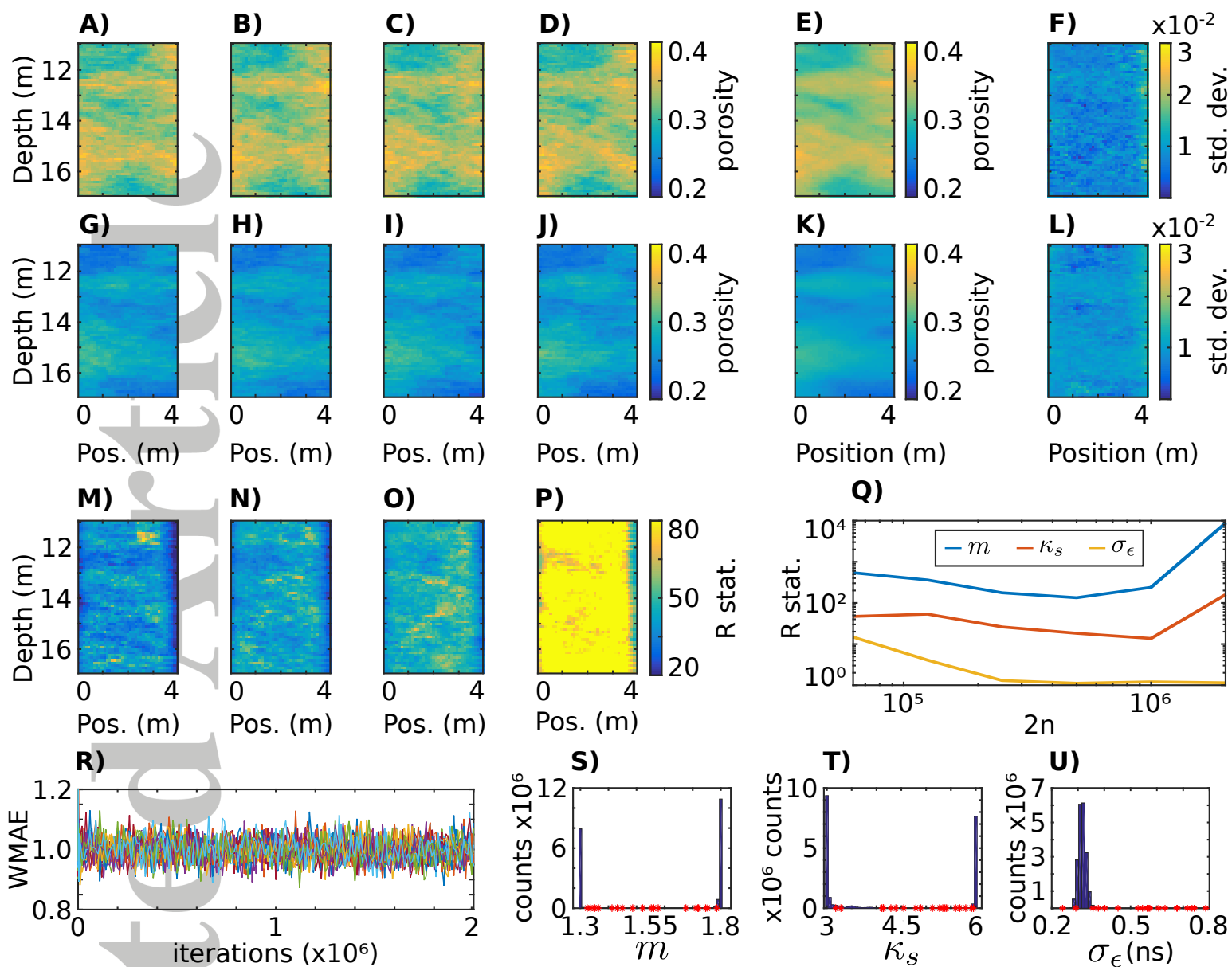


Figure 7.

