Probabilistic inference of multi-Gaussian fields from indirect hydrological data using circulant embedding and dimensionality reduction

Eric Laloy,[1,3] Niklas Linde,[2], Diederik Jacques[1] and Jasper A. Vrugt[3,4]

Eric Laloy, Institute for Environment, Health and Safety, Belgian Nuclear Research Centre, Mol, Belgium. (elaloy@sckcen.be)

Niklas Linde, Applied and Environmental Geophysics Group, Institute of Earth Sciences, University of Lausanne, Lausanne, Switzerland.

Diederik Jacques, Institute for Environment, Health and Safety, Belgian Nuclear Research Centre, Mol, Belgium.

Jasper A. Vrugt, Department of Civil and Environmental Engineering, University of California Irvine, Irvine, CA, USA.

[1]Institute for Environment, Health and

Abstract. We present a Bayesian inversion method for the joint inference of high-dimensional multi-Gaussian hydraulic conductivity fields and associated geostatistical parameters from indirect hydrological data. We combine Gaussian process generation via circulant embedding to decouple the variogram from grid cell specific values, with dimensionality reduction by interpolation to enable Markov chain Monte Carlo (MCMC) simulation. Using the Matérn variogram model, this formulation allows inferring the conductivity values simultaneously with the field smoothness (also called Matérn shape parameter) and other geostatistical parameters such as the mean, sill, integral scales and anisotropy direction(s) and ratio(s). The proposed dimensionality reduction method systematically honors the underlying variogram

Safety, Belgian Nuclear Research Centre,
Mol, Belgium.

[2]Applied and Environmental Geophysics
Group, Institute of Earth Sciences,
University of Lausanne, Lausanne,
Switzerland.

[3]Department of Civil and Environmental
Engineering, University of California Irvine,
Irvine, USA.

[4]Department of Earth Systems Science,
University of California Irvine, Irvine, USA.

and is demonstrated to achieve better performance than the Karhunen-Loève expansion. We illustrate our inversion approach using synthetic (error corrupted) data from a tracer experiment in a fairly heterogeneous 10,000-dimensional 2D conductivity field. A 40-times reduction of the size of the parameter space did not prevent the posterior simulations to appropriately fit the measurement data and the posterior parameter distributions to include the true geostatistical parameter values. Overall, the posterior field realizations covered a wide range of geostatistical models, questioning the common practice of assuming a fixed variogram prior to inference of the hydraulic conductivity values. Our method is shown to be more efficient than sequential Gibbs sampling (SGS) for the considered case study, particularly when implemented on a distributed computing cluster. It is also found to outperform the method of anchored distributions (MAD) for the same computational budget.

1. Introduction

High-parameter dimensionality poses considerable challenges for the inversion of groundwater flow and transport data [e.g., *Kitanidis*, 1995; *Hendricks-Franssen et al.*, 2009; *Laloy et al.*, 2013; *Zhou et al.*, 2014, and references therein]. What is more, conceptual and structural inadequacies of the subsurface model and measurement errors of the model input (boundary conditions) and output (calibration) data introduce uncertainty in the estimated parameters and model simulations. Another important source of uncertainty originates from sparse data coverages that rarely contain sufficient information to uniquely characterize the subsurface at a spatial resolution deemed necessary for accurate modeling. This results in an ill-posed inverse problem with many different sets of model parameter values that fit the data acceptably well. Inversion methods should consider this inherent uncertainty and provide an ensemble of model realizations that accurately span the range of possible models that honor the available calibration data and prior information.

Hydraulic conductivity (K) fields are typically assumed to be stationary and log-normally distributed with a spatial structure determined by a two-point geostatistical model or variogram [e.g., *Rubin*, 2003]. Unfortunately, a lack of (sufficient) point K measurements (if any) makes it difficult to estimate directly the geostatistical parameters (mean, sill, variogram model, integral scales and anisotropy factors) from variographic analysis [*Ortiz-Deutsch*, 2002; *Nowak et al.*, 2010]. Simultaneous (inverse) inference of conductivity values and associated geostatistical parameters is therefore attractive yet computationally challenging. Indeed, only a few studies can be found in the literature that have attempted simultaneous estimation using global and probabilistic search meth-

ods. For example, *Jafarpour and Tarrahi* [2011] used the Ensemble Kalman Filter [EnKF, *Evensen*, 2003] to estimate jointly the conductivity field and associated geostatistical properties. However, their attempt was not particularly successful, and they attributed this lack of success to a complex and nonunique relationship between the parameters of interest and available flow data. On the contrary, *Jardani et al.* [2012] used Bayesian inversion and backed out successfully the transmissivities at 72 pilot-points together with 45 leakage coefficients, and the sill, and correlation range of a 2D spherical isotropic variogram. However, their case study was made relatively simple by assuming a fairly smooth multi-Gaussian field with narrow ranges for the two unknown variogram parameters. *Fu and Gómez-Hernández* [2009] introduced a Markov chain Monte Carlo (MCMC) simulation method that iteratively produces local perturbations of a multi-Gaussian parameter field by resimulation of only a fraction of the field at each realization. This approach cannot directly cope with variogram uncertainty, but *Hansen et al.* [2012] and *Hansen et al.* [2013a, b] proposed methodological extensions to enable joint inference of the values of a multi-Gaussian field and its associated variogram properties. Their method, known as sequential Gibbs sampling (SGS), was applied to the joint estimation of a 2D velocity field and corresponding correlation lengths using crosshole ground penetrating radar tomography data [*Hansen et al.*, 2013b]. However, the variogram inference only concerned two parameters (i.e., the range in two different directions), and the inverse solution was stabilized by using a Gaussian prior. Furthermore, variogram inference with SGS simply involves sampling of the prior variogram distribution, which may not be efficient if the prior variogram uncertainty range is large and/or the information content of the calibration data is high. Lastly, the method of anchored distributions (MAD) introduced by

Zhang and Rubin [2008] and *Rubin et al.* [2010] simultaneously derives variogram parameters and hydraulic conductivity/transmissivity values at selected locations (the so called "anchors") within the hydrogeologic domain of interest. In the examples presented by *Rubin et al.* [2010] and *Murakami et al.* [2010], the estimated geostatistical properties are restricted to the mean, sill and range of a 2D exponential isotropic model. Perhaps more importantly, the MAD framework relies on plain Monte Carlo (MC) simulation and is thus computationally very demanding, particularly in high-dimensional parameter spaces [e.g., *Murakami et al.*, 2010]. Indeed, MC simulation is very inefficient if the prior distribution is large with respect to the size of the posterior distribution.

Dimensionality reduction of the parameter space can help solving high-dimensional inverse problems. This includes methods such as the MAD technique described above or the Karhunen-Loève (KL) transform [*Loève*, 1977]. The latter is widely used in subsurface hydrology to represent multi-Gaussian parameter fields. The KL transform uses the covariance function to describe a spatially Gaussian process in a reduced basis [e.g., *Zhang and Zu*, 2004; *Li and Cirpka*, 2006; *Laloy et al.*, 2013]. The base functions are the eigenfunctions of the covariance function multiplied by the square root of the associated eigenvalues. The sorted eigenvalues and corresponding eigenfunctions can then be truncated, thereby leading to a reduced parameter space if the number of dominant eigenvalues is smaller than the number of simulation grid points. Overall, the smoother the covariance kernel the larger the parameter reduction. Hence, the KL expansion cannot cope efficiently with rough random fields and/or short integral scales. In such cases, the number of base functions needed for accurate field reconstruction may approach the size of the original discretized domain. Another difficulty arises from the fact that numerical

estimation of the required eigenfunctions and eigenvalues of the considered covariance kernel can be CPU-demanding for large grids [though different kernel-specific solutions exist for speeding up efficiency, see, e.g, *Zhang and Zu*, 2004; *Li and Cirpka*, 2006].

Here we present a novel Bayesian inversion approach for the simultaneous estimation of high-dimensional hydraulic conductivity fields and associated two-point geostatistical properties from indirect hydrologic data. Our method uses Gaussian process generation via circulant embedding [*Dietrich and Newsam*, 1997] to decouple the variogram from grid cell specific values, and implements dimensionality reduction by interpolation to enable MCMC simulation with the DREAM$_{(ZS)}$ algorithm [*Vrugt et al.*, 2009; *Laloy and Vrugt*, 2012]. We use the Matérn function to infer the conductivity values jointly with the field smoothness and other geostatistical parameters (mean, sill, integral scales, anisotropy direction(s) and anisotropy ratio(s)). Moreover, conditioning on direct point conductivity measurements (if any) is straightforward. We illustrate our method using synthetic, error corrupted, data from a tracer simulation experiment involving a 10,000 dimensional 2D conductivity field.

This paper is organized as follows. Section 2 presents the different elements of our inversion approach, and demonstrates the merits of the proposed dimensionality reduction method by comparison against the widely used KL transform. This is followed in section 3 with numerical experiments involving a fairly heterogeneous reference field. Our numerical experiments involve benchmark tests against standard implementations of the state-of-the-art SGS and MAD techniques. Section 4 takes a closer look at the required CPU-time and provides further analysis of the performance of our method. Finally, section 5 concludes this paper with a summary of the most important findings.

2. Methods

For ease of understanding, we proceed first by describing the general circulant embedding technique in section 2.1. Next, section 2.2 details our proposed dimensionality reduction approach.This step is of utmost importance as it reduces significantly the dimensionality of the parameter space used by the circulant embedding, thereby enabling Bayesian inference via MCMC simulation. Section 2.3 then compares for different levels of dimensionality reduction the proposed approach against the KL transform, before sections 2.4, 2.5 and 2.6 present the other ingredients of our inversion approach.

2.1. Stationary Gaussian Process Generation via Circulant Embedding

We use the stationary Gaussian process generation through circulant embedding of the covariance matrix proposed by *Dietrich and Newsam* [1997] [see also the excellent review by *Kroese and Botev*, 2013]. We provide herein a short description of this methodology for a 2D domain but extension to a 3D grid is straightforward. Further details on the method can be found in the cited references.

Multi-Gaussian field generation over a regularly meshed 2D grid can be implemented as follows. If $\mathbf{Y}_0$, the zero-mean stationary Gaussian field to be generated, is of size $m \times n$ then:

1. Build a $(2m-1) \times (2n-1)$ symmetric covariance matrix $\mathbf{S}$ for a given covariance kernel $\mathbf{C}$. The entries of $\mathbf{S}$ are the 2-point covariances between any point in the $(2m-1) \times (2n-1)$ domain and the domain center point. The dimensions $(2m-1) \times (2n-1)$ represent the minimal length for which a symmetric nonnegative definite circulant matrix $\mathbf{S}$ can be found.

2. Compute the $\mathbf{\Omega}$ matrix of eigenvalues as

$$\mathbf{\Omega} = \frac{\text{real}\left[\text{FFT2}\left(\text{FFTSHIFT}\left\{\mathbf{S}\right\}\right)\right]}{(2m-1)\times(2n-1)}, \tag{1}$$

where real[ ] denotes the real part, FFT2 signifies the two-dimensional fast Fourier transform, and FFTSHIFT is a function that swaps the first quadrant of a matrix with the third and the second quadrant with the fourth.

3. Make sure that all elements $\Omega_{i,j}$ where $i = 1 \cdots 2m-1$ and $j = 1 \cdots 2n-1$ are greater than zero (nonnegative embedding). Negative eigenvalues might appear when the integral scale of $\mathbf{C}$ becomes large compared to the domain size. For instance, *Dietrich and Newsam* [1997] showed that for a two-dimensional domain of size $m \times m$ the minimum ratios of the domain side length to the integral scale for which the embedding is non-negative are about 5.6 and 4.5 respectively, for Gaussian and exponential covariance models. In this work, we simply assume that the search range of the integral scale is bounded such that the maximum number of negative eigenvalues in $\mathbf{\Omega}$ is kept reasonably small, and negative values are set to zero if they occur.

4. Generate two $(2m-1)\times(2n-1)$ matrices with standard normal variates, say $\mathbf{Z}_1$ and $\mathbf{Z}_2$, and construct the complex Gaussian matrix

$$\underline{\mathbf{Z}} = \mathbf{Z}_1 + i\mathbf{Z}_2, \tag{2}$$

with $i = \sqrt{-1}$. Finally, two independent zero-mean stationary Gaussian realizations $\mathbf{Y}_0$ with prescribed covariance kernel $\mathbf{C}$ are given by the first $m \times n$ elements of real[$\mathbf{f}$] and imag[$\mathbf{f}$] where imag[ ] signifies the imaginary part and

$$\mathbf{f} = \text{FFT2}\left(\sqrt{\mathbf{\Omega}} \odot \underline{\mathbf{Z}}\right), \tag{3}$$

where $\odot$ denotes component-wise multiplication. The complex standard normal matrix $\underline{\mathbf{Z}}$ represents the uncorrelated noise component of the Gaussian field, decoupled from its covariance structure embedded into $\mathbf{\Omega}$.

The use of circulant embedding for generating multi-Gaussian fields (section 2.1) has several attractive features, one of which is that it enables joint inference by fast (de)coupling of the grid cell random numbers from their geostatistical properties. The computational complexity (cost) of circulant embedding is similar to that of the FFT method. The computational complexity of the FFT method is of order $\mathcal{O}(n \log(n))$ for a symmetric positive definite $n \times n$ covariance matrix, which compares very favorably to the computational complexity of $\mathcal{O}(n^3)$ for Cholesky decomposition. Yet statistical inference becomes more and more difficult with increasing dimensionality of the search space. This justifies the dimensionality reduction approach proposed in section 2.2.

