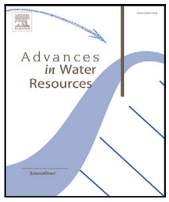
Advances in Water Resources 173 (2023) 104402

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres



Inference of geostatistical hyperparameters with the correlated pseudo-marginal method

Lea Friedli [a,*], Niklas Linde [a], David Ginsbourger [b], Alejandro Fernandez Visentini [a,c], Arnaud Doucet [d]

[a] Institute of Earth Sciences, University of Lausanne, Lausanne, Switzerland
[b] Institute of Mathematical Statistics and Actuarial Science, University of Bern, Bern, Switzerland
[c] Bureau de Recherches Géologiques et Minières, Orléans, France
[d] Department of Statistics, Oxford University, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Geostatistics
Hyperparameters
Bayesian inversion
Latent variable model
Likelihood estimation
Non-ergodicity

ABSTRACT

We consider non-linear Bayesian inversion problems targeting the geostatistical hyperparameters of a random field describing hydrogeological or geophysical properties given hydrogeological or geophysical data. This problem is of particular importance in the non-ergodic setting as there are no analytical upscaling relationships linking the data to the hyperparameters, such as, mean, standard deviation, and integral scales. Full inversion of the hyperparameters and the local properties of the field (typically involving many thousands of unknowns) brings substantial computational challenges, such that simplifying model assumptions (e.g., homogeneity or ergodicity) are typically made. To prevent the errors resulting from such simplified assumptions while also circumventing the burden of high-dimensional full inversions, we use a pseudo-marginal Metropolis–Hastings algorithm that treats the random field as latent variables. In this random effects model, the intractable likelihood of observing the data given the hyperparameters is estimated by Monte Carlo averaging over realizations of the random field. To increase the efficiency of the method, low-variance approximations of the likelihood ratio are obtained by using importance sampling and by correlating the samples used in the proposed and current steps of the Markov chain. We assess the performance of this correlated pseudo-marginal method by considering two representative inversion problems involving diffusion-based and wave-based physics, respectively, in which we infer the hyperparameters of (1) hydraulic conductivity fields using apparent hydraulic conductivity data in a data-poor setting and (2) fracture aperture fields using borehole ground-penetrating radar (GPR) reflection data in a more data-rich setting. For the first test case, we find that the correlated pseudo-marginal method generates similar estimates of the geostatistical mean as classical rejection sampling, while an inversion assuming ergodicity provides biased estimates. For the second test case, we find that the correlated pseudo-marginal method estimates the hyperparameters well, while rejection sampling is computationally unfeasible and a simplified model assuming homogeneity leads to biased estimates.

1. Introduction

The scale dependence of most environmental processes poses significant challenges for hydrogeological and geophysical modeling (e.g., Klemeš, 1983; Blöschl and Sivapalan, 1995). The governing partial differential equations (PDEs) traditionally employed to describe fluid flow, chemical or electrical transport (Neuman and Di Federico, 2003) are solved at some support volume scale assumed to be a "Representative Elementary Volume" (REV; Hill, 1963). That is, it is assumed that smaller-scale heterogeneity averages out and can be represented (with regard to the process under consideration) by averaged physical or chemical properties. In practice, the conditions necessary for the existence of a REV are often not met because geological media exhibit heterogeneity over a wide range of scales (Neuman and Di Federico, 2003). Errors occurring when only partially accounting for or ignoring heterogeneity generally grow with the non-linearity of the physical or chemical process under study and can result in misleading predictions (e.g., Dentz et al., 2011; Yu and Michael, 2021). For this reason, it is essential to characterize and account for the statistical properties of small-scale heterogeneity even when targeting mean properties.

* Correspondence to: Géopolis, Quartier Mouline, 1015 Lausanne, Switzerland.
E-mail address: lea.friedli@unil.ch (L. Friedli).

https://doi.org/10.1016/j.advwatres.2023.104402
Received 21 March 2022; Received in revised form 3 February 2023; Accepted 3 February 2023
Available online 14 February 2023

We consider non-linear inversion problems targeting geostatistical hyperparameters (e.g., mean, standard deviation, integral scale and anisotropy factor) of a random field describing hydrogeological or geophysical properties given indirect data. This problem setting is applicable when the main properties of interest are the hyperparameters and not the local field properties. The geostatistical literature is full of studies (e.g., Rehfeldt et al., 1992; Hess et al., 1992; Bohling et al., 2016) focusing on hyperparameter estimation based on direct data (e.g., permeability data along boreholes), but much less work has considered indirect data (e.g., pressure data, tracer breakthrough data) as in the present study. In what follows, we only discuss this latter case. One of the first approaches considering unknown hyperparameters in such an inversion setting was the quasi-linear geostatistical approach by Kitanidis (1995), which optimizes the hyperparameters along with the spatial field. Another approach enabling joint inference of a Gaussian random field and its variogram parameters relied on so-called sequential Gibbs sampling (Hansen et al., 2012; Hansen et al., 2013a; Hansen et al., 2013b). Zhao and Luo (2021) applied an iterative approach based on principal components which is updating biased or unknown hyperparameters while solving a non-linear inversion problem. Recently, Wang et al. (2022) proposed an hierarchical Bayesian inversion targeting first global variables (such as hyperparameters but also physical variables) and later the posterior of the whole field (referred to as spatial variables). Note that none of these studies focus on inferring the hyperparameters only.

We rely on a Bayesian framework and infer the hyperparameters' posterior probability density function (PDF) given indirect hydrogeological or geophysical measurements. To sample from the posterior, we apply a Markov chain Monte Carlo (MCMC) method building on the Metropolis–Hastings algorithm (MH; Hastings, 1970; Metropolis et al., 1953). The basic procedure of the MH algorithm in this setting is to propose iteratively a new set of hyperparameters, which are then accepted or rejected based on their prior probabilities and likelihoods. We consider synthetic experimental setups in which the hydrogeological or geophysical data average over a random field realization that is either ergodic or non-ergodic. A random field must be stationary to be ergodic, but not vice versa. Stationarity implies that the distribution does not change with position. Ergodicity, on the other hand, implies that the field realization is much larger than the characteristic scale of heterogeneity. By the so-called ergodic setting, we consider data that average over a scale that is much larger than the field's scale of heterogeneity such that the effects of small-scale fluctuations average out. Consequently, the data do not depend on the local properties of a given random field realization, but on the hyperparameters only. By the non-ergodic setting, we refer to cases when the data averaging takes place over a scale that is smaller or comparable to the scale of heterogeneity. This implies that the data depend not only on the hyperparameters but also on the random field realization on which measurements are made. That is, variations between field realizations in terms of magnitudes and locations of high and low property values lead to different data responses as the fluctuations do not average out. Broadly speaking, such behavior is expected when the physical response is averaging over length scales that are less than some ten correlation lengths of the parameter field. In the non-ergodic setting, there are no analytical upscaling relationships linking the data to the hyperparameters of interest. If relationships assuming ergodicity or assumptions of homogeneity are employed in such a case, bias is likely to occur in the inferred hyperparameters (e.g., Visentini et al., 2020; Shakas and Linde, 2017). We suggest that most measurements in hydrogeology and geophysics take place in such a non-ergodic setting.

Equivalent properties derived from measurements of one type of physics (e.g., the equivalent aperture describing fluid flow) generally do not represent the equivalent property for another type of physics (e.g., the equivalent aperture of thermal transport; e.g., Tsang, 1992). This disparity occurs as soon as the underlying physics is non-linear, implying for instance that equivalent mean properties do not correspond to arithmetic mean properties (e.g., Jougnot et al., 2018; Shakas and Linde, 2015). One solution to this problem that is pursued in the present study is to instead infer hyperparameters while accounting for small-scale heterogeneity. In this way, it is possible to use estimates derived from one type of physics to make predictions for another type of physics. In many ergodic settings, upscaling theory provides relevant relationships between hyperparameters and equivalent properties (e.g., Renard and De Marsily, 1997; Torquato and Haslach, 2002; Sanchez-Vila et al., 2006), while no such relationships are available in the non-ergodic setting.

One way to infer hyperparameters in the non-ergodic setting by MCMC methods is to parameterize the field by hyperparameters and white noise to describe the local properties (as e.g. in Laloy et al., 2015, Hunziker et al., 2017 and Xiao et al., 2021). The corresponding full inversion problem involves typically many thousands of parameters, for which either an efficient MH proposal scheme has to be designed (e.g., Xiao et al., 2021) or dimensionality reduction arguments have to be invoked (e.g., Laloy et al., 2015, Rubin et al., 2010). While the first approach is very challenging (curse of dimensionality, e.g., Robert et al., 2018), the second approach may lead to biased estimates (Laloy et al., 2015). An example of the application of dimensionality reduction relevant to the current study is Shakas et al. (2018) who inferred fracture aperture distribution and geometry by combining GPR forward modeling with flow-and-transport simulations. Even if this study provided reasonable estimates of the statistical properties, it was plagued by a low acceptance rate, slow mixing of the chains and no formal convergence despite a large number of iterations.