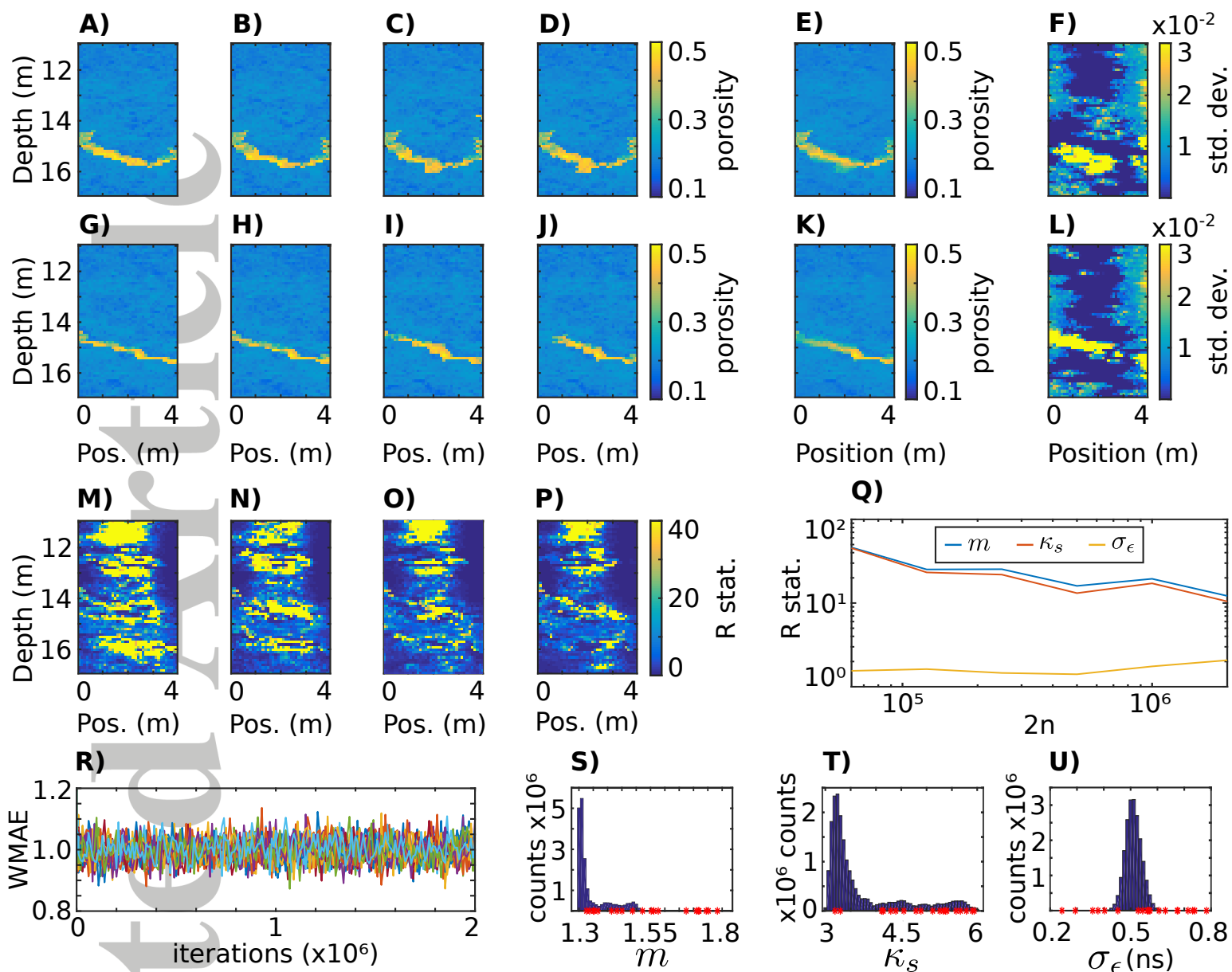


Figure 8.