Lastly, it is worth noting that an alternative FFT-based decoupling method was proposed by *Le Ravalec et al.* [2000]. Their FFT moving average (or FFT-MA) approach has been developed independently from the standard technique by *Dietrich and Newsam* [1997], though it also makes use of FFT and the circulant embedding property of covariance matrices. The FFT-MA generator is used as a basic building block by *Hansen et al.* [2013a, b] for joint inversion of field properties and variogram parameters. Yet our experience with FFT-MA suggests the need for a larger embedding domain than for the standard approach to produce consistent multi-Gaussian fields with long integral scales. Coupling FFT-MA with dimensionality reduction also led to reconstructed models that are less accurate than those derived from coupling of dimensionality reduction with the classical circulant embedding approach (see also section 2.3).

2.2. Dimensionality Reduction by One-dimensional FFT Interpolation

We take advantage of one-dimensional interpolation to significantly reduce the dimensionality of the parameter space. More specifically, we employ one-dimensional interpolation by the FFT method to resample two lower-dimensional vectors of standard normal variates, say $\mathbf{r}_1$ and $\mathbf{r}_2$, to two sets of $(2m-1)(2n-1)$ equally spaced points, hereafter referred to as $\mathbf{z}_1$ and $\mathbf{z}_2$. The reconstructed vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ can then be reshaped into the matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ (see equation (2)). In this work, $\mathbf{r}_1$ and $\mathbf{r}_2$ have a dimensionality that is one to two orders of magnitude lower than that of $\mathbf{z}_1$ and $\mathbf{z}_2$.

The original low-dimensional $\mathbf{r}_1$ and $\mathbf{r}_2$ vectors are thus transformed to the Fourier domain using FFT and then transformed back with $(2m-1)(2n-1)$ points to produce $\mathbf{z}_1$ and $\mathbf{z}_2$, respectively. We deliberately chose FFT over other methods such as linear interpolation, as this approach was found to better preserve, during resampling, the unit variance of the standard normal distribution. In other words, if $\mathbf{r}_1$ and $\mathbf{r}_2$ are $\sim \mathcal{N}(\mathbf{0}, \mathbf{1})$, then the variances of $\mathbf{z}_1$ and $\mathbf{z}_2$ were found to be closer to unity when one-dimensional interpolation is performed by FFT. In contrast, linear interpolation generally led to a variance reduction.

To eliminate short lag autocorrelation, the elements of $\mathbf{z}_1$ and $\mathbf{z}_2$ are permuted randomly after interpolation from $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. This permutation step is necessary as the circulant embedding method breaks down if neighboring values of $\mathbf{z}_1$ and $\mathbf{z}_2$ are correlated. Therefore, we use pre-selected permutation schemes to independently permute the elements of $\mathbf{z}_1$ and $\mathbf{z}_2$. Of the two field realizations real[$\mathbf{f}$] and imag[$\mathbf{f}$] produced by equation (3), we only use real[$\mathbf{f}$] herein. This is because, for inference, each $\mathbf{r} = [\mathbf{r}_1, \mathbf{r}_2]$

vector must be associated with a single multi-Gaussian field and corresponding simulated data set.

Our dimensionality reduction approach can thus be briefly summarized as follows:

1. Perform circulant (periodic) embedding of the covariance function at the desired resolution. This gives the circulant matrix $\mathbf{S}$.

2. Fourier transform $\mathbf{S}$ to obtain the matrix of eigenvalues $\mathbf{\Omega}$.

3. Generate two low-dimensional vector of real-valued (standard normal) random numbers, $\mathbf{r}_1$ and $\mathbf{r}_2$. The larger the dimensionalty reduction, the smaller the selected sizes of $\mathbf{r}_1$ and $\mathbf{r}_2$.

4. Using one-dimensional FFT interpolation, resample $\mathbf{r}_1$ and $\mathbf{r}_2$ to two vectors of real-valued random numbers of the right size, $\mathbf{z}_1$ and $\mathbf{z}_2$.

5. Randomly permute the elements of $\mathbf{z}_1$ and $\mathbf{z}_2$ to eliminate the short-lag autocorrelation caused by the interpolation.

6. Reshape the $\mathbf{z}_1$ and $\mathbf{z}_2$ vectors into the two matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively.

7. Fourier transform the product of the complex Gaussian matrix $\underline{\mathbf{Z}} = \mathbf{Z}_1 + i\mathbf{Z}_2$ with the square-root of $\mathbf{\Omega}$ to obtain two fields in the spatial domain (a real-valued and an imaginary one).

Lastly, we would like to stress that our dimensionality reduction approach is fundamentally different from the KL transform. This latter method describes multi-Gaussian fields in a reduced basis that reproduces the large scale variations only. The proposed approach, on the contrary, does not favor one length scale over another, and will lead to reconstructed fields that consistently honor the selected variogram independently of the

number of "super parameters" or dimensionality reduction variables (e.g., elements of $\mathbf{r}$) considered. This is demonstrated in the next section.

2.3. Effects of the Dimensionality Reduction

We first investigate the trade-off between dimensionality reduction and the accuracy of reconstruction, that is, the degree to which the statistical properties of the reconstructed field match those derived from direct generation of the original field. To highlight the essential differences between our approach and the KL expansion, the latter is included in our analysis.

An anisotropic exponential variogram with short integral scales is considered for reconstruction of a 100 $\times$ 100 field (that is, 10,000 grid points). This variogram model characterizes the log conductivity of the reference field used in our inversions (Table 1 and Figure 1a). The grid point mean and variance distributions and the average experimental variograms calculated from 1000 field realizations are analyzed to assess the performance of the dimensionality reductions. The number of variables of each dimensionality reduction, hereafter referred as DR variables, corresponds to the length of the $\mathbf{r}$-vector, $\mathbf{r} = [\mathbf{r}_1, \mathbf{r}_2]$. For the KL transform, the dimensionality reduction variables are the coefficients that multiply the base functions [see, e.g., *Zhang and Zu*, 2004; *Li and Cirpka*, 2006; *Laloy et al.*, 2013, for details] and we refer to these coefficients as KL variables.

Figures 2 and 3 depict the corresponding results for 100, 250 and 1000 DR (Figure 2) and KL (Figure 3) variables. The mean of the reconstructed field (not shown) is not affected by dimensionality reduction, yet the grid point variances clearly are (Figures 2a-2c and 3a-3c). As the number of DR variables increases and dimensionality reduction becomes less important, the distribution of the grid point variance gets narrower and closer

to the statistical fluctuations derived from direct simulation of 1000 standard normal fields (Figures 2a-c). A similar trend is observed for the KL transform (Figures 3a-c), though with much more irregular and over-dispersed variance distributions. Indeed, the proposed approach appears to honor the prescribed variogram independently of the selected number of DR variables (panels 2d-2f). The spurious correlations introduced by dimensionality reduction do not noticeably affect the 2-point correlation structure of the reconstructed field. This is explained by the fact that the fixed permutation scheme used herein causes the (artificial) additional correlations to be distributed independently from the lag (separation) distance between two points. This permutation scheme also has a desired byproduct which is that it simplifies the model reduction error to random noise during inversion. As a consequence of the above, the associated (randomly chosen) field realizations are visually similar to their counterparts derived from direct field generation (compare Figure 1a with Figures 2g-i). Perhaps not surprisingly, the KL is unable to honor the variogram of the reference field even when 1000 KL variables are used (Figures 3d-f), and the generated fields are overly smooth (Figures 3g-i).

We repeated the above analysis for the same geostatistical model except for the integral scale along the major axis of anisotropy that we fixed to a five times larger value, that is, 3.33 m. The main results of this analysis (not shown) are that our proposed approach still outperforms the KL. Even when considering 1,000 KL variables, the KL transform was found to produce over-smoothed fields given the selected exponential variogram model whereas the associated correlation lengths remain slightly overestimated (not shown).

We conclude from this analysis that the proposed dimensionality reduction approach (1) is well suited to reconstruction of multi-Gaussian fields, and (2) outperforms the KL

transform in cases of short and moderate-lag correlation(s). For computational tractability, we use 250 DR variables in our first MCMC trial. Though larger values would ensure less bias in the grid point variance, we consider the deviations of Figure 2b to be acceptable. In a second step, a MCMC trial with 1000 DR variables is performed, and the posterior distributions resulting from using 250 and 1000 DR variables are compared.

2.4. Conditioning to Point Conductivity Measurements

The unconditional (approximately) multi-Gaussian field realizations generated by our method can easily be conditioned on point measurements via kriging [e.g., *Chilès and Delfiner*, 1999]. This reproduces the actual point measurements and preserves the prescribed variogram. Details of this procedure can be found in Appendix A. For the sake of brevity, however, we do not condition on point conductivity measurements in the present paper.

2.5. The Matérn Variogram

We use the *Matérn* [1960] variogram model to describe the geostatistical properties of the field. This function is given by

$$\gamma\left(|\mathbf{h}|\right) = v\left[1 - \frac{1}{\Gamma\left(\nu\right)2^{\nu-1}}\left(\frac{|\mathbf{h}|}{\alpha}\right)^{\nu} K_{\nu}\left(\frac{|\mathbf{h}|}{\alpha}\right)\right], \tag{4}$$

where $\alpha > 0$ is the scale (or range) parameter, $\Gamma\left(\cdot\right)$ represents the gamma function, $K_{\nu}\left(\cdot\right)$ denotes the modified Bessel function of the second kind and order ν, and $|\mathbf{h}|$ signifies the norm of the lag distance vector $\mathbf{h}$. The Matérn function is equivalent to the exponential model for $\nu = 0.5$, the *Whittle* [1954] model for $\nu = 1$ and approaches the Gaussian model for $\nu \to \infty$. The integral scale, I, measures spatial persistence and is defined as [e.g.,

Rubin, 2003]

$$I = \frac{1}{v}\int_0^\infty C\left(|\mathbf{h}|\right) d\,|\mathbf{h}|\,, \tag{5}$$

with covariance function $C\left(|\mathbf{h}|\right) = -\gamma\left(|\mathbf{h}|\right) + v$. The integral scale of the Matérn model depends on the values of α and the shape parameter ν. For a fixed value of α, I becomes larger if ν increases [e.g., *Pardo-Iguzquiza and Chica-Olmo*, 2008]. This dependence might complicate inference of I. Fortunately, numerical simulations with different values of α and ν demonstrates that the ratio of I to α is constant for a given value of ν. For any value of ν, the following fitted polynomial function can be used to derive α from I

$$\frac{I}{\alpha} = -0.0014\nu^6 + 0.0242\nu^5 - 0.1745\nu^4 + 0.6558\nu^3 - 1.4377\nu^2 + 2.4506\nu + 0.0586. \tag{6}$$

This allows us to simultaneously infer I and ν. The coefficient of determination (squared correlation coefficient) associated with equation (6) is 0.999986.

2.6. Joint Inference of Conductivity Fields and Variogram Parameters

In the Bayesian paradigm, the unknown model parameters, $\boldsymbol{\theta}$, are viewed as random variables with a posterior probability density function (pdf), $p\left(\boldsymbol{\theta}|\mathbf{d}\right)$, given by

$$p\left(\boldsymbol{\theta}|\mathbf{d}\right) = \frac{p\left(\boldsymbol{\theta}\right)p\left(\mathbf{d}|\boldsymbol{\theta}\right)}{p\left(\mathbf{d}\right)} \propto p\left(\boldsymbol{\theta}\right)L\left(\boldsymbol{\theta}|\mathbf{d}\right), \tag{7}$$

where $L\left(\boldsymbol{\theta}|\mathbf{d}\right) \equiv p\left(\mathbf{d}|\boldsymbol{\theta}\right)$ signifies the likelihood function of $\boldsymbol{\theta}$. The normalization factor $p\left(\mathbf{d}\right) = \int p\left(\boldsymbol{\theta}\right)p\left(\mathbf{d}|\boldsymbol{\theta}\right)d\boldsymbol{\theta}$ is obtained from numerical integration over the parameter space so that $p\left(\boldsymbol{\theta}|\mathbf{d}\right)$ is a proper probability density function and integrates to unity. The quantity $p\left(\mathbf{d}\right)$ is generally difficult to estimate in practice but is not required for parameter inference. In the remainder of this paper, we will thus focus on the unnormalized density $p\left(\boldsymbol{\theta}|\mathbf{d}\right) \propto p\left(\boldsymbol{\theta}\right)L\left(\boldsymbol{\theta}|\mathbf{d}\right)$. As an exact analytical solution of $p\left(\boldsymbol{\theta}|\mathbf{d}\right)$ is not available in most practical cases, we resort to MCMC simulation to generate samples from the posterior

pdf [see, e.g., *Robert and Casella*, 2004]. The state-of-the-art DREAM$_{(ZS)}$ [*ter Braak and Vrugt*, 2008; *Vrugt et al.*, 2009; *Laloy and Vrugt*, 2012] algorithm is used to approximate the posterior distribution. A detailed description of this sampling scheme including a proof of ergodicity and detailed balance can be found in the cited references. Various contributions in hydrology and geophysics (amongst others) have demonstrated the ability of DREAM$_{(ZS)}$ to successfully recover high-dimensional target distributions [*Laloy et al.*, 2012, 2013; *Linde and Vrugt*, 2013; *Rosas-Carbajal et al.*, 2014; *Laloy et al.*, 2014; *Lochbühler et al.*, 2014, 2015].