Instead of a full inversion, we here target the hyperparameters of interest only. Since the local properties of the field influence the observations in the non-ergodic setting, the field is considered a latent (unobservable) variable. Due to the random effect the unobservable field has on the data, we speak of a random effects model. To implement a MH algorithm inferring the hyperparameters only, we have to evaluate the likelihood of observing the data given the currently proposed set of hyperparameters. In a random effects model, this likelihood has generally no analytical form and is, therefore, referred to as intractable. The pseudo-marginal (PM) method introduced by Beaumont (2003) and studied by Andrieu and Roberts (2009) relies on an unbiased estimator of this intractable likelihood function that is based on averaging over Monte Carlo samples of the latent variables. This implies that after proposing a new set of hyperparameters, different field realizations with the same hyperparameters are sampled. Then, the likelihood of each field realization can be calculated and the intractable likelihood function is estimated by averaging over the obtained values. Beaumont (2003) demonstrates that using such a non-negative and unbiased estimator of the likelihood within the MH algorithm results in an algorithm that draws samples from the same target distribution as when using the true likelihood. In the PM method, a high variance of the log-likelihood ratio estimator has a very strong adverse impact on performance, but achieving a low variance often comes at the price of using an excessive number of samples in the Monte Carlo averaging. To obtain an efficient algorithm balancing these two aspects, it has been shown that the standard deviation of the log-likelihood estimator should be around 1.2–1.5 (Doucet et al., 2015). This can be ensured by (1) properly choosing the number of samples used in the Monte Carlo averaging and by (2) applying importance sampling to draw the realizations of the latent variables. In the context of state-space models, the number of samples has to increase linearly with the number of observations, which is computationally impractical in data-rich settings (Deligiannidis et al., 2018). To address this problem, the correlated pseudo-marginal (CPM) by Deligiannidis et al. (2018) correlates the draws of latent variables between two subsequent iterations, thereby, reducing the number of Monte Carlo draws needed to ensure low-variance log-likelihood ratio approximations.

The pseudo-marginal and correlated pseudo-marginal methods have hardly been studied in hydrogeological and geophysical settings. In Friedli et al. (2022), the CPM method was shown to outperform other competing approaches to lithological tomography (Bosch, 1999), in which geophysical data are used to directly infer (hydro)geological properties of interest. Friedli et al. (2022) considered a very high dimensionality of the target and latent variables under the assumption of known hyperparameters. Here, the interest is instead placed on inferring few hyperparameters while accounting for the effects of thousands of latent variables. This leads to a very different model setting and study objectives than Friedli et al. (2022). We assess the performance of the CPM method with two synthetic test cases in which we infer the hyperparameters describing (1) hydraulic property fields using equivalent (apparent) hydraulic conductivity data and (2) fracture aperture fields using borehole ground-penetrating radar (GPR) reflection data. The two test cases are chosen to be representative for transmission problems governed by diffusion (e.g., groundwater flow, heat transport, electrical conduction) and reflection problems governed by wave-based physics (e.g., GPR, seismics and acoustics). In the first test case, we consider a very data-poor setting and are mainly interested in the geostatistical mean of the field. By comparing the CPM results with those of an MH algorithm that replaces the forward solver with an analytical upscaling relationship that assumes ergodicity, we show that assuming a simplified model can lead to strongly biased estimates of the hyperparameters in the non-ergodic setting. We also demonstrate that the CPM results are in agreement with those obtained by rejection sampling, which is computationally feasible for this very data-poor example. In the second test case, we consider much more data and show that the CPM method provides accurate estimates of the geostatistical mean and other hyperparameters. Additionally, we show how these hyperparameters describing aperture properties inferred from GPR data allow us to predict fracture transmissivity.

This paper is structured as follows. Section 2 introduces the CPM methodology in the considered context. Section 3 presents the first test case based on measurements across a hydraulic conductivity field and Section 4 presents the second test case in which borehole GPR data are used to infer the hyperparameters of fracture aperture fields. This is followed by a discussion in Section 5 and conclusions in Section 6.

2. Methodology

The methodology section starts by presenting the considered random effects model and the chosen notation of Gaussian random fields (Section 2.1). Bayesian inference and MCMC algorithms are then described (Section 2.2) before introducing the correlated pseudo-marginal method (Section 2.3) and giving a brief introduction into rejection sampling (Section 2.4). It ends with a description of the performance assessment metrics used to evaluate the results (Section 2.5).

2.1. Random effects model

We are interested in a random field describing hydrogeological or geophysical property distributions. A random field (spatial stochastic process) $\boldsymbol{X}(d,\omega)$ with $\omega \in \Omega$ is a family of random variables indexed by the spatial location $d \in \mathcal{D} \subset \mathbb{R}^2$ (Chiles and Delfiner, 2009). For fixed $\omega = \omega_0$, $\boldsymbol{X}(\cdot,\omega_0)$ is a realization of the random field with ω referring to the "randomness" of the field. For a fixed location $d = d_0$, $\boldsymbol{X}(d_0,\cdot)$ is a real-valued random variable. For simplicity, in the following we write $\boldsymbol{X}(\cdot)$ to indicate $\boldsymbol{X}(\cdot,\omega)$. The "true" hydrogeological or geophysical property field is considered a realization of the underlying random field. We are interested in inferring the hyperparameters $\boldsymbol{\theta}$ parameterizing the geostatistical distribution of the random field $\boldsymbol{X}(\cdot)$.

We consider a Gaussian random field (GRF) $\boldsymbol{X}(\cdot)$ for which all finite-dimensional distributions are multivariate Gaussians (Chiles and Delfiner, 2009). Its distribution is determined by the mean and the covariance function. We assume the mean function $\mu_{\boldsymbol{\theta}}(\cdot)$ to be constant even if it would be straightforward to employ a non-stationary function. For the covariance function $C_{\boldsymbol{\theta}}(\cdot,\cdot)$, we apply the powered exponential expressed here in isotropic form:

$$C_{\boldsymbol{\theta}}(b,b') = \sigma^2 exp\left(-\left(\frac{\|b-b'\|}{I}\right)^{2H}\right), \tag{1}$$

whereby $\|b\| = \sqrt{b^T b}$ denotes the Euclidean norm, σ the standard deviation, I the integral scale and H the Hurst exponent (with $0 < H \leq 1$). For $H = 0.5$, the powered exponential covariance function reduces to the classical exponential covariance function and for $H = 1$ to the Gaussian (squared exponential) covariance function. We also consider geometric anisotropy (e.g., Chiles and Delfiner, 2009), for which the covariance depends not only on the Euclidean distance but also on the direction between the considered positions. We assume a known anisotropy angle of 90 degrees and refer to the integral scale in the vertical direction as I^y, which, multiplied by the anisotropy factor λ, gives the integral scale in the horizontal direction I^x.

To infer the P hyperparameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_P)$, we have access to T measurements $\boldsymbol{y} = (y_1, y_2, \dots, y_T)$. As generally there exists no upscaling relationship linking the hyperparameters to the measurements, we formulate the problem with a random effects model using the latent random field $\boldsymbol{X}(\cdot)$:

$$\boldsymbol{X} \sim f_{\boldsymbol{\theta}}(\cdot) \quad \boldsymbol{Y}|\boldsymbol{X} \sim g_{\boldsymbol{\theta}}(\cdot|\boldsymbol{X}). \tag{2}$$

For the latent random field $\boldsymbol{X}(\cdot)$ we use a discretized representation on a $(D \times D)$-grid, whereby we assume the grid cells to be representative elementary volumes (REV) for the governing physical process. We consider a setting in which the number of target hyperparameters P is much smaller than the number of latent variables (grid cells) D^2. The measurements are described by the random variable $\boldsymbol{Y} = \mathcal{G}(\boldsymbol{X}) + \varepsilon_{\boldsymbol{\Theta}}$ with $\mathcal{G} : \mathbb{R}^{D^2} \rightarrow \mathbb{R}^T$ denoting the physical forward solver and $\varepsilon_{\boldsymbol{\Theta}}$ the observational noise. While $\boldsymbol{Y}$ refers to the random variable, $\boldsymbol{y}$ denotes the "true" measurements considered to be a realization of $\boldsymbol{Y}$.

Assuming the latent random field to be Gaussian, we write $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \varphi_{D^2}(\boldsymbol{x}; \boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ with $\varphi_{D^2}(\cdot; \boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ denoting the PDF of a D^2-variate normal distribution with mean vector $\boldsymbol{\mu}_{\boldsymbol{\theta}} = (\mu_{\boldsymbol{\theta}}(b_i))_{1 \leq i \leq D^2}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = (C_{\boldsymbol{\theta}}(b_i, b_j))_{1 \leq i,j \leq D^2}$ specified by the hyperparameters $\boldsymbol{\theta}$. Furthermore, we assume the observational noise $\varepsilon_{\boldsymbol{\Theta}}$ to be Gaussian, such that $\boldsymbol{Y}|\boldsymbol{X} \sim g_{\boldsymbol{\theta}}(\cdot|\boldsymbol{X})$ is distributed with the PDF $g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}) = \varphi_T(\boldsymbol{y}; \mathcal{G}(\boldsymbol{x}), \boldsymbol{\Sigma}_{\boldsymbol{Y}})$, with $\boldsymbol{\Sigma}_{\boldsymbol{Y}}$ being a diagonal matrix with the variance of the observational noise on its diagonal. To generate a realization of the D^2-dimensional GRF $\boldsymbol{X}(\cdot)$ with mean vector $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, we rely on a pixel-based parameterization,

$$\boldsymbol{X} = \boldsymbol{\mu}_{\boldsymbol{\theta}} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{1/2} \boldsymbol{Z}, \tag{3}$$

with $\boldsymbol{Z}$ denoting a D^2-dimensional random vector consisting of $i.i.d.$ standard normal distributed variables.