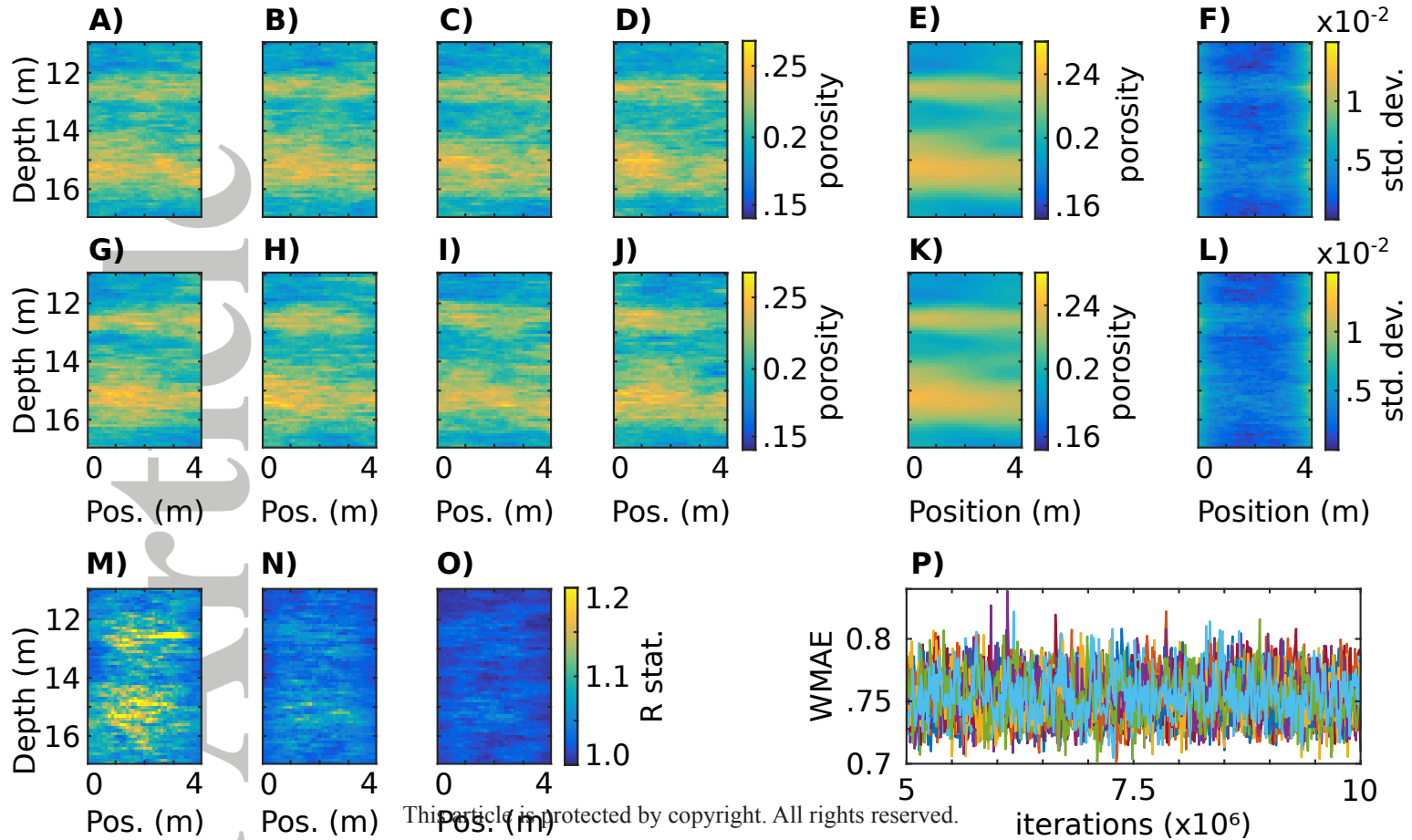


Figure 9.