Under Gaussian and stationarity assumptions, the field geostatistical properties and pixel/voxel random number values that jointly define the (base ten) log conductivity field, $\log_{10}(\mathbf{K})$, can be inferred simultaneously using MCMC simulation. We use the Matérn function to infer field smoothness jointly with the standard normal variates and other geostatistical parameters. The following geostatistical parameters are sampled together with $\mathbf{r}_1$ and $\mathbf{r}_2$: (I) m, the mean, (II) v, the variance, (III) I_M, the integral scale along the major axis of anisotropy, (IV) R_I, the ratio of the integral scale along the minor axis of anisotropy (I_m) to the integral scale along the major axis of anisotropy, (V) A, the anisotropy direction or angle (rotation anti-clockwise from the z-axis), and (VI) ν, the shape parameter of the Matérn function. To build the covariance kernel, $\mathbf{C}$, we used the mGstat geostatistical toolbox in MATLAB (`http://mgstat.sourceforge.net/`).

If we assume the N-vector of residual errors (differences between the measured and simulated data), $\mathbf{e}$, to be Gaussian distributed, uncorrelated and with constant variance, σ_e^2, the likelihood function of $\boldsymbol{\theta}$ can be written as

$$L(\boldsymbol{\theta}|\mathbf{d}) = \left(\frac{1}{\sqrt{2\pi\sigma_e^2}}\right)^N \exp\left(-\frac{1}{2}\sigma_e^{-2}\sum_{i=1}^{N}\left[d_i - F_i(\boldsymbol{\theta})\right]^2\right), \quad (8)$$

where $\mathbf{d} = (d_1, \ldots, d_N)$ is a set of measurements, and $F(\boldsymbol{\theta})$ is a deterministic "forward" model. The standard deviation of the residuals, σ_e (kg m^{-3}), is jointly inferred with the other unknown variables, and thus $\boldsymbol{\theta} = [\sigma_e, m, v, I_M, R_I, A, \nu, \mathbf{r}_1, \mathbf{r}_2]$ (see Table 1). The number of parameters sampled with MCMC is thus equivalent to the number of DR variables plus seven. This equates to a total of 257 parameters for the first MCMC trial with 250 DR variables, and 1007 parameters for the second trial with 1000 DR variables.

The standard normal distribution of $\mathbf{z}_1$ and $\mathbf{z}_2$ (and thus $\mathbf{r}_1$ and $\mathbf{r}_2$) can be enforced by the use of a standard normal prior

$$p(\mathbf{r}) = \frac{\exp\left(-\frac{1}{2}\mathbf{r}^T\mathbf{r}\right)}{\sqrt{2\pi}^N}, \tag{9}$$

in which the superscript T signifies transpose and $\mathbf{r} = [\mathbf{r}_1, \mathbf{r}_2]$. The variogram is assumed to be largely unknown and thus characterized by a wide prior, details of which will follow in section 3.

3. Case Studies

3.1. Model Setup

The 100 × 100 modeling domain lies in the $x - z$ plane with a grid cell size of 0.2 m (Figure 1a). Steady state groundwater flow is simulated using MaFloT [*Künze and Lunati*, 2012] assuming no flow boundaries at the top and bottom and fixed head boundaries on the left and right sides of the domain so that a lateral head gradient of 0.025 (-) is imposed, with water flow in the x-direction. For the tracer experiment, we consider two different boreholes that are 20 m apart. A conservative tracer with concentration of 1 kg m^{-3} is applied into the fully screened left borehole using a step function. The background solute concentration is assumed to be 0.01 kg m^{-3}. Ignoring density effects,

conservative transport of the tracer through the subsurface is simulated with MaFloT using open boundaries on all sides, and longitudinal and transverse dispersivities of 0.1 and 0.01 m, respectively. Solute transport was monitored during a period of 10 days with concentration measurements made every day at nine different depths (2, 4, 6, 8, 10, 12, 14, 16, and 18 meters) in the borehole at the right hand side. The total number of observations is therefore 90. These measurement data were then corrupted with a Gaussian white noise using a standard deviation equivalent to 5% of the mean observed concentration. This led to a root-mean-square-error (RMSE) of 0.039 kg m^{-3} between error-free and noisy data (Figure 1b).

3.2. Inference of an Heterogeneous Random Field with Short Integral Scales

Our case study considers a reference log conductivity field with an exponential variogram model and fairly short integral scales compared to the domain size of 20 × 20 m (Figure 1a). The values of the geostatistical parameters are: $\nu = 0.5$, $I_M = 0.67$ m and $R_I = 0.25$ ($I_m = 0.17$ m). Furthermore, we assume value of m = -3 and $v = 1$ for the mean and variance of the log-conductivity field, whereas the anisotropy angle, A, is set to 75 degrees. In the absence of prior information about the geostatistical parameters (with exception of the ranges of the search space), we assumed either uniform or *Jeffreys* [1946] (that is, log-uniform) truncated individual priors that span a wide range of values. We selected a bounded uniform prior for m and a bounded Jeffreys prior for v. This is a common choice in the inference of multi-Gaussian fields [e.g., *Box and Tiao*, 1973; *Rubin et al.*, 2010]. We chose bounded uniform priors for I_M, A, and R_I, and a bounded Jeffreys prior for ν. Also, a Jeffreys prior is selected for σ_e. Table 1 summarizes the prior distribution and

corresponding ranges of each parameter. For completeness, we also list the true values of the geostatistical parameters used to generate the reference field.

We estimate the posterior distribution of the parameters using MCMC simulation with DREAM$_{(ZS)}$. Default values of the algorithmic variables are used. Yet the number of Markov chains was increased to eight and the number of crossover values (geometric series) set to 25 to enhance the MCMC search capabilities for this high-dimensional parameter space. To further increase the acceptance rate of proposals, we decreased the default jump rate of DREAM$_{(ZS)}$ by a factor of four. Of course, we could have tuned the jump rate automatically in DREAM$_{(ZS)}$ to achieve a certain desired acceptance rate of proposals but choose this simpler approach. Convergence of the sampled Markov chains was monitored using the potential scale reduction factor, $\hat{R}$ [*Gelman and Rubin*, 1992]. This statistic compares for each parameter of interest the average within-chain variance to the variance of all the chains mixed together. The closer the values of these two variances, the closer to one the value of the $\hat{R}$ diagnostic. Values of $\hat{R}$ smaller than 1.2 are commonly deemed to indicate convergence to a limiting distribution. Our simulation results indicate that convergence is achieved after about 400,000 forward model evaluations (FEs) (shown later). Visual inspection of the sampled likelihood values suggests however, that far fewer model evaluations are needed for every chain to locate the posterior distribution.

Figure 4a presents the evolution of the $\hat{R}$ statistic calculated from the last 90% of the samples in each chain, and Figure 4b depicts a trace plot of the sampled RMSE values for each of the eight Markov chains. The average acceptance rate (AR) is about 33.3 % (not shown). All chains appear to sample stable RMSE (and thus likelihood) values after approximately 40,000 FEs (Figure 4b). However, another 360,000 FEs are required before

all $\hat{R}$ values are smaller than 1.2 and official convergence can be declared (Figure 4a). It is not surprising that the sampled RMSE values stabilize much faster than the associated values of the $\hat{R}$ diagnostic. The chains converge rapidly to a point in the posterior but many more function evaluations are required to fully explore this distribution and satisfy requirements for convergence.

Marginal distributions of σ_e, the geostatistical parameters and r_1 and r_{250} are depicted in Figure 5 using kernel density smoothing. The prior distribution is also shown. The standard deviation of the residuals, σ_e, and the field mean of the log conductivity, m, appear very well resolved. Despite a 40-times dimensionality reduction, the posterior distributions of the geostatistical parameters contain their true values used to create the reference conductivity field. The posterior modes are somewhat removed from the true values, especially for A and to a lesser extent ν. This is due to measurement errors and the use of a reduced parameter space which inevitably introduces some bias in the sampled posterior distribution.

Figure 6 displays the reference conductivity field and eight randomly chosen samples from the posterior distribution. The posterior conductivity fields differ substantially from each other, yet all of them produce simulation results that are in (statistical) agreement with the observed data. The geostatistical properties of realization V are relatively similar to those of the reference field. The other posterior fields include fairly different spatial statistics (e.g. realizations III, IV and VI).

To investigate the bias introduced by the dimensionality reduction in the posterior estimates we would need to compare our results for the DREAM$_{(ZS)}$ trial with 250 DR variables against those of DREAM$_{(ZS)}$ for the original parameter space. However, such a sam-

pling run is computationally intractable. Instead, we performed a trial with DREAM$_{(ZS)}$ using 1000 DR variables and thus 1007 parameters.

The results of this more complex run are fairly similar to those of our trial with 250 DR variables. Again, about 30,000 - 40,000 FEs are required to reach stable values of the RMSE (not shown), yet a larger computational budget of approximately 1 million FEs is required for this 1007 dimensional search space before convergence to a limiting distribution can be officially declared (not shown). This is more than twice the number of FEs needed for the previous trial with 250 DR variables, but arguably rather efficient considering the about fourfold increase in parameter dimensionality. The AR of DREAM$_{(ZS)}$ is rather large (45%) but results in a good mixing of the individual chains. Perhaps most importantly, the marginal distributions of the geostatistical parameters plotted in Figures 5b-g (dashed red line) are in good agreement with their counterparts derived from the DREAM$_{(ZS)}$ trial with 250 DR variables (solid blue line). For the trial with 1000 DR variables as well, the marginal distributions include the true values used to generate the reference log-conductivity field. Note though that the distributions of most of the geostatistical parameters have become somewhat more peaky, most noticeably for A. The distributions of m, v, I_M and ν are also more centered on their true values.

Altogether we conclude that the posterior distribution of the geostatistical parameters is only weakly affected by dimensionality reduction, let alone the maximum a posteriori (MAP) values which are better resolved as dimensionality reduction decreases.

The field realizations resulting from the 1007-dimensional posterior distribution are in strong qualitative agreement with the counterparts derived from the 257-dimensional

posterior distribution, but with less variation in the anisotropy angle, and a slight tendency towards smoother fields (Figure 7).

3.3. Comparison With Other Posterior Sampling Methods

3.3.1. Comparison Against Sequential Gibbs Sampling

Now that we have discussed the main elements of our inversion methodology we are left with a comparison against state-of-the-art methods in the literature. As a first test, we consider the SGS method for variogram estimation as implemented in the SIPPI 0.94 toolbox [*Hansen et al.*, 2013a, b]. This open source MATLAB package is described in detail in the cited references, and interested readers are referred to these publications. We used default settings for the algorithmic variables, and the same prior distribution for σ_e and the geostatistical parameters as used in our numerical experiments described previously. Furthermore, we added to the SIPPI toolbox the Matérn variogram as this function was not yet incorporated in the toolbox.

Before proceeding with our results, we would like to emphasize that SGS (or the very similar but independently developed iterative spatial resampling (ISR) scheme by *Mariethoz et al.* [2010]) is a powerful MCMC algorithm for sampling from complex geologic prior models [e.g, *Mariethoz et al.*, 2010; *Hansen et al.*, 2012]. This method creates candidate points by conditioning a field realization drawn from the prior to a randomly chosen set of points from the current state (and hence model/field) of the Markov chain. Nonetheless, SGS with variogram inference suffers one important drawback and that is that it relies completely on sampling from the prior distribution of the variogram parameters [see *Hansen et al.*, 2013b, for details]. Moreover, SGS uses a single Markov chain in pursuit of the posterior distribution. This not only makes it difficult to rapidly explore

multi-dimensional parameter spaces, but also complicates assessment of convergence, and effective use of multi-processor resources. Pre-fetching [*Brockwell*, 2006] and multi-try Metropolis sampling [*Liu et al.*, 2000] offer some options for distributed, multi-core implementation of single chains. What is more, the use of a single chain increases chances of premature convergence. To mitigate this risk, it is generally recommended to perform several independent trials and verify whether the different chains have converged to the approximate same limiting distribution. In contrast to SGS, DREAM$_{(ZS)}$ is embarrassingly parallel and thus readily amenable to multi-processor distributed computation which should drastically reduce the required CPU-time for posterior exploration. In the case studies presented herein, we ran each of the eight different Markov chains of DREAM$_{(ZS)}$ on a different processor. This significantly reduced the CPU-time required for posterior exploration, details of which will be presented in section 4 of this paper.

For proper convergence assessment, we performed three (independent) SGS trials using starting points drawn randomly from the prior distribution. Because of the associated computational costs, we terminated the calculation after a total of 500,000 FEs (that is, 166,667 FEs per Markov chain) and the comparison with our method is made on the basis of the same computational budget of 500,000 FEs. We used default settings of SGS and individually sampled, with equal probability, the different geostatistical parameters and vector of (standard normal) field values. Each iteration produces a candidate log conductivity field as follows. With probability 1/8, either a new 10,000-dimensional vector of standard normal variates, say $\mathbf{g}$, is produced, or the current geostatistical model is updated by replacing one of the six geostatistical parameters with a random draw from its prior, or a new value of σ_e is sampled from $p(\sigma_e)$. When $\mathbf{g}$ is updated, the candidate

model (proposal) is obtained by conditioning on a fraction, ϕ, of locations randomly chosen from the current model. The value of ϕ is adapted during burn-in to achieve a targeted acceptance rate, which we set to 20%. The upper bound of ϕ was set to 1 (i.e., totally different proposal) whereas its lower bound was fixed to 0.001, that is, only 10 of the 10,000 $\log_{10}(K)$ values are perturbed per iteration.