2.2. Bayesian inference with Markov chain Monte Carlo

Bayes' theorem specifies the posterior PDF $p(\boldsymbol{\theta}|\boldsymbol{y})$ of the model parameters $\boldsymbol{\theta}$ conditioned on the measurements $\boldsymbol{y}$ as,

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})} \propto p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta}), \tag{4}$$

where $p(\boldsymbol{\theta})$ denotes the prior PDF of the model parameters, $p(\boldsymbol{y}|\boldsymbol{\theta})$ the likelihood function and $p(\boldsymbol{y})$ the evidence (assumed positive). If there is no analytical form of the posterior PDF but it is possible to evaluate the unnormalized entity for some value of $\boldsymbol{\theta}$, MCMC methods (see, e.g., Robert and Casella, 2013) can be applied to generate realizations drawn proportionally from the posterior PDF. The basic procedure behind MCMC algorithms is to propose new values for the target parameters, which are then accepted or rejected with a given probability. The Metropolis–Hastings (MH; Metropolis et al., 1953; Hastings, 1970) method is a well-known example. At iteration j, it proceeds as follows:

First, new values for the target parameters $\theta^{(j)}$ are proposed using the model proposal density $q(\cdot|\theta^{(j-1)})$. Then, the acceptance probability,

$$\alpha_{MH}\left(\theta^{(j-1)},\theta^{(j)}\right) = \min\left\{1, \frac{q(\theta^{(j-1)}|\theta^{(j)})p(\theta^{(j)}|\boldsymbol{y})}{q(\theta^{(j)}|\theta^{(j-1)})p(\theta^{(j-1)}|\boldsymbol{y})}\right\} \tag{5}$$

$$= \min\left\{1, \frac{q(\theta^{(j-1)}|\theta^{(j)})p(\theta^{(j)})p(\boldsymbol{y}|\theta^{(j)})}{q(\theta^{(j)}|\theta^{(j-1)})p(\theta^{(j-1)})p(\boldsymbol{y}|\theta^{(j-1)})}\right\}, \tag{6}$$

is calculated and the proposed $\theta^{(j)}$ is accepted (if $\alpha_{MH}(\theta^{(j-1)},\theta^{(j)}) \geq V$) or rejected (if $\alpha_{MH}(\theta^{(j-1)},\theta^{(j)}) < V$) on the basis of a draw of a uniformly distributed random variable $V \sim \text{Unif}([0,1])$. If the proposed $\theta^{(j)}$ is rejected, the MCMC chain remains at the old position ($\theta^{(j)} = \theta^{(j-1)}$).

In order to evaluate the acceptance probability in Eq. (6), the value of the likelihood function $\theta \mapsto p(\boldsymbol{y}|\theta^{(j)})$ has to be calculated. In a random effects model (see Section 2.1), the likelihood function is given by,

$$p(\boldsymbol{y}|\theta) = \int g_\theta(\boldsymbol{y}|\boldsymbol{x}) f_\theta(\boldsymbol{x}) d\boldsymbol{x}. \tag{7}$$

This integral often does not admit an analytical form making the direct implementation of the MH algorithm impossible and specific algorithms such as the correlated pseudo-marginal method are needed (outlined in Section 2.3 below).

2.2.1. MCMC proposal scheme

To achieve an efficient MCMC algorithm, one needs a suitable proposal density $q(\cdot|\theta^{(j-1)})$. Even in an inversion targeting only few parameters, one has to choose the direction and size of the model proposal steps carefully. Too large steps lead to a low acceptance rate, while too small steps lead to very slow exploration of the target space; both of these situations lead to an algorithm needing an unnecessarily large number of iterations until convergence (see Section 2.5 below for the assessment of convergence).

To generate model proposals, we apply the adaptive Metropolis algorithm of Haario et al. (2001), in which the covariance matrix describing the Gaussian proposal distribution is updated during the MCMC run. Despite the adaptation, the algorithm is ensured to be ergodic, although not Markovian (Haario et al., 2001). The Gaussian proposal distribution at iteration j is expressed as $q(\theta^{(j)}|(\theta^{(0)},\theta^{(1)},\dots\theta^{(j-1)})) = \varphi_d(\theta^{(j-1)},\mathbf{C}^{(\mathbf{j})})$, with

$$\mathbf{C}^{(\mathbf{j})} = \begin{cases} \mathbf{C}^{(\mathbf{0})} & j \leq j_0 \\ s_P(Cov(\theta^{(0)},\theta^{(1)},\dots,\theta^{(j-1)}) + \epsilon\mathbb{I}_P) & j > j_0, \end{cases} \tag{8}$$

denoting the evolving covariance matrix. During the first j_0 iterations, the method uses an initial covariance matrix $\mathbf{C}^{(\mathbf{0})}$ selected according to available prior knowledge. After this initial period, the covariance matrix is updated with $\mathbf{C}^{(\mathbf{j})} = s_P(cov(\theta^{(0)},\theta^{(1)},\dots,\theta^{(j-1)}) + \epsilon\mathbb{I}_P)$, where s_P is a parameter depending on the dimension of the target space (Haario et al. (2001) use $s_P = (2.4)^2/P$ as in Gelman et al., 1996), $\epsilon > 0$ is a small constant and $\mathbb{I}_P$ denotes the identity matrix of dimension P. To ensure an efficient calculation, Haario et al. (2001) use the recursion formula,

$$\mathbf{C}^{(\mathbf{j+1})} = \frac{(j-1)}{j}\mathbf{C}^{(\mathbf{j})} + \frac{s_P}{j}\left(j\overline{\theta^{(j-1)}}\;\overline{\theta^{(j-1)T}} - (j+1)\overline{\theta^{(j)}}\;\overline{\theta^{(j)T}} + \theta^{(j)}\theta^{(j)T} + \epsilon\mathbb{I}_P\right), \tag{9}$$

with $\overline{\theta^{(j)}} = 1/(j+1)\sum_{i=0}^{j}\theta^{(i)}$ and $\theta^{(i)}$ considered to be column vectors.

For a target parameter θ_i with bounded support $[a,b]$, one has to make sure that the proposed value lies within the considered interval. Therefore, we apply fold boundary handling implying that a proposal which passes one boundary of the support is re-entered through the other boundary (Vrugt, 2016), that is, similar to periodic boundary conditions in numerical simulations.

2.3. Pseudo-marginal and correlated pseudo-marginal method

In Section 2.2, we explained that the considered random effects model has an intractable likelihood function. The pseudo-marginal and correlated pseudo-marginal methods presented below provide a solution to this in the form of Monte Carlo estimations of the likelihood function. To illustrate the presented concepts, a flow chart describing the basic procedure of the correlated pseudo-marginal method is depicted in Fig. 1.

2.3.1. Pseudo-marginal method

A MH algorithm employing a non-negative unbiased estimator of the likelihood function samples realizations of the same target distribution as one using the true likelihood (Beaumont, 2003). To exploit this remarkable property, Beaumont (2003) proposes a MH algorithm estimating, at each iteration, an intractable likelihood function using Monte Carlo averaging over samples of the latent variables. This approach was termed the pseudo-marginal (PM) method and analyzed theoretically by Andrieu and Roberts (2009). The efficiency of the PM method depends mainly on the variability of the likelihood estimator. When only one brute force Monte Carlo sample of the latent variables is used to estimate the likelihood, the algorithm is likely to suffer from a low acceptance rate caused by the high variability of the log-likelihood estimator. This happens when the likelihood estimator can take very different values for different realizations of the latent variables. In our setting, this is the case if different local properties of the latent random field $\boldsymbol{X}(\cdot)$ lead to very different data responses even if the hyperparameters of the fields are the same. The variance of the log-likelihood estimator can be reduced by (1) using many samples of the latent variables and (2) selecting a well-working importance sampling (IS; e.g. Owen and Zhou, 2000) scheme to draw them from. The PM method proposes the following unbiased estimator for the likelihood $p(\boldsymbol{y}|\theta)$ of Eq. (7),

$$\hat{p}_N(\boldsymbol{y}|\theta) = \frac{1}{N}\sum_{n=1}^{N} w_\theta(\boldsymbol{y}|\boldsymbol{X}_n), \quad \text{with} \quad w_\theta(\boldsymbol{y}|\boldsymbol{X}_n) = \frac{g_\theta(\boldsymbol{y}|\boldsymbol{X}_n)f_\theta(\boldsymbol{X}_n)}{m_\theta(\boldsymbol{X}_n)}, \tag{10}$$

where $\boldsymbol{X}_n \overset{i.i.d}{\sim} m_\theta(\cdot)$ for $n = 1,2,\dots,N$ with $m_\theta(\cdot)$ denoting the importance density function.

To derive the importance density $\boldsymbol{x} \mapsto m_\theta(\cdot)$, we follow the approach of Friedli et al. (2022). Therefore, we choose a distribution which is nearly proportional to $\boldsymbol{x} \mapsto g_\theta(\boldsymbol{y}|\boldsymbol{x})f_\theta(\boldsymbol{x})$ (see e.g., Owen and Zhou, 2000 referring to the results of Kahn and Marshall, 1953). Since it holds that $p(\boldsymbol{x}|\theta,\boldsymbol{y}) \propto g_\theta(\boldsymbol{y}|\boldsymbol{x})f_\theta(\boldsymbol{x})$, we approximate the importance density with a Gaussian expression of $\boldsymbol{x} \mapsto p(\boldsymbol{x}|\theta,\boldsymbol{y})$. For details, see Appendix A.

2.3.2. Correlated pseudo-marginal method

The efficiency of the PM method depends strongly on the number of latent variable samples N used to estimate the likelihood function. If this number is too low, the variability of the log-likelihood ratio estimator is likely to be high and the MH algorithm suffers from an impractically low acceptance rate (Beaumont, 2003). In the context of state-space models, Deligiannidis et al. (2018) show that N needs to increase linearly with the number of data T, thereby, often implying prohibitively high computational costs. For this reason, Deligiannidis et al. (2018) adapted the PM method by correlating the draws of latent variables used in the current and proposed step of the MH algorithm. The resulting correlated pseudo-marginal method (CPM method; illustrated in Fig. 1) leads to a better performance as the variance of a ratio of estimators is reduced when positively correlating the estimators of the denominator and numerator (Koop, 1972). For a standard normal distributed latent variable $\boldsymbol{Z}$, the CPM method draws a correlated realization of the nth latent variable in iteration j by,

$$\boldsymbol{Z}_n^{(j)} = \rho\boldsymbol{Z}_n^{(j-1)} + \sqrt{1-\rho^2}\epsilon, \text{ with } \rho \in (0,1) \text{ and } \epsilon = (\epsilon_1,\epsilon_2,\dots,\epsilon_L), \epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1). \tag{11}$$