With an average (adapted) value of ϕ reaching its lower bound of 0.001, the mean AR values for each of the eight individual sampling steps are 20.7, 1.0, 7.9, 5.1, 9.4, 4.9, 13.4, and 7.0%, for $\mathbf{g}$, m, v, I_M, A, R_I, ν, and σ_e, respectively. Both SGS and the proposed sampling method successfully fit the data to the prescribed error level (Figures 8a and 8b), but the SGS method needs somewhat fewer function evaluations to do so. Nevertheless, our proposed inversion approach is more effective and efficient in exploring the posterior target, as shown by the respective Markov chain trajectories of DREAM$_{(ZS)}$ and SGS for the field variance (Figures 8c and 8d) and integral scale along the major anisotropy axis (Figures 8e and 8f). Even after a total of 500,000 FEs or 166,667 FEs per chain, the three SGS trials do not converge appropriately to the true value of I_M. The DREAM$_{(ZS)}$ algorithm on the contrary needs about 200,000 FEs (that is, 25,000 FEs per chain) to converge to the reference value of 0.67. Moreover, SGS has difficulty in sampling the correct value of v as well. Two of the three chains are somewhat stuck near its lower bound.

A convergence check of the three chains sampled by SGS is provided in Figure 9a. For convenience we plot only the evolution of the $\hat{R}$ statistic of σ_e and the six variogram parameters. In practice, SGS samples 10,007 parameters, and hence convergence can only be formally declared if all sampled parameters fall below the threshold value of

1.2 (horizontal black line). Nevertheless, it is evident that SGS is unable to converge adequately within the allowed computational budget. Even after 500,000 FEs several of the plotted trajectories remain well above the $\hat{R}$-threshold value of 1.2. As a consequence, the corresponding posterior distribution is quite inaccurate. The posterior mode of v is close to its lower bound, well removed from the reference value (Figure 9b). What is more, a multimodal posterior distribution is observed for I_M with true value that falls in a region with lower posterior probability (Figure 9c). Finally, the posterior fields sampled by SGS exhibit considerably more correlation at different spatial lags than its counterpart derived from our proposed approach (Figure 9d).

3.3.2. Comparison Against the Method of Anchored Distributions

The MAD method [see *Rubin et al.*, 2010, for an extensive description] is especially designed for inference of (multi-)Gaussian parameter fields. It differs from classical Bayesian inference methods in the treatment of the likelihood function, $L(\boldsymbol{\theta}|\mathbf{d}) \equiv p(\mathbf{d}|\boldsymbol{\theta})$. Whereas SGS and our proposed inversion method describe $p(\mathbf{d}|\boldsymbol{\theta})$ as a parametric probability distribution of the residuals (e.g, equation (8)) that is specified a-priori, MAD takes $p(\mathbf{d}|\boldsymbol{\theta})$ as the conditional probability density of the simulated data given a parameter set $\boldsymbol{\theta}$. This is done by approximating $p(\mathbf{d}|\boldsymbol{\theta})$ from an ensemble of conditional simulations, using a nonparametric approach whenever computationally tractable. This method has several advantages, one of which is that it avoids making strong and sometimes unjustified assumptions about the properties of the residual errors. In practice, however, and because of computational constraints, at least some parametric (Gaussian) assumptions often need to be made about $p(\mathbf{d}|\boldsymbol{\theta})$ a-priori [e.g., *Murakami et al.*, 2010; *Over et al.*, 2015]. Another feature of MAD is that it uses basic Monte Carlo simulation to solve for the posterior

parameter distribution. Once the anchor locations have been defined, the inferred parameters are drawn randomly from their (marginal) prior distribution for a pre-specified number of times. This is not very efficient, especially if the posterior distribution constitutes only a small part of the prior distribution. Hence, even with the assumption of a Gaussian distribution for $p(\mathbf{d}|\boldsymbol{\theta})$, the total number of forward model evaluations required by MAD will typically be on the order of several millions [e.g., *Murakami et al.*, 2010; *Over et al.*, 2015]. This requires the use of many processors on a distributed computing network [on the order of several thousands, e.g., *Murakami et al.*, 2010].

MAD distinguishes between two types of inferred variables: variogram parameters, $\boldsymbol{\theta}_{\mathrm{V}}$ and conductivity values at selected locations or anchor sets, $\boldsymbol{\theta}_{\mathrm{K}}$. If no direct point conductivity measurements are considered in the inference, the posterior distribution $p(\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}}|\mathbf{d})$ reduces to

$$p(\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}}|\mathbf{d}) \propto p(\boldsymbol{\theta}_{\mathrm{V}})\, p(\boldsymbol{\theta}_{\mathrm{K}}|\boldsymbol{\theta}_{\mathrm{V}})\, p(\mathbf{d}|\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}}), \tag{10}$$

where $p(\boldsymbol{\theta}_{\mathrm{V}})$ denotes the prior distribution of the variogram parameters, $p(\boldsymbol{\theta}_{\mathrm{K}}|\boldsymbol{\theta}_{\mathrm{V}})$ signifies the prior anchor distribution given a variogram parameter vector $\boldsymbol{\theta}_{\mathrm{V}}$, and $p(\mathbf{d}|\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}})$ is the likelihood function of $\{\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}}\}$. While $p(\boldsymbol{\theta}_{\mathrm{V}})$ and $p(\boldsymbol{\theta}_{\mathrm{K}}|\boldsymbol{\theta}_{\mathrm{V}})$ can be derived analytically, numerical estimation of $p(\mathbf{d}|\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}})$ is a complicated task. MAD proceeds as follows. First, define the anchor locations. Then sample $p(\boldsymbol{\theta}_{\mathrm{V}})$ n_V times and for each of the resulting $\boldsymbol{\theta}_{\mathrm{V}}$ vectors, sample $p(\boldsymbol{\theta}_{\mathrm{K}}|\boldsymbol{\theta}_{\mathrm{V}})$ n_K times. Third, create an ensemble of n_f random fields, $\boldsymbol{\Psi}$, from each of the $n_V \times n_K$ $\{\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}}\}$ parameter sets using a direct conditioning method such as the one described in Appendix A. The n_f fields of $\boldsymbol{\Psi}$ thus exactly honor $\boldsymbol{\theta}_{\mathrm{K}}$ and are distributed according to $\boldsymbol{\theta}_{\mathrm{V}}$, and result into n_f simulated data vectors, $\boldsymbol{\Delta}$. Finally, the likelihood, $p(\mathbf{d}|\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}})$, of the parameter set under evaluation,

$\{\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}}\}$, is approximated by fitting a multivariate density distribution to the multivariate frequency distribution of the simulated data stored in $\boldsymbol{\Delta}$. For this, one can use non-parametric kernel density estimation (possibly after reduction of the data vector) or assume a Gaussian parametric model for $\mathbf{d}$. The total number of forward model calls is therefore equal to $n_V \times n_K \times n_f$.

We used 49 anchors on a regular grid (Figure 10a) and set n_V to 500, n_K to 12 and n_f to 100, leading to a total computational budget of 600,000 FEs. The 500 samples from $p\left(\boldsymbol{\theta}_{\mathrm{V}}\right)$ were drawn randomly using Latin hypercube sampling. Our values for n_V, n_K and n_f are based on the work of *Murakami et al.* [2010] who used 44 anchor locations, $n_V = 3000$, $n_K = 12$ and $n_f = 250$ for a grid of approximatively similar dimension as in our numerical experiments herein, but with a much smaller prior variogram uncertainty. Obviously, the best choice for the MAD algorithmic parameters is problem-dependent and our settings might not be optimal. We would like to stress, however, that a total of 600,000 FEs for MAD is justified. Indeed, a computational budget of only 500,000 FEs was assigned to our proposed inversion approach (section 3.2) and the SGS method (section 3.3.1).

Figure 10b presents a histogram of the RMSE values derived from the 600,000 forward model calls. The minimum RMSE sampled by MAD is 0.051 kg m^{-3} (vertical black line) which is not only significantly larger than the true value of 0.039 kg m^{-3} for the reference field but also outside the posterior distribution of RMSE values sampled by our approach and SGS (see Figures 4b and 8a-b). No matter how the likelihoods of the $i = 1, \cdots, n_V \times n_K$ $\{\boldsymbol{\theta}_{\mathrm{V}}, \boldsymbol{\theta}_{\mathrm{K}}\}_i$ parameter sets are estimated, inference from these 600,000 forward runs can thus only be flawed. Obviously, appropriate (random) sampling of the prior parameter space would require much larger values of n_V and n_K.

4. Discussion

Some remarks about the presented study are in order. Due to time constraints, the different MCMC and MAD trials were only performed once. Repeated sampling runs with different random seeds would provide a more accurate benchmark of our inversion methodology. Nevertheless, the results presented herein inspire confidence in the effectiveness and efficiency of our proposed Bayesian inversion method.

The computational requirement is an important issue. Considering a serial calculation framework, we find that the serial implementation of our approach outperforms both SGS and MAD for the case study considered herein. The DREAM$_{(ZS)}$ sampler is however designed specifically so that it is embarrassingly parallel and thus can take maximum advantage of multi-processor resources. We did so herein using an 8-core workstation, assigning each of the eight interacting Markov chains to a different processor. This resulted in a six times speed up of the calculations. The advantages of the proposed multi-core approach are hence evident. Note also that if more parallel cores are available, then the search efficiency can be further increased by using the multi-try variant of DREAM$_{(ZS)}$ [*Laloy and Vrugt*, 2012].

Another crucial point is the inevitable trade-off between model truncation (dimensionality reduction) and the accuracy of the posterior field realizations. The MCMC search is performed within a truncated model space that constitutes only a subset of the original model space. By construction, not all possible posterior models can therefore be represented by the truncated posterior pdf. Model truncation might also bias the posterior distribution by shifting the probability mass away from the original MAP values. In addition, for a given truncation level the peakedness of the likelihood function will influence

the quality of approximation of the original posterior distribution by the truncated posterior distribution. A dimensionality reduction with a factor 40 was shown to work well for the considered case study. The resulting posterior distribution was found to be in good agreement with the distribution stemming from a dimensionality reduction with a factor 10. As model space truncation or peakedness of the likelihood further increases, the distribution and associated parameter uncertainties will nevertheless be increasingly corrupted. The combined effects of dimensionality reduction and measurement data quality on the accuracy of the estimated target distribution deserve further analysis.

The *Gelman and Rubin* [1992] potential scale reduction factor was computed using the last 90% of the generated samples in each Markov chain evolved by DREAM$_{(ZS)}$. This value differs from the default of 50% used in DREAM$_{(ZS)}$, but is warranted in each of our case studies because the joint chains converge to stable RMSE values within less than 10% of the assigned computational budget. Indeed, the sampled RMSE (and thus likelihood) values appropriately converge within 30,000 - 40,000 FEs. The use of 90% of the chains is equivalent to a burn-in of 50,000 (trial with 250 DR variables) and 100,000 (trial with 1000 DR variables) samples, well beyond when the posterior distribution has been located.

This study considers a two-dimensional flow and transport modeling domain. Extension of the proposed approach to 3D domains is straightforward and will be investigated in future work. Extension to pluri-Gaussian simulation [e.g., *Lantuéjoul et al.*, 2002] for inference of categorical conductivity fields also seems promising.

5. Conclusions

This paper presents a novel Bayesian inversion scheme for the simultaneous estimation of high-dimensional multi-Gaussian conductivity fields and associated geostatistical

properties from indirect hydrological data. Our method merges Gaussian process generation via circulant embedding [*Dietrich and Newsam*, 1997] to decouple the variogram from grid cell specific values, with dimensionality reduction by interpolation to facilitate Markov chain Monte Carlo (MCMC) simulation with the DREAM$_{(ZS)}$ algorithm [*ter Braak and Vrugt*, 2008; *Vrugt et al.*, 2009; *Laloy and Vrugt*, 2012]. We use the Matérn variogram model to infer the conductivity values simultaneously with field smoothness (or Matérn shape parameter) and other geostatistical parameters (mean, sill, integral scales and anisotropy factor(s)). The proposed dimensionality reduction approach systematically honors the prescribed variogram and is shown to outperform the Karhunen-Loève [*Loève*, 1977] transform. Our inverse method is demonstrated using synthetic, error corrupted, data from a flow and transport model involving a fairly heterogeneous 10,000-dimensional multi-Gaussian conductivity field. Despite a reduction of the parameter space by a factor of 40, the measurement data were fitted to the prescribed noise level while the derived posterior parameter distributions always included the true geostatistical parameter values. A comparison between the posterior distributions derived for a 257- (40-times reduction) and 1007-dimensional (10-times reduction) parameter space, respectively, indicated that the bias introduced by the dimensionality reduction in the posterior estimates is rather small. This inspires confidence in the effectiveness of the approach. The posterior field uncertainty encompassed a large range of different geostatistical models, which calls into question the common practice in hydrogeology of fixing the variogram model before inversion. For the considered case study, the serial version of our method appears to be more computationally efficient than both the SGS algorithm of *Hansen et al.* [2012, 2013a, b] and the MAD method of *Rubin et al.* [2010]. The advantages of the proposed approach

are even more apparent when executed on a distributed computing network. A 6-times reduction in CPU-time was observed with DREAM$_{(ZS)}$ using parallel evaluation of the eight different Markov chains. Future work will investigate the application of the proposed approach to 3D modeling domains, and to pluri-Gaussian simulations for inference of categorical field structures.

Appendix A

Conditioning an unconditional simulation of a random field can be easily performed via kriging. Kriging-based geostatistical methods are extensively described in the literature [e.g., *Chilès and Delfiner*, 1999; *Ren et al.*, 2005; *Huang et al.*, 2011].

If the values of $Y(\mathbf{x})$ are known at locations $\mathbf{x}_i, i = 1 \cdots n_y$, the value of $Y(\mathbf{x})$ at any arbitrary location $\mathbf{x}$, $Y_{\mathrm{kr}}(\mathbf{x})$, can be predicted unbiasedly as

$$Y_{\mathrm{kr}}(\mathbf{x}) = \sum_{i=1}^{n_y} \lambda_i Y(\mathbf{x}_i) \tag{A1}$$

in which the kriging weights λ_i depend solely on the prescribed variogram. Now if we have an unconditional realization, $Y_{\mathrm{uc}}(\mathbf{x})$, the corresponding random field conditioned to the n_y observed values, Y_{c}, is given by

$$Y_{\mathrm{c}}(\mathbf{x}) = Y_{\mathrm{uc}}(\mathbf{x}) + [Y_{\mathrm{kr}}(\mathbf{x}) - Y_{\mathrm{kr-u}}(\mathbf{x})], \tag{A2}$$

where $Y_{\mathrm{kr-u}}$ is obtained by kriging (equation (A1)) using the unconditional simulated values at the n_y data locations, and the same kriging weights are used for determining $Y_{\mathrm{kr}}(\mathbf{x})$ and $Y_{\mathrm{kr-u}}(\mathbf{x})$. Equation (A2) implies that, at each conditioning data location, the unconditional simulated value is taken out and replaced by the conditioning datum. In the vicinity of a conditioning data location, the kriging operator smooths the change between the conditioning data and the unconditional simulated values outside the range

of kriged values. The conditioning is therefore exact at data locations whereas beyond the correction range, the conditional simulated values will be the unconditional simulated values.

Using matrix notation, the λ_i in equation (A1) are obtained for a random field with dimension $m \times n$ as

$$\boldsymbol{\lambda} = \mathbf{C}_{\mathrm{df}}^{T} \mathbf{C}_{\mathrm{dd}}^{-1}, \tag{A3}$$

where $\mathbf{C}_{\mathrm{df}}$ is the $n_y \times (m \times n)$ matrix of covariances between data and target field values, $\mathbf{C}_{\mathrm{dd}}$ is the data-to-data covariance matrix, and the superscripts T and -1 denote transpose and inverse matrix operations, respectively.

Acknowledgments.

We would like to thank the Associate Editor Olaf Cirpka and three anonymous reviewers for their useful comments and suggestions which significantly helped to improve the manuscript. We are grateful to Thomas Mejer Hansen and coworkers for sharing online their mGstat and SIPPI toolboxes. We also like to thank Rouven Künze for providing us with the MaFloT simulator. A MATLAB code of the approach proposed in this study is available from the first author (`elaloy@sckcen.be`). The general-purpose DREAM$_{(\mathrm{ZS})}$ algorithm is available from the fourth author (`jasper@uci.edu`).

References

Box, G. E. P., and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley.

Brockwell, A. E. (2006), Parallel Markov chain Monte Carlo simulation by pre-fetching, *J. Comp. Graph. Stat., 15*, 246–261, doi:10.1198/106186006X100579.

Chan G., and A.T.A. Wood (1997), Algorithm AS 312: An algorithm for simulating stationary Gaussian random fields, *J. Roy. Stat. Soc. C-App., 46*, 171–181.

Chilès, J.-P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, Wiley.

Dietrich, C. R., and G. H. Newsam (1997), Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix, *SIAM J. Sci. Comput., 18*, 1088–1107.

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367.

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

Gelman, A. G., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci., 7*, 457–472.

Hansen, T. M., K. C. Cordua, and K. Mosegaard (2012), Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling, *Comput. Geosci.*, *16*, 593–611, doi:10.1007/s10596-011-9271-1.

Hansen, T. M., K. C. Cordua, M. C. Looms, and K. Mosegaard (2013a), SIPPI: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 1 - Methodology, *Comput. Geosci.*, *52*, 470–480, doi:10.1016/j.cageo.2012.09.004.

Hansen, T. M., K. C. Cordua, M. C. Looms, and K. Mosegaard (2013b), SIPPI: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 2 - Application to crosshole GPR tomography, *Comput. Geosci.*, *52*, 481–492, doi:10.1016/j.cageo.2012.10.001.

Hendricks-Franssen, H.J., A. Alcolea, M. Riva, M. Bakr, N. van der Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse modelling of groundwater flow: application to the characterisation of well catchments. *Adv. Water Resour., 32*(6), 851–72, doi:10.1016/j.advwatres.2009.02.011.

Huang, J. W., G. Bellefleur, and B. Milkereit (2010), CSimMDMV: A parallel program for stochastic characterization of multi-dimensional, multi-variant, and multi-scale distribution of heterogeneous reservoir rock properties from well log data, *Comput. Geosci.*, *37*, 1763–1776, doi:10.1016/j.cageo.2010.11.012.

Jafarpour B., and M. Tarrahi (2011), Assessing the performance of the ensemble Kalman filter for subsurface flow data integration under variogram uncertainty, *Water Resour. Res.*, *47*(5), doi:10.1029/2010WR009090.

Jardani, A., J. P. Dupont, A. Revil, M. Massei, M. Fournier, and B. Laignel (2012), Geostatistical inverse modeling of the transmissivity field of a heterogeneous alluvial aquifer under tidal influence, *J. Hydrol.*, *472*, 287–300.

Jeffreys, H. (1946), An invariant form for the prior probability in estimation problems, *Proc. R. Soc. London*, *A186*, 453–461.

Kitanidis, P. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, *31*(10), 2411-9, doi:10.1029/95WR01945.

Kroese, D. P., and Z. I. Botev (2013), Spatial Process Generation, in *Lectures on Stochastic Geometry, Spatial Statistics and Random Fields, Volume II: Analysis, Modeling and Simulation of Complex Structures*, edited by V. Schmidt, Springer-Verlag.

Künze R, and I. Lunati (2012), An adaptive multiscale method for density-driven instabilities, *J. Comput. Phys.*, *231*, 5557–5570.

Laloy, E., and J. A. Vrugt (2012), High-dimensional posterior exploration of hydrologic models using multiple-try DREAM$_{(ZS)}$ and high performance computing, *Water Resour. Res.*, *48*, W01526, doi:10.1029/2011WR010608.

Laloy, E., N. Linde, and J. A. Vrugt (2012), Mass conservative three-dimensional water tracer distribution from Markov chain Monte Carlo inversion of time-lapse ground-penetrating radar data, *Water Resour. Res.*, *48*, W07510, doi:10.1029/2011WR011238.

Laloy, E., B. Rogiers, J. A. Vrugt, D. Mallants, and D. Jacques (2013), Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov Chain Monte Carlo simulation and polynomial chaos expansion, *Water Resour. Res.*, *49*, 2664–2682, doi:10.1002/wrcr.20226.

Laloy, E., J. A. Huisman, and D. Jacques (2014), High-resolution moisture profiles from full-waveform probabilistic inversion of TDR signals, *J. Hydrol.*, *519*, 2121–2135, doi:10.1016/j.jhydrol.2014.10.005.

Lantuéjoul, C. (2002), *Geostatistical simulation: models and algorithms*, Springer.

Le Ravalec, M., B. Noetinger, B., and L. Y. Hu (2000), The FFT moving average (FFT-MA) generator: an efficient numerical method for generating and conditioning Gaussian simulations, *Math. Geol.*, *32*(6),701–723.

Li, W. and O. A. Cirpka (2006), Efficient geostatistical inverse methods for structured and unstructured grids, *Water Resour. Res., 42*, W06402, doi:10.1029/2005WR004668.

Linde, N., and J. A. Vrugt (2013), Distributed soil moisture from crosshole ground-penetrating radar travel times using stochastic inversion, *Vadose Zone J.*, *12*(1), doi:10.2136/vzj2012.0101.

Liu, J. S., F. Liang, and W. H. Wong (2000), The multiple-try method and local optimization in Metropolis sampling, *J. Am. Stat. Assoc., 95*(449), 121–134, doi:10.2307/2669532.

Lochbühler, T., S. J. Breen. R. L. Detwiler, J. A. Vrugt, and N. Linde (2014), Probabilistic electrical resistivity tomography for a CO2 sequestration analog, *J. Appl. Geophys.*, *107*, 80–92, doi:10.1016/j.jappgeo.2014.05.013.

Lochbühler, T., J. A. Vrugt, M. Sadegh, and N. Linde (2015), Summary statistics from training images as prior information in probabilistic inversion, *Geophys. J. Int.*, *201*, 157–171. doi:10.1093/gji/ggv008.

Loève, M. (1977), *Probability Theory, Fourth edition*, Springer-Verlag, New York.

Mariethoz, G., P. Renard, and J. Caers (2010), Bayesian inverse problem and optimization with iterative spatial resampling, *Water Resour. Res., 46*, W11530, doi:10.1029/2010WR009274.

Matérn, B. (1960), Spatial Variation, *Meddelanden fran Statens Skogsforskningsinstitut*, *49*5 (second ed., 1986, Lecture notes in Statistics, no. 36, Springer).

Murakami, H., X. Chen, M. S. Hahn, Y. Liu, M. L. Rockhold, V. R. Vermeul, J. M. Zachara, and Y. Rubin (2010), Bayesian approach for three-dimensional aquifer characterization at the Hanford 300 Area, *Hydrol. Earth Syst. Sci.*, *14*, 19892001, doi:10.5194/hess-141989-2010.

Nowak, W., F. P. J. de Barros, and Y. Rubin (2010), Bayesian geostatistical design: Task-driven optimal site investigation when the geostatistical model is uncertain, *Water Resour. Res.*, *46*, W03535, doi:10.1029/2009WR008312.

778 Ortiz, J. C, and C. L. Deutsch (2002), Calculation of uncertainty in the variogram, *Math.*
779 *Geol.*, *34*(2), 169–183.

780 Over, M., W., U. Wollschläger, C. A. Osorio-Murillo, and Y. Rubin (2015),
781 Bayesian inversion of Mualem-van Genuchten parameters in a multi-layer soil pro-
782 file: a data-driven, assumption-free likelihood function, *Water Resour. Res.*, *51*,
783 doi:10.1002/2014WR015252.

784 Pardo-Iguzquiza, E., and M. Chica-Olmo (2008), Geostatistics with the Matern semivar-
785 iogram model: A library of computer programs for inference, kriging and simulation.
786 *Comput. Geosci.*, *34*(9), 1073–1079.

787 Ren, W., L. Cunha, and C. V. Deutsch (2005), Preservation of multiple point structure
788 when conditioning by kriging, in *Geostatistics Banff 2004*, Quantitative Geology and
789 Geostatistics Vol. 4, edited by O. Leuangthong and C.V. Deutsch, Springer.

790 Robert, C. P., and G. Casella (2004), *Monte Carlo statistical methods*, second edition.
791 Springer.

792 Rosas-Carbajal, M., N. Linde, T. Kalscheuer, and J. A. Vrugt (2014), Two-dimensional
793 probabilistic inversion of plane-wave electromagnetic data: Methodology, model con-
794 straints and joint inversion with electrical resistivity data, *Geophys. J. Int.*, *196*, 1508–
795 1524, doi: 10.1093/gji/ggt482.

796 Rubin, Y. (2003), *Applied Stochastic Hydrology*, Oxford University Press.

797 Rubin, Y., X. Chen, H. Murakami, and M. Hahn (2010), A Bayesian approach for inverse
798 modeling, data assimilation, and conditional simulation of spatial random fields, *Water*
799 *Resour. Res.*, *46*, W10523, doi:10.1029/2009WR008799.

ter Braak, C. J. F., and J. A. Vrugt (2008), Differential evolution Markov chain with snooker updater and fewer chains, *Stat. Comput.*, *18*, 435–46, doi:10.1007/s11222-008-9104-9.

Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, D. Higdon, B. A. Robinson, J. M. Hyman (2009), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling., *Int. J. Nonlinear Sci. and Numer. Simul.*, *10*, 273–290.

Whittle, P. (1954), On stationary processes in the plane, *Biometrika*, *41*, 439–449.

Zhang, D., and Z. Lu (2004), An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loève and polynomial expansions, *J. Comput. Phys., 194*(2), 773–794.

Zhang, Z., and Y. Rubin (2008), MAD: A new method for inverse modeling of spatial random fields with applications in hydrogeology, *Eos Trans. AGU*, *89*(53), Fall Meet. Suppl., Abstract H44C07.

Zhou, H., J. Gómez-Hernández, and L. Liangping (2014), Inverse methods in hydrogeology: evolution and recent trends, *Adv. Water Resour., 63*, 22–37, doi:10.1016/j.advwatres.2013.10.014.

Table 1. Bounds of the Jeffreys (J), uniform (U) and standard normal (N) prior distributions used in our case study. The last column lists the true values of σ_e and the geostatistical parameters.