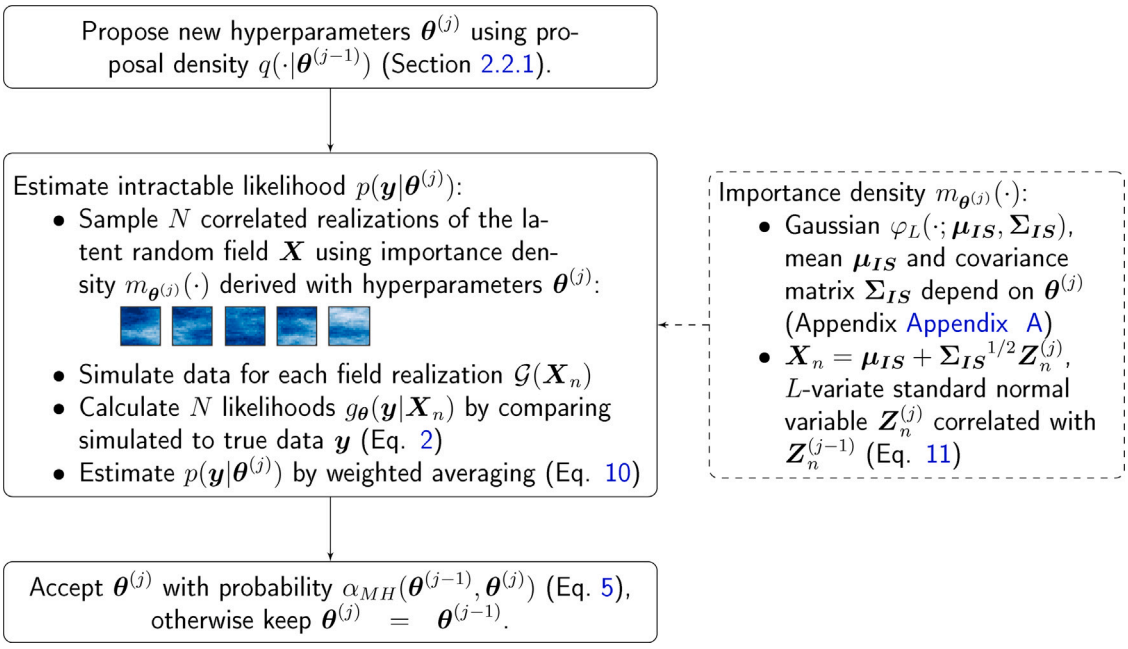


Fig. 1. Flow chart illustrating the CPM method with importance sampling at iteration j.

As numerous distributions can be obtained by transformations from standard normal variates, the general applicability of the CPM method is not limited by the uncorrelated Gaussian assumption (e.g. Chen et al., 2018). For example, in our two test cases that will follow, we generate correlated Gaussian latent variables $\boldsymbol{X}$ with mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ (or $\boldsymbol{\mu}_{\boldsymbol{IS}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{IS}}$) by transforming correlated standard-normally-distributed variables $\boldsymbol{Z}$ using Eq. (3). We stress that the proposed latent variables $\boldsymbol{Z}_n^{(j)}$ are only saved if $\boldsymbol{\theta}^{(j)}$ is accepted, otherwise we keep $\boldsymbol{Z}_n^{(j)} = \boldsymbol{Z}_n^{(j-1)}$ as for $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}$ in the MH algorithm.

The CPM method has two additional parameters compared to the standard MH algorithm: the latent variable sample size N and the correlation parameter ρ. Deligiannidis et al. (2018) propose to select N and ρ such that the variance of the log-likelihood ratio estimator,

$$W = \log\left(\hat{p}_N^{(j)}(\boldsymbol{y}|\boldsymbol{\theta})\right) - \log\left(\hat{p}_N^{(j-1)}(\boldsymbol{y}|\boldsymbol{\theta})\right), \tag{12}$$

takes values between 1.0 and 2.0 for $\boldsymbol{\theta}$ fixed in a region of high posterior probability mass. In practice, decreasing the variance of the estimator requires (1) more samples of the latent field or (2) a higher correlation of the samples making the exploration of the latent space slower. The range of 1.0 to 2.0 ensures a reasonable trade-off between the variance of the estimator, the exploration of the latent space (which would be slowed down by high ρ) and the computational cost (increases with increasing N). The region of $\boldsymbol{\theta}$ with high posterior mass can be chosen based on an initial MCMC run with N and ρ selected according to available prior knowledge. This choice can be inefficient, but will anyway give some first information. In practice, we first fix the number of samples N such that it is smaller than the number of available parallel processors. Then, we test a range of values for ρ and select one leading to $Var(W)$ being between 1.0 and 2.0.

2.4. Rejection sampling

Rejection sampling (RS; Ripley, 1987) is a basic Monte Carlo technique to generate independent samples from the posterior PDF. While it often suffers from an unfeasibly low acceptance rate, it is an exact sampling method (e.g., Robert and Casella, 2013) proceeding as follows:

1. Sample $\boldsymbol{\theta}$ from its prior distribution $p(\boldsymbol{\theta})$.
2. Sample u from a uniform distribution over $[0, 1]$.
3. Accept $\boldsymbol{\theta}$ if $u \leq \frac{p(\boldsymbol{y}|\boldsymbol{\theta})}{S_L}$, where S_L is the supremum of the likelihood function.

For our random effects model (Section 2.1), we estimate the intractable likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$ by sampling one brute-force realization $\boldsymbol{x}$ of the latent variable field $\boldsymbol{X} \sim f_{\boldsymbol{\theta}}(\cdot)$ with hyperparameters $\boldsymbol{\theta}$,

$$\begin{aligned}\hat{p}(\boldsymbol{y}|\boldsymbol{\theta}) &= g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}) = \varphi_T(\boldsymbol{y};\mathcal{G}(\boldsymbol{x}),\boldsymbol{\Sigma}_{\boldsymbol{Y}})\\ &= \det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{Y}})^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{y}-\mathcal{G}(\boldsymbol{x}))^T\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\mathcal{G}(\boldsymbol{x}))\right).\end{aligned} \tag{13}$$

In practice, an important challenge of RS methods is the need to estimate a tight bound S_L for the likelihood function. The most conservative choice is to assume a perfect data fit such that for our Gaussian likelihood function above we get $S_L = \det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{Y}})^{-1/2}$, but this will typically lead to an acceptance rate being close to zero. If we assume the errors to be equal to the standard deviation of the observational noise, we get $S_L = \det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{Y}})^{-1/2}\exp\left(-\frac{1}{2}T\right)$, which might lead to some bias as some realizations are likely to have higher likelihoods. One further possibility is to use the maximum likelihood value of the prior samples. To achieve this, RS is run by first saving all sampled prior realizations and their corresponding likelihood values. From this database, the maximum likelihood value is determined and all samples are assessed using this value. To obtain some accepted prior samples of $\boldsymbol{\theta}$ while ensuring an accurate estimate, we combine the second and the third approach and use the maximum of those two values as the supremum S_L.

2.5. Performance assessment

To assess if the CPM algorithm has converged, we use the $\hat{R}$-statistic of Gelman and Rubin (1992) comparing the within-chain variance with the between-chain variance of the second half of the MCMC chains. We follow the convention that the $\hat{R}$-statistic has to be smaller or equal to 1.2 for all model parameters. We also consider the acceptance rates (AR), which are aimed to be between 15% and 30% as proposed by Vrugt (2016).

We evaluate the amount of information gained by the inversion by comparing the marginal prior and posterior PDFs of the hyperparameters. This is achieved using the Kullback–Leibler divergence

(KL divergence; Kullback and Leibler, 1951) expressing the distance between two PDFs $z \mapsto p_1(z)$ and $z \mapsto p_2(z)$ (assumed positive) by,

$$KL(p_1 \parallel p_2) = \int p_1(z) \log\left(\frac{p_1(z)}{p_2(z)}\right) dz. \tag{14}$$

If $KL(p_1 \parallel p_2) = 0$, this means that the two PDFs are equal (almost everywhere), while an increasing value indicates diverging distributions. For example, for a standard normal PDF $p_2(\cdot)$, a KL divergence $KL(p_1 \parallel p_2) = 0.1$ is obtained by reducing the standard deviation within a centered standard normal $p_1(\cdot)$ to 0.7 and a KL divergence $KL(p_1 \parallel p_2) = 1$ is obtained by reducing the standard deviation to 0.23. To approximate the posterior PDFs, we apply kernel density estimation to the posterior samples (with manually adapted bandwidth).

To assess the quality of the posterior estimates, we use histograms to visually compare the marginal distributions with the true underlying values. Additionally, we evaluate the accuracy of the obtained posterior samples for each hyperparameter θ_i ($i \in \{1, 2, \dots, P\}$) numerically by applying a so-called scoring rule (Krüger et al., 2021). A scoring rule assesses the accuracy of a predictive PDF $z \mapsto p(z)$ with respect to a true value θ by accounting for both the statistical consistency between predictions and observations (calibration) and the sharpness of the prediction (Gneiting and Raftery, 2007). For our test cases, we employ the logarithmic score (logS; Good, 1952) defined by,

$$\text{logS}(p, \theta) = -\log p(\theta), \tag{15}$$

that is related to the Kullback–Leibler divergence (Gneiting and Raftery, 2007). If we compare two posterior estimates, the one with the lower score is favored. In practice (as for the KL divergence), we use a kernel density estimate of the posterior samples, which depends on the choice of the kernel and the bandwidth of the kernel smoothing window. We use a Gaussian kernel with manually adapted bandwidth. Our testings show that the choice influences the specific score values, but that the main results in terms of comparisons and conclusions are robust. If the posterior samples do not include the true value of θ, the density estimate of $p(\theta)$ can be numerically zero resulting in a logarithmic score of infinity. The logarithmic score is also available for multivariate densities, thereby, allowing evaluation of the estimated joint posterior PDFs of the hyperparameters.