Parameter	Units	Prior	Prior range	True value
σ_e	kg m^{-3}	J	0.01 - 0.3	0.039
m	-	U	(-4) - (-2)	-3
v	-	J	0.5 - 2	1
I_M	m	U	0.2 - 2	0.67
A	degree	U	60 - 120	75
R_I	-	U	0.1 - 0.5	0.25
ν	-	J	0.1 - 5	0.5
r	-	N	-	-

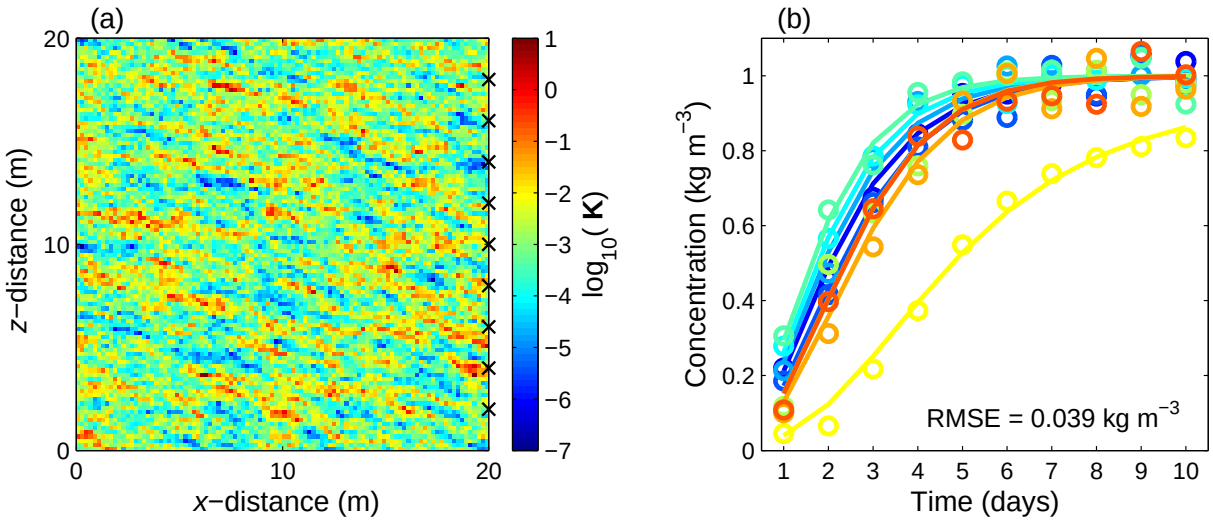

Figure 1. (a) Reference log conductivity field and (b) simulated transport data (b) used in the inversions. Each black cross in the right-hand side borehole denotes a sampling location. Each color in the right plot corresponds to a different sampling location (black cross) in the right-hand side borehole. Lines represent the uncorrupted data and circles signify the noise-contaminated data that are used as measurements for the inversion. The noise level is 0.039 kg m^{-3}.

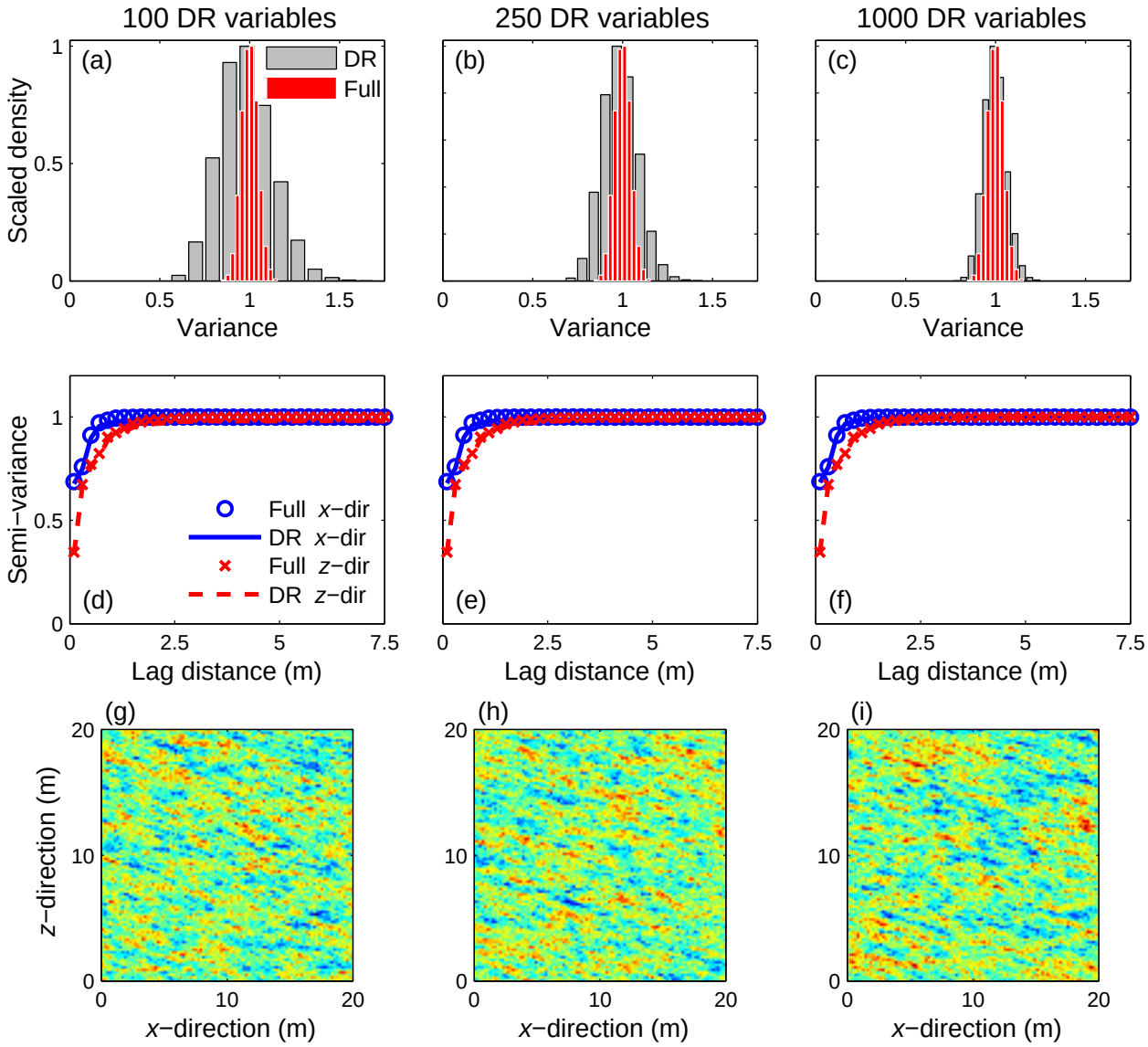

Figure 2. Grid point variance distribution (subplots a, b, c) and experimental variograms in both the horizontal (x-dir) and vertical (z-dir) directions (subplots d, e, f) calculated from 1000 realizations of an autocorrelated multi-Gaussian field with zero-mean and the reference variogram used for the inversions (see last of column Table 1 and Figure 1a), using (a,d) 100-dimensional, (b-e) 250-dimensional and (c-f) 1000-dimensional $\mathbf{r} = [\mathbf{r}_1, \mathbf{r}_2]$ vectors, and randomly chosen field realizations (subplots g-i) derived with (g) 100-dimensional, (h) 250-dimensional and (i) 1000-dimensional $\mathbf{r} = [\mathbf{r}_1, \mathbf{r}_2]$ vectors. The gray and red bins in subplots a, b, and c denote the variance distributions obtained from our dimensionality reduction approach and from directly generating autocorrelated multi-Gaussian fields with the prescribed variogram, respectively. The red histograms in subplots a, b and c thus represent the natural statistical fluctuations. The (plain and dashed) lines and (circle and cross) symbols in subplots d-f signify average experimental variogram values in the x- (blue color) or z-direction (red color) for each of the considered lags, derived from our dimensionality reduction approach (DR) and from directly generating autocorrelated multi-Gaussian fields with prescribed variogram (Full), respectively.

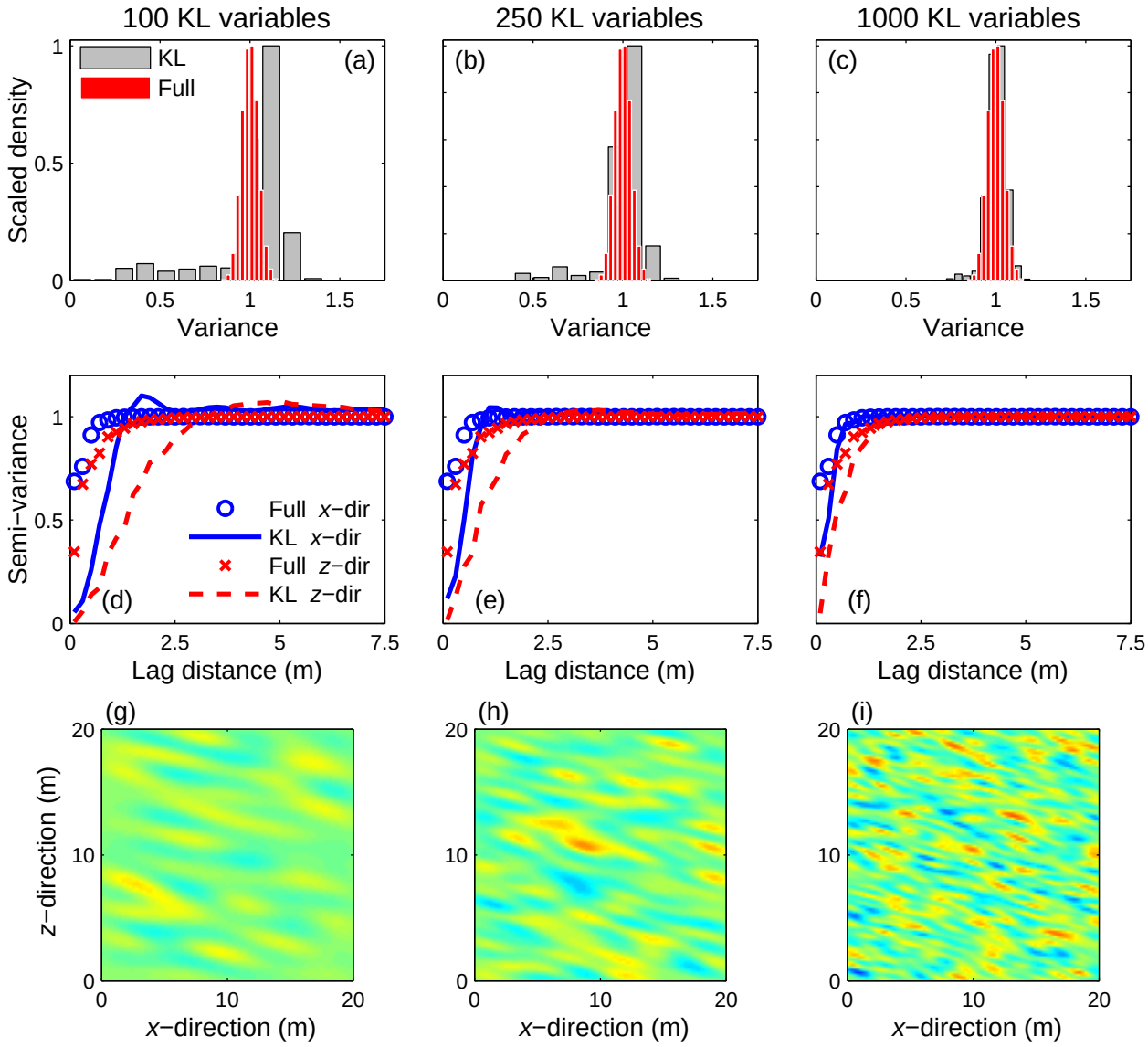

Figure 3. Grid point variance distribution (subplots a, b, c) and experimental variograms in both the horizontal (x-dir) and vertical (z-dir) directions (subplots d, e, f) calculated from 1000 realizations of an autocorrelated multi-Gaussian field with zero-mean and the reference variogram used for the inversions (see last of column Table 1 and Figure 1a), using (a,d) 100, (b-e) 250 and (c-f) 1000 KL variables, and randomly chosen field realizations (subplots g-i) derived with (g) 100, (h) 250 and (i) 1000 KL variables. The gray and red bins in subplots a, b, and c denote the variance distributions obtained from the KL transform and from directly generating autocorrelated multi-Gaussian fields with prescribed variogram, respectively. The red histograms in subplots a, b and c thus represent the natural statistical fluctuations. The (plain and dashed) lines and (circle and cross) symbols in subplots d-f signify average experimental variogram values in the x- (blue color) or z-direction (red color) for each of the considered lags, derived from KL expansion (KL) and from directly generating autocorrelated multi-Gaussian fields with prescribed variogram (Full), respectively.

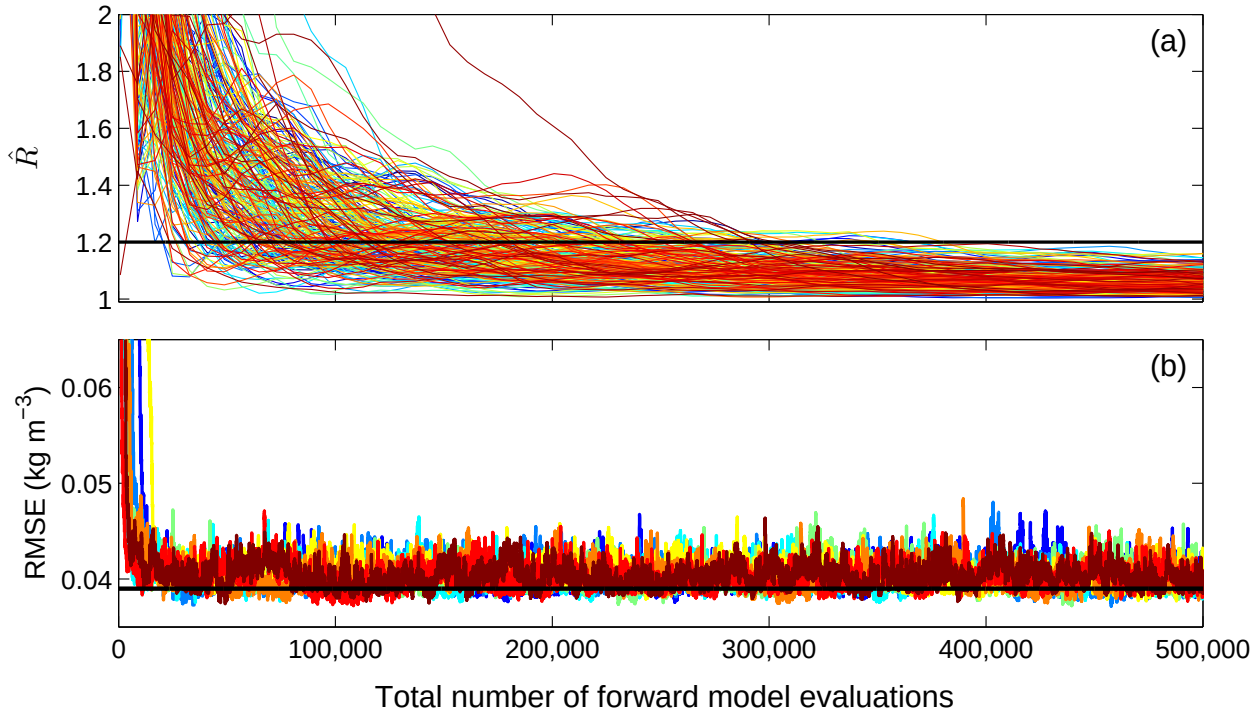

Figure 4. (a) Evolution of the $\hat{R}$ convergence metric [*Gelman and Rubin*, 1992] for the proposed method with 250 DR variables. Each of the 257 sampled parameters is coded with a different color. The chains have officially converged if the plotted lines fall below the horizontal black line. (b) Trace plot of the sampled RMSE values. Each of the eight Markov chains is coded with a different color. The horizontal black line denotes the true RMSE of 0.039 kg m^{-3}.

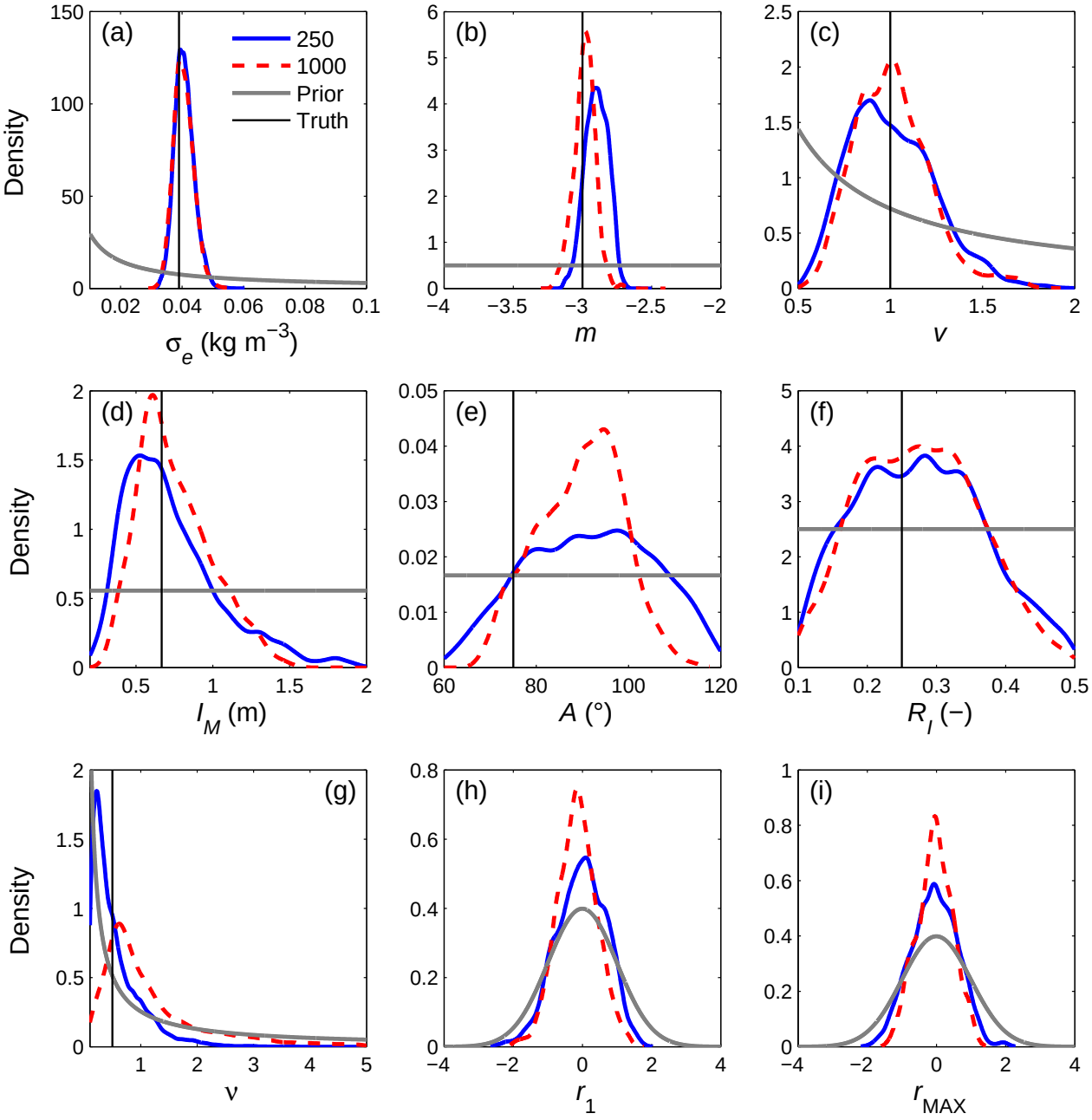

Figure 5. Marginal prior and posterior distributions of (a) the standard deviation of the residual errors, (b-g) the six geostatistical parameters, and (h) the first (r_1), and last (r_{MAX}) elements of the **r** vector for the inversion with 250 (blue lines) and 1000 (dashed red lines) DR variables. The distributions are derived from kernel density smoothing using the last 90% of the samples generated by DREAM$_{(\mathrm{ZS})}$. The parameter values used for creating the reference field are separately indicated with a vertical black line. Since the reference field was generated without dimensionality reduction, there are no true values for r_1 and r_{MAX}. The r_{MAX} parameter is either r_{250} (blue line) or r_{1000} (red dashed line), whereas the respective r_1 and r_{MAX} distributions derived from the inversions with 250 and 1000 DR variables are grouped into the same plots for visual convenience only. There is no reason for these distributions to be similar across inversions.

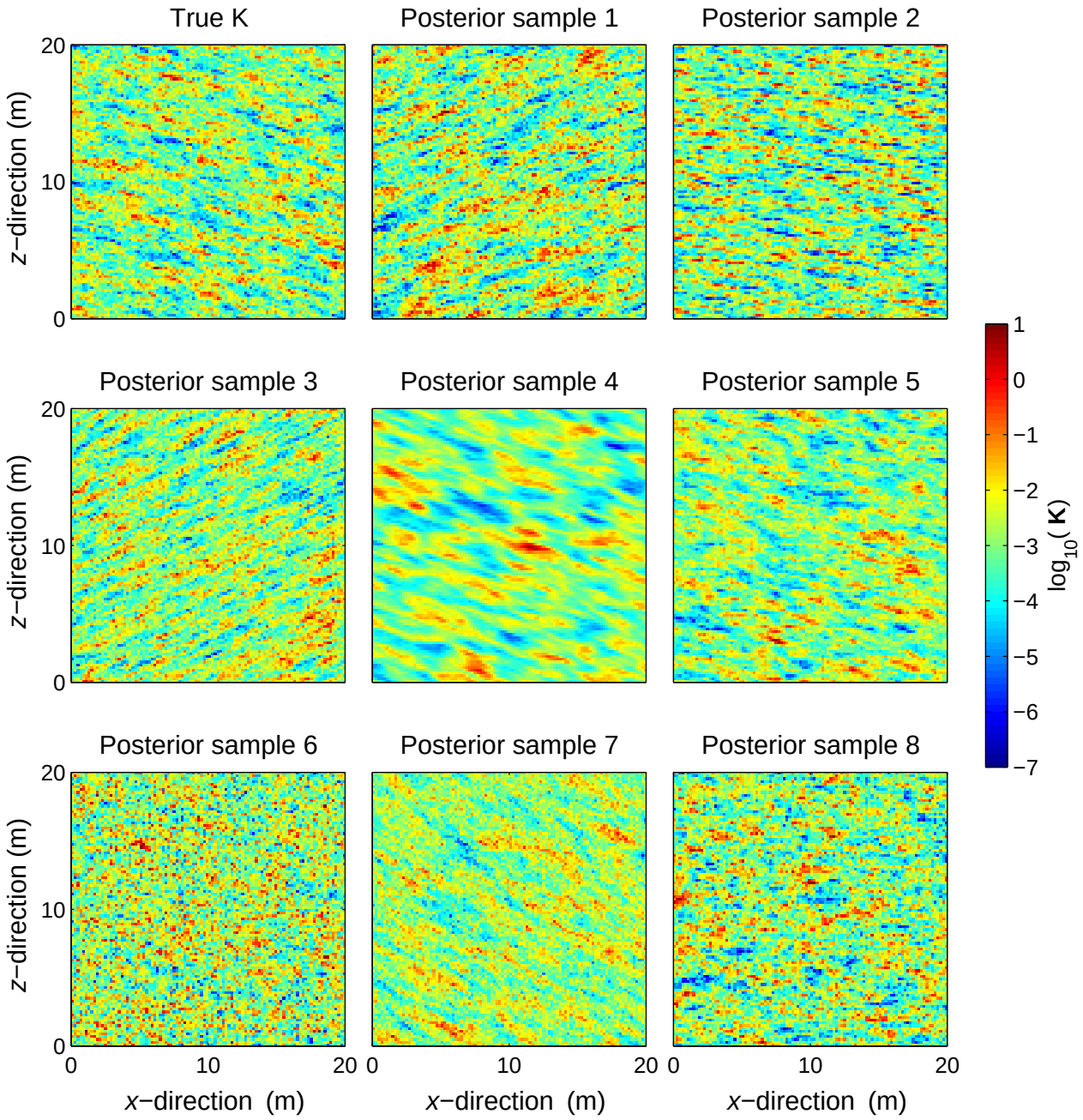

Figure 6. True (reference) model and eight (randomly chosen) realizations of the DREAM$_{(ZS)}$ derived posterior distribution for 250 DR variables.