3. Test case 1: Hydraulic conductivity field

Hydraulic conductivity is a key hydrogeological property. Particularly in contamination studies, the spatial variation of hydraulic conductivity plays an important role as it has a major influence on solute movement Butler (2005). Visentini et al. (2020) rely on time-lapse electrical resistance data during a tracer test to demonstrate that measurements of equivalent electrical properties, at a given scale, can be used to infer hyperparameters of the hydraulic conductivity field below this scale. Here, we seek to infer the hyperparameters of a log-hydraulic conductivity field in a data-poor setting involving only horizontally- and vertically- averaged equivalent hydraulic conductivity data. With such limited data, it is tempting to ignore heterogeneity or rely on upscaling relations valid for ergodic fields, as there is little hope that the data can constrain the field or its hyperparameters well. This example is used to demonstrate that ignoring heterogeneity or assuming ergodicity leads to significant errors when estimating the geostatistical mean. Furthermore, this data-poor setting allows for comparisons with rejection sampling (Section 2.4), thereby, demonstrating that the CPM method targets the right posterior of the hyperparameters.

3.1. Data and inversion setting

We target a 1 m × 1 m log-hydraulic conductivity field distributed according to a Gaussian random field $GRF(\mu_\theta(\cdot), C_\theta(\cdot,\cdot))$ with constant mean $\mu_\theta(\cdot)$ and exponential covariance function $C_\theta(\cdot,\cdot)$ (Eq. (1) with $H = 0.5$). We allow geometric anisotropy (Section 2.1) and denote the integral scale in the vertical direction (depth) as I^y and the anisotropy factor as λ. Together with the mean and standard deviation of the log-field, this forms the hyperparameters $\boldsymbol{\theta} = (\mu, \sigma, I^y, \lambda)$. Although we are mainly interested in the mean, we infer the other hyperparameters along with it, thereby, accounting for the possible non-ergodicity of the field. The log-hydraulic conductivity field (natural logarithm) is generated on a 100 × 100 grid (cell size is 1 cm) using a pixel-based approach (Section 2.1).

3.1.1. Synthetic data generation

We generate noise-contaminated synthetic data in both an ergodic and a non-ergodic setting. For the ergodic setting, we create one "true" field realization from which we obtain noise-contaminated data by assuming the field to be isotropic and use $\boldsymbol{\theta} = (ln(10^{-4}), 0.5, 0.03\ \text{m}, 1)$ and for the non-ergodic, anisotropic case we choose $\boldsymbol{\theta} = (ln(10^{-4}), 1.5, 0.1\ \text{m}, 3)$. The true log-hydraulic conductivity fields are shown in Fig. 2. Due to the discretization of the field chosen to limit the number of grid cells, the ergodic field is only nearly ergodic, implying that the generated data will vary somewhat when considering different field realizations with the true hyperparameters. In what follows, we will refer to it as ergodic except when a more specific designation is needed.

For the simulated measurements, we impose a hydraulic pressure gradient along either the horizontal or the vertical direction of the target field and observe a flux across one boundary. This information can then be used to calculate the equivalent horizontal and vertical hydraulic conductivities, given by Visentini et al. (2020),

$$K^H = \frac{1}{\triangle \Phi^H} \int_{\Gamma^H} -K(\mathbf{x}) \nabla_x \Phi^H(\mathbf{x}) d\mathbf{x}, \tag{16}$$

$$K^V = \frac{1}{\triangle \Phi_V} \int_{\Gamma^V} -K(\mathbf{x}) \nabla_y \Phi^V(\mathbf{x}) d\mathbf{x}, \tag{17}$$

where $K(\mathbf{x})$ denotes the hydraulic conductivity at position $\mathbf{x}$ with $\mathbf{x} = (x, y)^T$ referring to the 2-D position vector. Furthermore, $\triangle \Phi^H = \triangle \Phi^V = 1$ kPa denotes the constant hydraulic pressure difference imposed along the horizontal and vertical direction, respectively and $\Phi^H(\mathbf{x})$ and $\Phi^V(\mathbf{x})$ the resulting hydraulic head. Finally, Γ^H and Γ^V refer to integration paths separating the left and right and the top and bottom boundaries, respectively.

For the ergodic and isotropic field (Fig. 2(a)), we obtain equivalent hydraulic conductivities of $K^H = K^V = 9.2 \times 10^{-5}$ m/s and for the non-ergodic anisotropic field (Fig. 2(b)) we get an equivalent horizontal hydraulic conductivity of $K^H = 6.6 \times 10^{-5}$ m/s and an equivalent vertical hydraulic conductivity of $K^V = 4.8 \times 10^{-5}$ m/s. Finally, we add *i.i.d.* relative errors $\varepsilon_{\mathcal{O}}$ to the data pairs using a centered Gaussian distribution with a standard deviation given by 3% of the corresponding values.

3.1.2. Inversion settings and prior assumptions

The CPM method is implemented running three chains in parallel with adaptive proposals (Section 2.2.1) using an initialization period of $j_0 = 100$ where $\mathbf{C}^{(\mathbf{0})}$ is a diagonal matrix with $(0.008, 0.008, 0.002, 0.2)$ along its diagonal. For the prior PDFs of the first three hyperparameters, we use Uniform distributions: for the mean μ, we use the interval $[\log(10^{-5}), \log(10^{-3})]$, a range of standard deviation σ in-between $[0, 2]$ and for the integral scale I^y we assume [0 m, 0.5 m]. To account for the anisotropy factor λ being asymmetric around one, we employ a log-Uniform distribution with boundaries $[0.1, 10]$.

To tune N and ρ in the CPM method (Section 2.3.2), we consider the variance of the log-likelihood ratio estimator W (Eq. (12)). Fig. 3 depicts the dependence of the variance of W on the correlation ρ for ten and fifty samples ($N = 10, 50$) of the latent variable $\boldsymbol{X}$ for both the ergodic and the non-ergodic data setting. To evaluate the variances, we fix $\boldsymbol{\theta}$ at values having high posterior probability and draw realizations of the field by both sampling from its prior PDF $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ (noIS) and using importance sampling (IS, Appendix A). In the ergodic setting (Fig. 3(a)), all considered cases lead to variances of W being close to the target

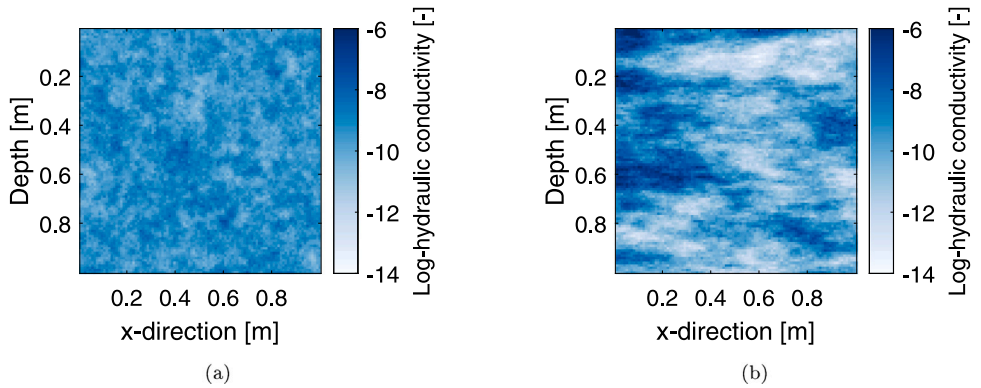


Fig. 2. "True" log-hydraulic conductivity fields of the first test case for (a) the ergodic setting and (b) the non-ergodic setting.

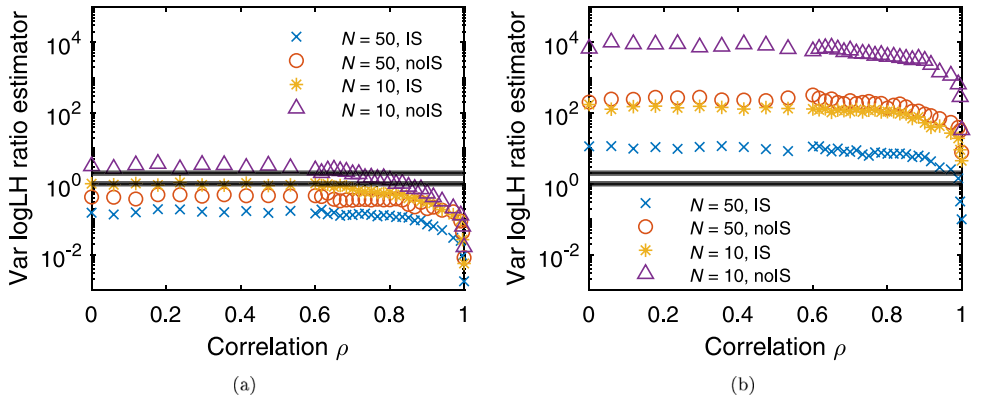


Fig. 3. Variance of the log-likelihood ratio estimator W for the first test case (hydraulic conductivity) as a function of the correlation ρ with θ being fixed at values with high posterior probability: (a) data generated with the ergodic setting and (b) data generated with the non-ergodic setting. The different markers refer to the number of latent variable samples and if sampling is made with or without importance sampling (IS). The black horizontal lines delimit the range between 1.0 and 2.0 recommended by Deligiannidis et al. (2018).

range between 1.0 and 2.0 recommended by Deligiannidis et al. (2018) even for $\rho = 0$. This is not surprising as in a purely ergodic setting, the realization of the random field does not influence the data. For the non-ergodic setting (Fig. 3(b)), the variance of W is up to 10^3 times higher and it is necessary to employ importance sampling. Of course, sampling from the prior could lead to variances of W being within the desired range, but the required values of N and ρ would lead to either excessively high computational costs at each iteration or very slow mixing in the draws of the latent variables. In the limit of $\rho = 1$, the variance of W is trivially equal to zero for all settings as we use the same latent variable samples in the first and second term of W, but this would lead to biased results. Initial MCMC runs showed that very diverse values of σ, I^y and λ have high posterior probabilities in both the ergodic and non-ergodic data settings as, in both cases, non-ergodic field realizations are sampled frequently by the CPM method. To ensure a controlled variance for all values θ with high posterior probability for both data settings, we perform importance sampling and choose $N = 50$ and $\rho = 0.975$ as it is appropriate for the more challenging non-ergodic settings.