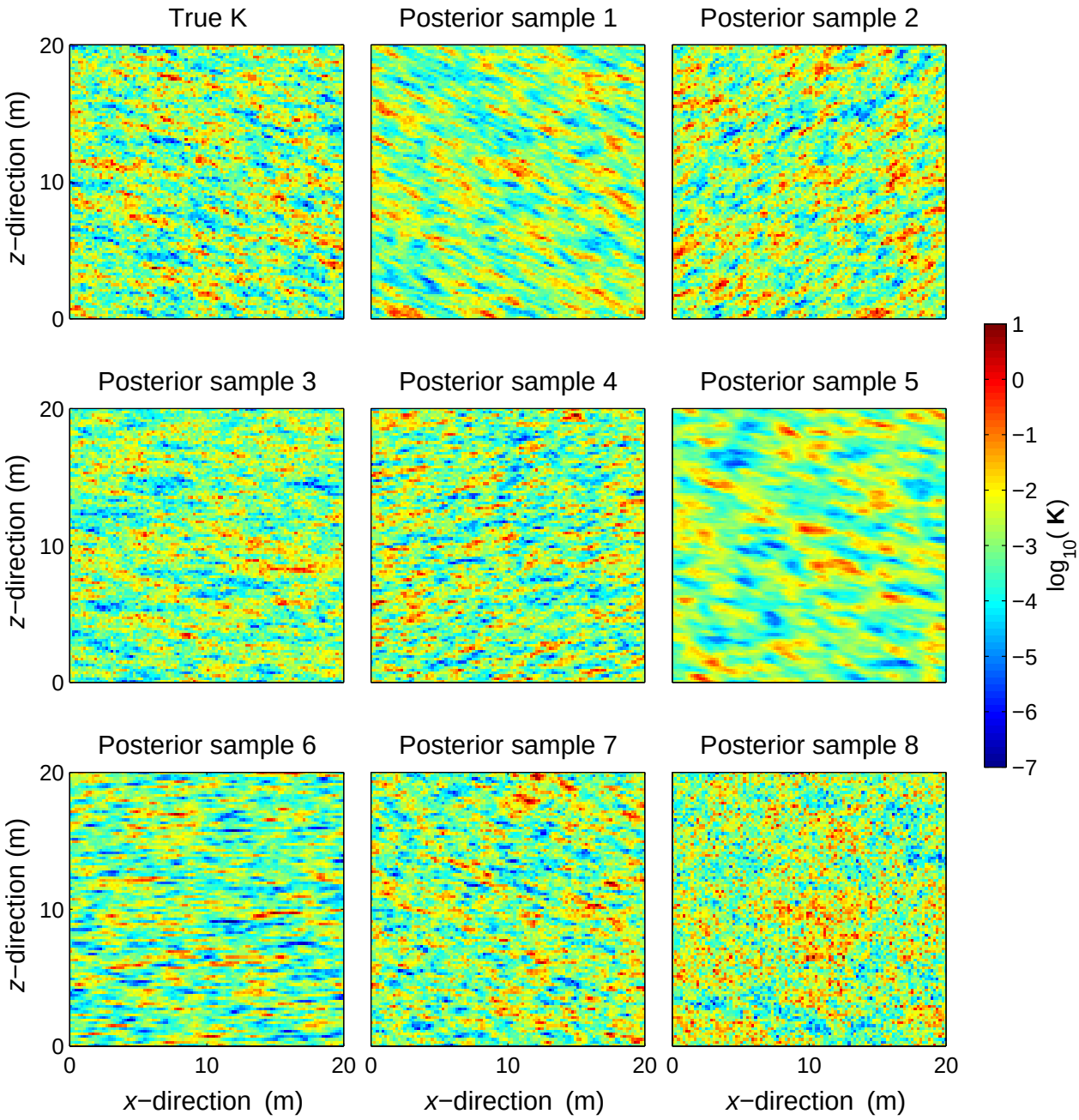

Figure 7. True (reference) model and eight (randomly chosen) realizations of the DREAM$_{(ZS)}$ derived posterior distribution for 1000 DR variables.

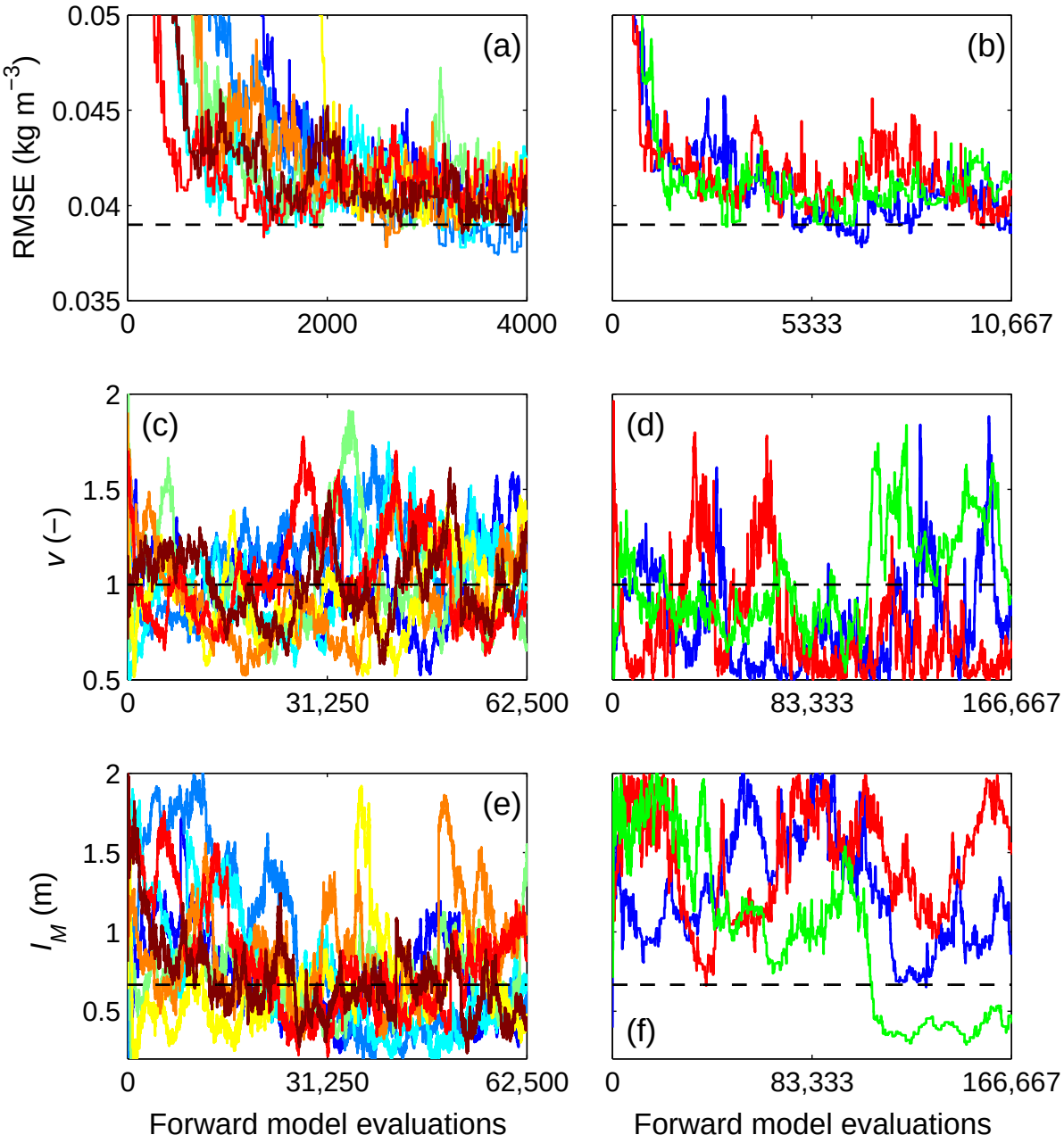

Figure 8. Trace plots of the sampled (a-b) RMSE, (c-d) v and (e-f) I_M values derived from our proposed inversion approach (left column: subplots a, c, and e) and the SGS method (right column: subplots b, d, and f) using 250 DR variables. The dashed horizontal lines depict the true parameter values of the reference log-conductivity field used herein. A visual comparison of the results demonstrates that the DREAM$_{(ZS)}$ (left-hand side) algorithm of the proposed inversion method exhibits an improved searching behavior compared to SGS (right-hand side).

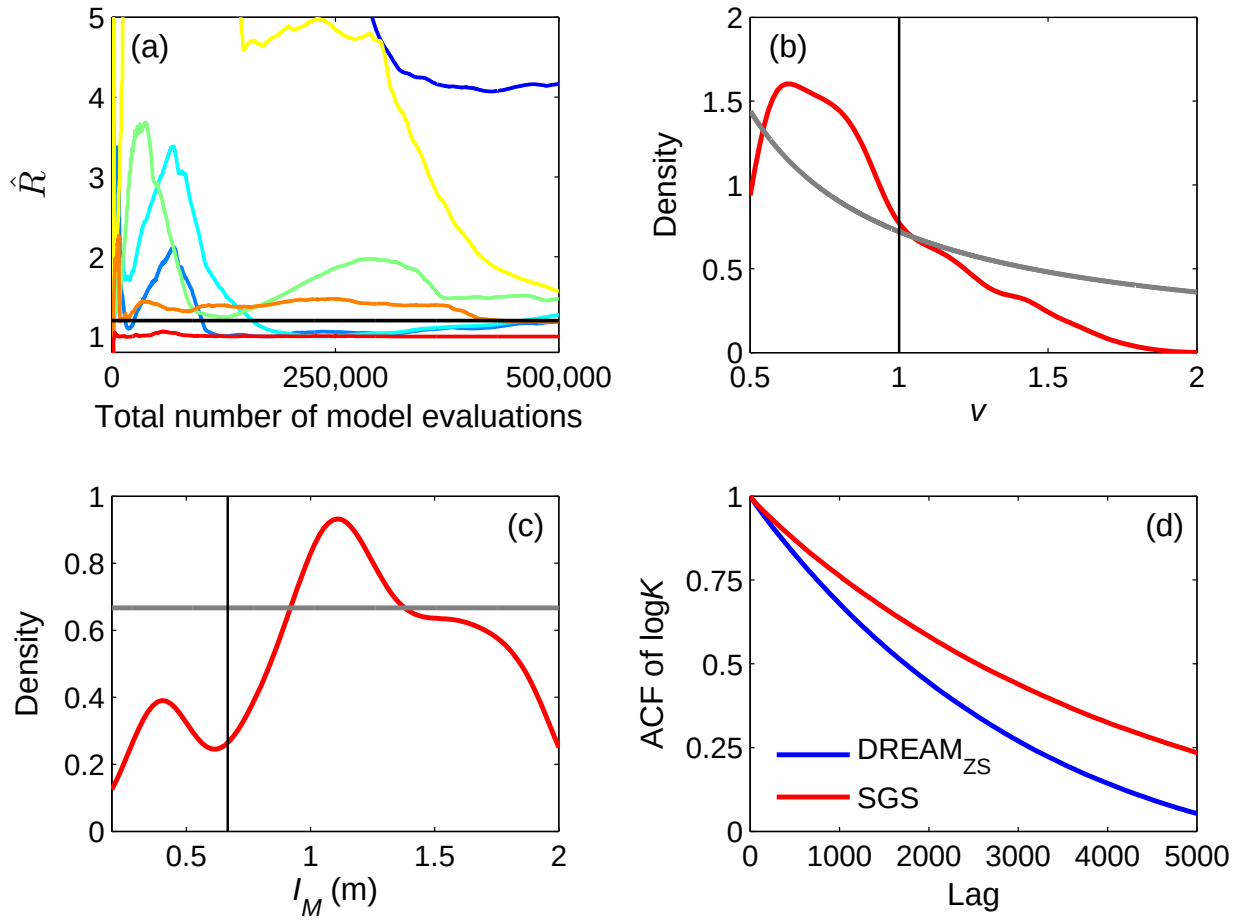

Figure 9. (a) Evolution of the $\hat{R}$ convergence diagnostic [*Gelman and Rubin*, 1992] for the total computational budget of 500,000 model evaluations of the three SGS trials. We only plot traces (color coded) of the six geostatistical parameters and σ_e. (b) and (c) Marginal posterior distributions (red lines) of v and I_M derived from the last 90% of the samples generated with the three different SGS trials. Kernel density smoothing is applied. The prior distribution is indicated by a gray line, and the true values by the vertical black lines. (d) Mean autocorrelation function (ACF) of the 10,000 log-conductivity grid values derived from DREAM$_{(ZS)}$ (blue line) or SGS (red line) for lags 0 to 5000. The lag-k autocorrelation is defined as the correlation between draws k lags apart. Listed statistics are computed for the last 50,000 log-conductivity fields sampled in the eight DREAM$_{(ZS)}$ chains or the three independent SGS chains. The average of the chains is presented.

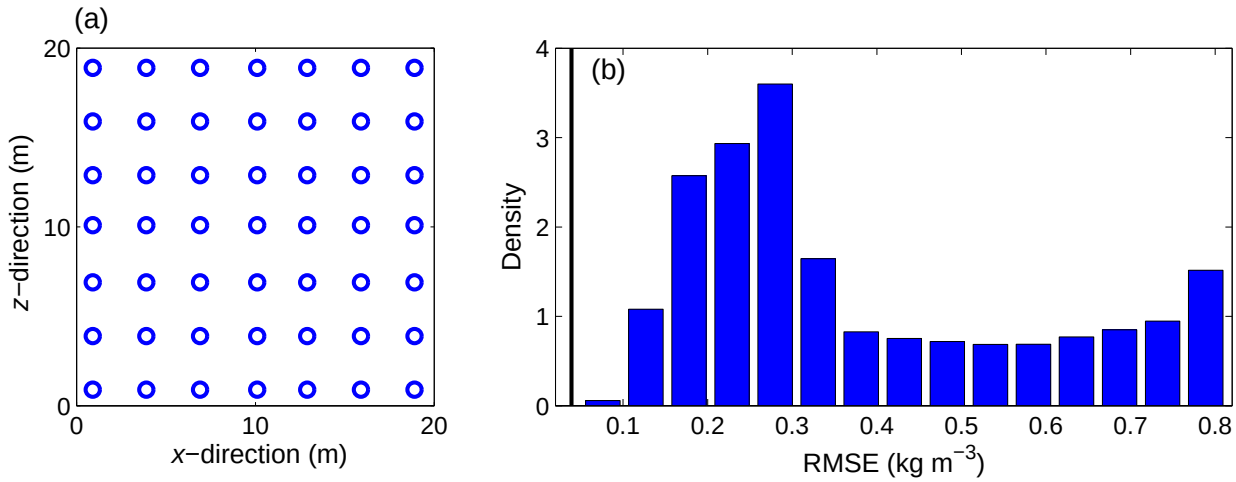

Figure 10. (a) Anchor locations and (b) RMSE distribution of the 600,000 forward runs performed by the MAD trial. The vertical black line in plot (b) denotes the true RMSE value of 0.039 kg m^{-3}. The RMSE values sampled with MAD are much larger than the true value. This simply demonstrates that the method cannot converge properly within the assigned computational budget of 600,000 model evaluations.