For comparison purposes, we also run rejection sampling (RS; see Section 2.4) and a MH inversion assuming the parameter field to be ergodic (referred to as simplified MH; Fig. 4). For RS, we use the same number of field samples with corresponding forward simulations as needed by the CPM method for convergence. For the simplified MH, we rely on equations presented by Gelhar and Axness (1983) for the equivalent hydraulic conductivities in a two-dimensional anisotropic infinite

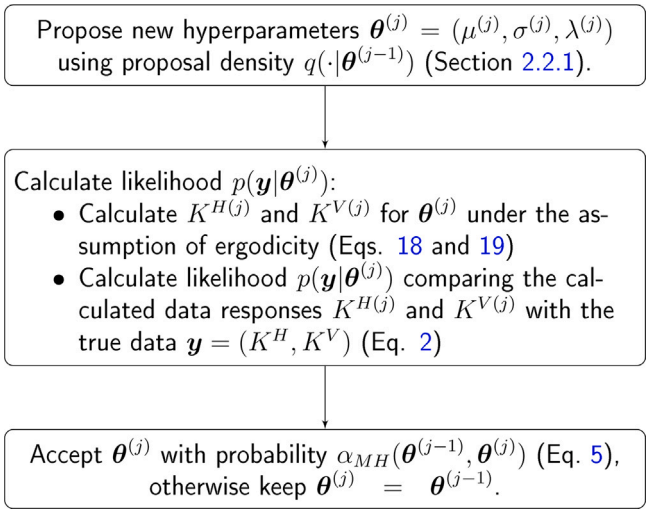


Fig. 4. Flow chart illustrating the simplified MH (assuming ergodicity) procedure for the first test case at iteration j.

domain (ensuring ergodicity) under mean uniform flow conditions:

$$K^H = K_G\Big(1 + \sigma^2\big(\frac{1}{2} - \frac{1}{1+\lambda}\big)\Big), \tag{18}$$

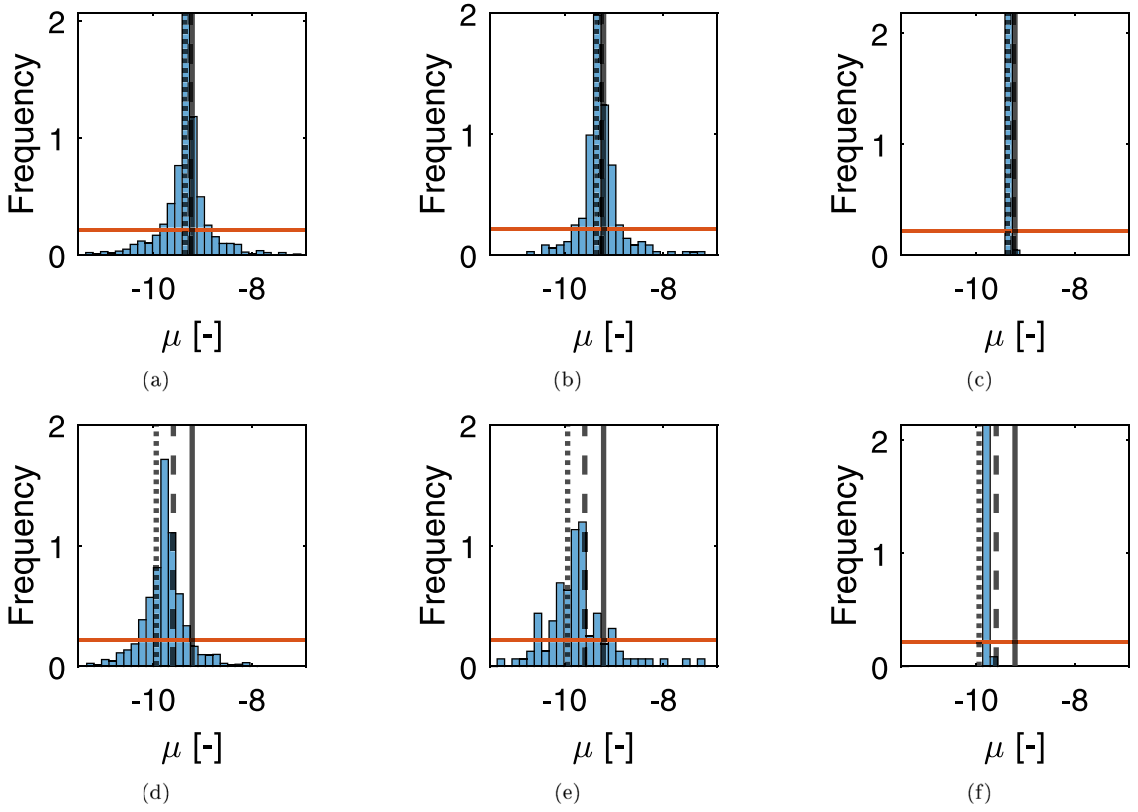


Fig. 5. Posterior samples obtained with the CPM method, rejection sampling (RS) and an inversion assuming ergodic conditions (simplified MH, Fig. 4) for the mean μ in the first test case (log-hydraulic conductivity): CPM results using the data generated with the (a) ergodic setting, (d) non-ergodic setting, RS using the data generated with the (b) ergodic setting and (e) non-ergodic setting and simplified MH using the data generated with the (c) ergodic setting and (f) non-ergodic setting. The dashed line is denoting $\log(K^H)$, the dotted line $\log(K^V)$, the solid vertical line indicates the true mean value and the red horizontal line the prior PDF.

$$K^V = K_G\Big(1 + \sigma^2(\frac{1}{2} - \frac{\lambda}{1+\lambda})\Big), \tag{19}$$

with K_G denoting the geometric mean of the linear hydraulic conductivity field $\mathbf{K} = \exp(\boldsymbol{X})$ (entry-wise exponential), which is the only parameter influencing the response for isotropic fields ($\lambda = 1$). It holds that $K_G = \exp(\overline{X})$ with $\overline{X}$ being the arithmetic mean of $\boldsymbol{X}$.

3.2. Results

We consider first the posterior PDF of the geostatistical mean μ obtained with the CPM method for the data generated with the ergodic setting (Fig. 2(a)). Only the samples obtained for the second half of the chains, after convergence has been declared (with respect to the $\hat{R}$-statistics, Table 1), are shown. The posterior PDF of the mean value μ in the ergodic data setting is centered around the true geostatistical mean and is clearly distinguishable from the Uniform prior PDF (Fig. 5(a)). This is confirmed by the correspondingly low logarithmic score emphasizing the accuracy of the posterior samples and the high KL divergence with respect to the prior PDF (Table 1). Comparison with the posterior samples obtained with RS (Fig. 5(b)) shows that both methods generate similar results with comparable KL divergences to the prior and almost equal logarithmic scores (Table 1). For the posterior samples obtained by assuming ergodicity (simplified MH, Fig. 4), we note a more compactly defined posterior than with CPM and RS with values of the mean being close to the true value (c.f., Figs. 5(a)–5(c)). Still, the logarithmic score of the mean is much higher than the one obtained with CPM and RS (Table 1), indicating that the samples generated under ergodic assumptions are not centered around the true geostatistical mean and are overconfident. This somewhat paradoxical result is a consequence of the data setting only being nearly ergodic, demonstrating the risk of getting biased and overconfident results even when the assumption of ergodicity is nearly fulfilled. The considered measurement scale is indeed 33 times larger than the integral scale.

For the non-ergodic data setting (Fig. 2(b)), the CPM-derived posterior distribution of the mean value μ contains the true value while being shifted towards the observed equivalent properties $\log(K^H)$ and $\log(K^V)$, leading to a higher logarithmic score than in the ergodic setting (Fig. 5(d) and Table 1). The posterior samples obtained with RS (Fig. 5(e)) are spread slightly wider than the ones of CPM, thereby, capturing more frequently the true value and leading to a lower KL divergence and a lower logarithmic score. We note that RS has an acceptance rate of 0.04% in this very data-poor and non-ergodic setting. In the non-ergodic setting, the simplified MH method leads to important errors in the estimated mean value (Fig. 5(f)). Indeed, the posterior samples are located around the (log-transformed) observed equivalent properties K^H and K^V and are removed from the true value of the geostatistical mean. This is reflected in a logarithmic score of infinity (see Table 1). Importantly, while the inversion assuming ergodicity solely samples mean values outside of the true value and has a very small posterior width, the CPM method includes the true value in the posterior samples (Figs. 5(d) and 5(f)).

For the other hyperparameters σ, I^y and λ inferred along with the mean μ, we get less well-resolved posterior estimates with both the CPM method and RS indicating that they are only weakly resolved by the available data. The corresponding plots are depicted in Appendix B.

4. Test case 2: Fracture aperture fields

Rock fractures play an important role as conduits (or barriers) for flow and solute transport. Their properties have often a major influence on hydrogeologic and geotechnical processes (National Research Council (NRC), 1996), but field characterization is inherently difficult.

Table 1
Table summarizing the results for the first test case (log-hydraulic conductivity) obtained with the CPM method, rejection sampling (RS) and the inversion assuming ergodicity (simplified MH; Fig. 4) for the ergodic (Fig. 2(a)) and the non-ergodic (Fig. 2(b)) data settings: convergence refers to the iteration in the MH methods with the $\hat{R}$-statistics being smaller than 1.2 for all parameters, the logarithmic score (LogS; Eq. (15)) assesses the accuracy of the posterior samples with respect to the true value and the KL divergence (Eq. (14)) is calculated for p_1 being the kernel density estimate gained with the (marginal) posterior samples and p_2 being the prior PDF. The bandwidth used for the marginal kernel density estimates of μ is 0.03 for CPM and RS and 0.005 for simplified MH.

Method Dataset	CPM Ergodic	RS Ergodic	Simp. MH Ergodic	CPM Non-ergodic	RS Non-ergodic	Simp. MH Non-ergodic
Convergence	1'300	–	4'000	2'200	–	10'000
AR	15%	0.11%	20%	15%	0.04%	15%
LogS μ	−0.33	−0.32	17.23	1.57	1.19	Inf
KL div. μ	1.07	1.16	3.92	1.05	0.66	4.01

The high contrast between the electrical properties of the filling of the fractures and the host rock (e.g., a water-filled fracture in granite host rock) leads to a very strong thin-bed response in ground penetrating radar (GPR) data. Quite remarkably, even sub-mm apertures yield measurable GPR responses even when the wavelength in the host rock may be on a metric scale. Imaging and characterization of fractures with GPR data has been studied extensively both from a theoretical perspective (e.g., Bradford and Deeds, 2006; Deparis and Garambois, 2008) and in controlled experiments (e.g., Grégoire and Hollender, 2004; Tsoflias et al., 2015). In these studies, it is typically either assumed that the aperture and material properties do not vary over the first Fresnel zone or that the influence of heterogeneous aperture fields average arithmetically in the acquired data. In a modeling study, Shakas and Linde (2017) assess this latter simplification by exploring a deterministic inversion in which the actual aperture field is heterogeneous at small scales, while it is assumed to be homogeneous when inferred for. Despite that the data can be fitted to the noise level, they find that the estimated apertures offer only reliable approximations of the arithmetic mean of the aperture field when the correlation length of the aperture heterogeneity is larger than the first Fresnel zone. Since fractures are known to be highly heterogeneous with self-affine properties, the study by Shakas and Linde (2017) suggest that many GPR-based estimations of mean apertures are biased and unreliable. They suggest that such heterogeneity needs to be explicitly accounted for, but they do not propose a solution. In this second data-rich test case, we will demonstrate how the CPM method can be used to obtain unbiased estimates of the mean aperture and statistics pertaining to the aperture field. We will then show how this information can be used to predict the equivalent hydraulic transmissivity of the fractures.

4.1. Data and inversion setting

We consider a 5 m × 5 m fracture aperture field $\boldsymbol{X}(\cdot)$ described as an isotropic Gaussian random field $GRF(\mu_{\boldsymbol{\theta}}(\cdot), C_{\boldsymbol{\theta}}(\cdot,\cdot))$ with constant mean $\mu_{\boldsymbol{\theta}}(\cdot)$ and powered exponential covariance function $C_{\boldsymbol{\theta}}(\cdot,\cdot)$ as specified in Eq. (1). With the CPM method, we target the mean μ, the standard deviation σ, the integral scale $I = I^x = I^y$ and the Hurst exponent H. The heterogeneous aperture field $\boldsymbol{X}(\cdot)$ is simulated using a pixel-based approach (Section 2.1) on a 50 × 50-dimensional grid ($D = 50$, cells of 10 cm side-lengths).

4.1.1. Synthetic data generation

The fracture aperture field from which data are generated is depicted in Fig. 6(a); the true hyperparameters are $\boldsymbol{\theta} = (\mu, \sigma, I, H) = (0.5\ \text{cm}, 0.1, 0.2\ \text{m}, 0.8)$. We only consider a single fracture in a model domain of 10 m × 10 m × 10 m (Fig. 6(b)). The background rock matrix is assumed to be homogeneous with a relative electrical permittivity of 9 and an electrical conductivity of 0.001 S/m. For the fracture, we assume a constant relative electrical permittivity of 81 and electrical conductivity of 0.1 S/m.

To generate the synthetic GPR reflection data, we rely on the effective-dipole method of Shakas and Linde (2015). This modeling framework combines analytical solutions for radiation in the matrix domain and dipole elements, corresponding to discretized sections of the fracture, radiating as electric dipoles modulated by the thin-bed reflection coefficients. A simple schematic of the method is represented in Fig. 6(c) (adapted from Fig. 3 of Shakas and Linde (2017)). We use two GPR reflection traces generated with sources and receivers located 5 m away from the fracture and with offsets of 0 m and 2 m (Fig. 6(b)). The source signal is assumed to be vertically-oriented with a source spectrum consisting of a Ricker wavelet with dominant wavelength of 100 cm. With a discretization of 10 cm of the aperture field, this results in 10 elements per dominant wavelength for which highly accurate simulations are expected (Shakas and Linde, 2017). The responses are generated in the frequency-domain using a frequency range from zero to 300 MHz with a sampling step size of 1 MHz. As in practice, the amplitude of the source wavelet is unknown, Shakas and Linde (2017) normalize the response values in the data generation and inversion. Here, we instead introduce an unknown factor c_A by which the responses are multiplied. This factor is equal to one for the true data and it is inferred within the inversion. This extends the target variables to $\boldsymbol{\theta} = (\mu, \sigma^2, l_c, H, c_A)$. Finally, for each of the 300 complex-valued numbers representing the electric field, we add independent realizations of Gaussian measurement noise $\varepsilon_{\mathcal{O}}$ with a standard deviation of 3% of the maximal value. The inversions are performed in the frequency-domain, but we present for visual purposes the two corresponding traces in the time-domain (Fig. 6(d)). For completeness, we also show the smoother traces (Fig. 6(e)) obtained by sampling over the same frequency range with a sampling step size of 0.1 MHz.

4.1.2. Inversion settings and prior assumptions

As in the first test case, we run an adaptive Metropolis–Hastings version of the CPM method with three chains in parallel. We specify $j_0 = 500$ and $\mathbf{C}^{(\mathbf{0})}$ as a diagonal matrix with 0.001 on its diagonal. Furthermore, to ensure a suitable acceptance rate, we decrease the step size by 50%. For the prior PDFs of the hyperparameters, we use Uniform distributions: for the mean μ we use $\text{Unif}[0\ \text{cm}, 1\ \text{cm}]$, for the standard deviation σ we use $\text{Unif}[0\ \text{cm}, 0.5\ \text{cm}]$, for the integral scale l_c we use $\text{Unif}[0\ \text{m}, 1\ \text{m}]$, for the Hurst exponent H we use $\text{Unif}[0.1, 1]$ and for the amplitude factor c_A we use $\text{Unif}[0.5, 2]$.

The importance sampling mean $\boldsymbol{\mu}_{\boldsymbol{IS}}$ for the latent aperture field $\boldsymbol{X}$ when the proposed $\boldsymbol{\theta}^{(j)}$ is the true hyperparameters is depicted in Fig. 7(a) (see formulas in Appendix A). Fig. 7(b) depicts the dependence of the variance of the log-likelihood ratio estimator W (Eq. (12)) on N and ρ. The importance sampling leads to a tremendous decrease of the variance of W (e.g., for $N = 1$ and $\rho = 0$, the variance of W is reduced from 10^6 to 10^2). Furthermore, increasing the number of latent variable samples N and the correlation parameter ρ also reduces the variance of W strongly. Following Fig. 7(b), we run the CPM algorithm with $N = 5$ and $\rho = 0.975$ in combination with importance sampling.

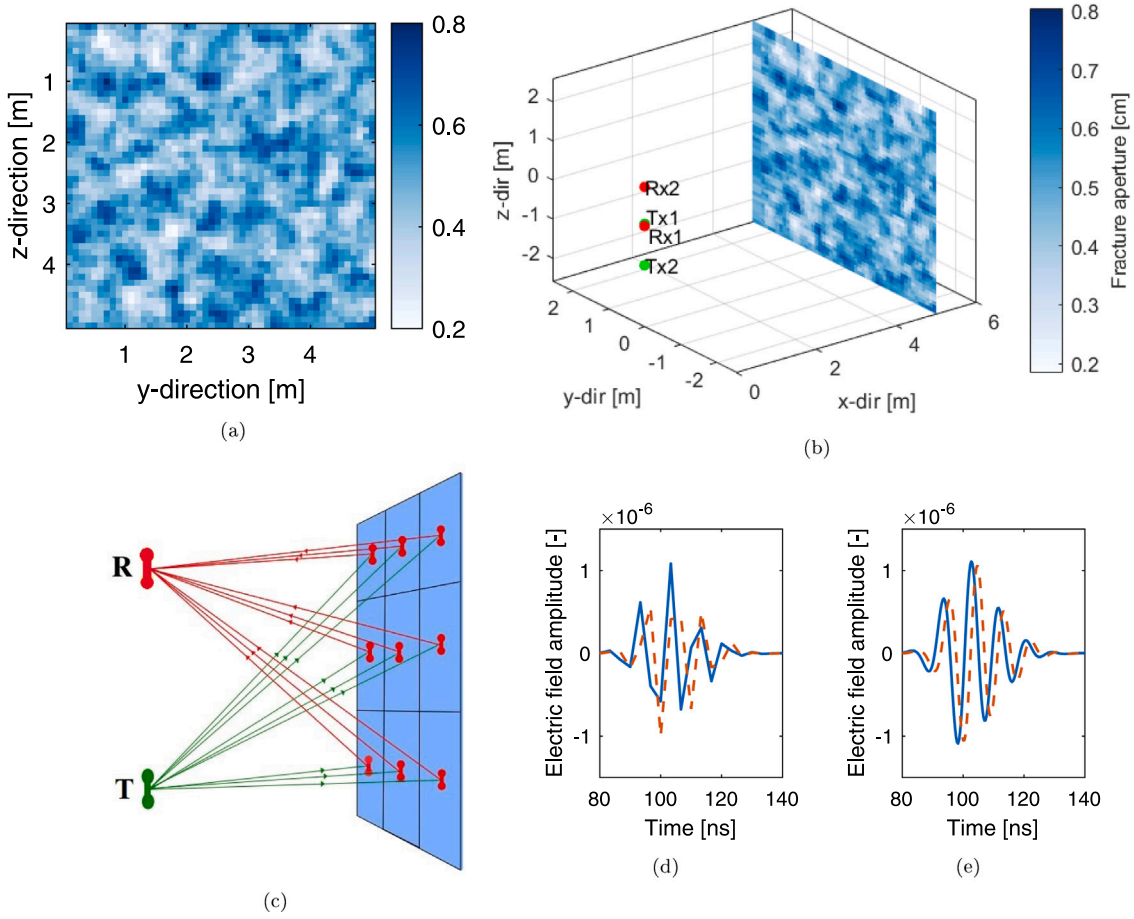


Fig. 6. (a) "True" fracture aperture field of the second test case with $\theta = (0.5\ \text{cm}, 0.1, 0.2\ \text{m}, 0.8, 1)$. (b) Model domain with aperture field and transmitter–receiver layout. (c) Schematic of the effective-dipole forward modeling framework (adapted from Fig. 3 of Shakas and Linde, 2017). GPR reflection traces (time-domain) with a sampling step size of (d) 1 MHz and (e) 0.1 MHz.

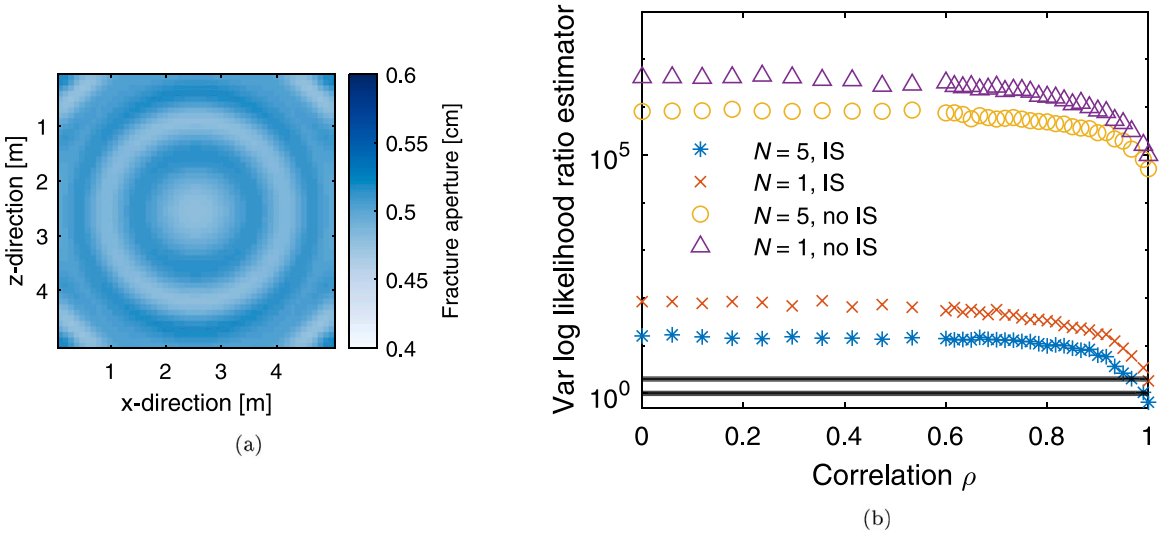


Fig. 7. (a) Importance sampling mean μ_{IS} of the aperture field for the true values of θ in the second test case (Appendix A) and (b) variance of the log-likelihood ratio estimator W as a function of the correlation ρ with θ being fixed at values with high posterior probability. The different markers refer to different numbers N of latent variable samples drawn with or without importance sampling (IS). The black horizontal lines delimit the range between 1.0 and 2.0 recommended by Deligiannidis et al. (2018).

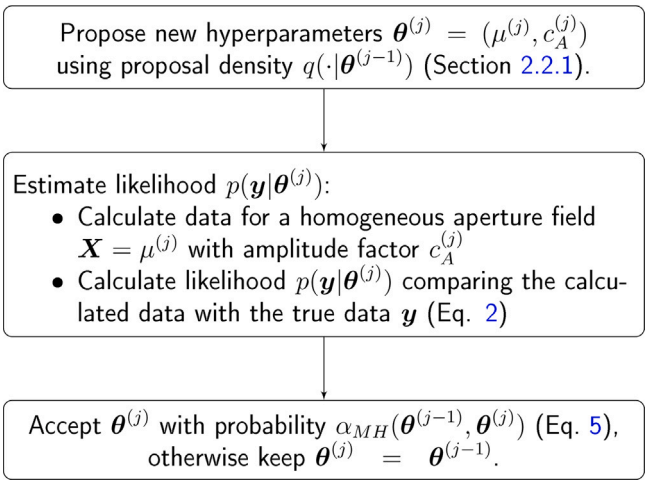


Fig. 8. Flow chart illustrating the homogeneous inversion procedure for the second test case at iteration j.

To place the results obtained with CPM into context, we compare them to those obtained with an inversion assuming the aperture field to be homogeneous (illustrated with a flow chart in Fig. 8). To achieve this, we only infer the mean aperture μ and the amplitude factor c_A, which is broadly similar to the inversion setting considered by Shakas and Linde (2017).

4.2. Results

The estimated marginal posterior PDFs of $\boldsymbol{\theta} = (\mu, \sigma^2, l_c, H, c_A)$ obtained with the CPM method are depicted in Figs. 9(a)–9(e). Convergence is reached within 10,000 iterations with respect to the $\hat{R}$-statistic and we display the results for the second half of the chains. The histograms depicting the posterior samples of the mean μ (Fig. 9(a)), standard deviation σ (Fig. 9(b)) and amplitude factor c_A (Fig. 9(e)) show the strongest concentration with respect to the prior and correspondingly high KL divergences (Table 2). The sample range for the integral scale I (Fig. 9(c)) and the Hurst exponent H (Fig. 9(d)) is equally wide as the respective prior PDFs and the corresponding KL divergences are rather small. Nonetheless, the integral scale is preferentially sampled in the region of the true value. As the values of the logarithmic score (Table 2) can generally not be compared between hyperparameters (different width of support), they will become of interest only in the comparison with a competing method.

Figs. 9(f) and 9(g) show the histograms of the posterior samples for the mean μ and the amplitude factor c_A obtained for the inversion assuming a homogeneous aperture (Fig. 8). The range of the samples is very narrow with high KL divergences with respect to the prior PDF (Table 2) but located far from the true parameter values. This results in infinite logarithmic scores (Table 2). As we have seen, the CPM method accounting for heterogeneity leads to posterior samples of the mean and amplitude factor that cover a wider range including the true values used to generate the data as reflected in lower logarithmic scores (Table 2). The estimates of the mean aperture μ and the amplitude factor c_A are highly correlated. Fig. 9(h) shows that under the assumption of knowing $c_A = 1$, the range of the samples obtained with CPM for the mean aperture would be more narrow and shifted towards the true value of 0.5 cm. We further see that the homogeneous inversion only explores a small part (and the wrong part) of the joint posterior model space, leading to a logarithmic score of infinity for the estimated joint posterior PDF (Table 2). For this second data-rich test case, rejection sampling is unfeasible as the acceptance rate is below 0.001%.

Table 2
Table summarizing the results for the second test case (aperture field, Fig. 6) obtained with the CPM method and the inversion assuming homogeneity (Fig. 8): convergence refers to the iteration with a $\hat{R}$-statistics being smaller than 1.2 for all parameters, the logarithmic score (logS; Eq. (15)) evaluates the accuracy of the marginal and joint posterior samples with respect to the true value and the KL divergence (Eq. (14)) is calculated for p_1 being the marginal kernel density estimate gained with the posterior samples and p_2 being the prior PDF. For the kernel density estimates of CPM, we use the following bandwidths: 0.01 (μ), 0.006 (σ), 0.02 (I), 0.04 (H), 0.01 (c_A) and 0.02 for both in the joint PDF of (μ, c_A). For the homogeneous inversion, we use 0.001 (μ), 0.001 (c_A) and 0.002 for both in the joint PDF of (μ, c_A).

Method	CPM	Homogeneous inv.
Convergence	10'000	1'000
Acceptance rate	15%	15%
LogS μ	0.60	Inf
LogS σ	−1.34	–
LogS I	−1.05	–
LogS H	−0.27	–
LogS c_A	0.51	Inf
LogS (μ, c_A)	−1.31	Inf
KL divergence μ	0.96	3.31
KL divergence σ	0.95	–
KL divergence I	0.35	–
KL divergence H	0.02	–
KL divergence c_A	1.29	3.90

4.2.1. Predictions of hydraulic transmissivity

To complement the results obtained for this GPR test case and to strengthen the link to hydrogeology, we use the aperture field estimates to derive equivalent hydraulic transmissivities. First, we use the inferred mean apertures obtained with the inversion assuming the field to be homogeneous (Fig. 9(f)). These aperture field realizations are used to derive hydraulic transmissivities at the fracture scale using the classical parallel plate model (Tsang, 1992),

$$T = (1/(12\eta))\mu^3, \tag{20}$$

with $\eta = 8.9 \times 10^{-4}$ Pa · s denoting the dynamic viscosity (25 degree C) and μ being the inferred mean aperture values (in meters). The resulting horizontal equivalent log-hydraulic transmissivities are shown in Fig. 10 (light gray). This result is now compared with the value obtained for the true aperture field under the assumption that the Reynolds equation is valid, implying that we can apply Eq. (20) locally to obtain a hydraulic transmissivity field and then solve numerically for the resulting effective transmissivity at the fracture scale. The results show that the true effective hydraulic transmissivity is roughly one order of magnitude smaller and that the posterior PDF of the homogeneous inversion (light gray) is nowhere close to include this value. This is reflected in a infinite logarithmic score. We then sample field realizations using the posterior PDFs of the hyperparameters inferred with the CPM method. The resulting equivalent log-hydraulic transmissivity values are shown in Fig. 10 (blue). This distribution is much wider, it includes the true value, and the mean is clearly shifted towards the true value. The corresponding logarithmic score is 0.23.

5. Discussion

Our two test cases presented in Sections 3 and 4 demonstrate the ability of the correlated pseudo-marginal method (CPM method; Fig. 1) to estimate the posterior PDF of the target field's hyperparameters (e.g., mean, standard deviation, integral scale, Hurst exponent and anisotropy factor) while accounting for the impact of small-scale heterogeneity within the estimate of the likelihood function. We further demonstrate that inversions invoking simplified assumption such as ergodicity or homogeneity lead to biased and overconfident results such that the inferred posteriors often do not include the true values. Compared to previous inversion approaches targeting hyperparameters (e.g., Laloy et al. (2015) and Xiao et al. (2021)), the CPM method

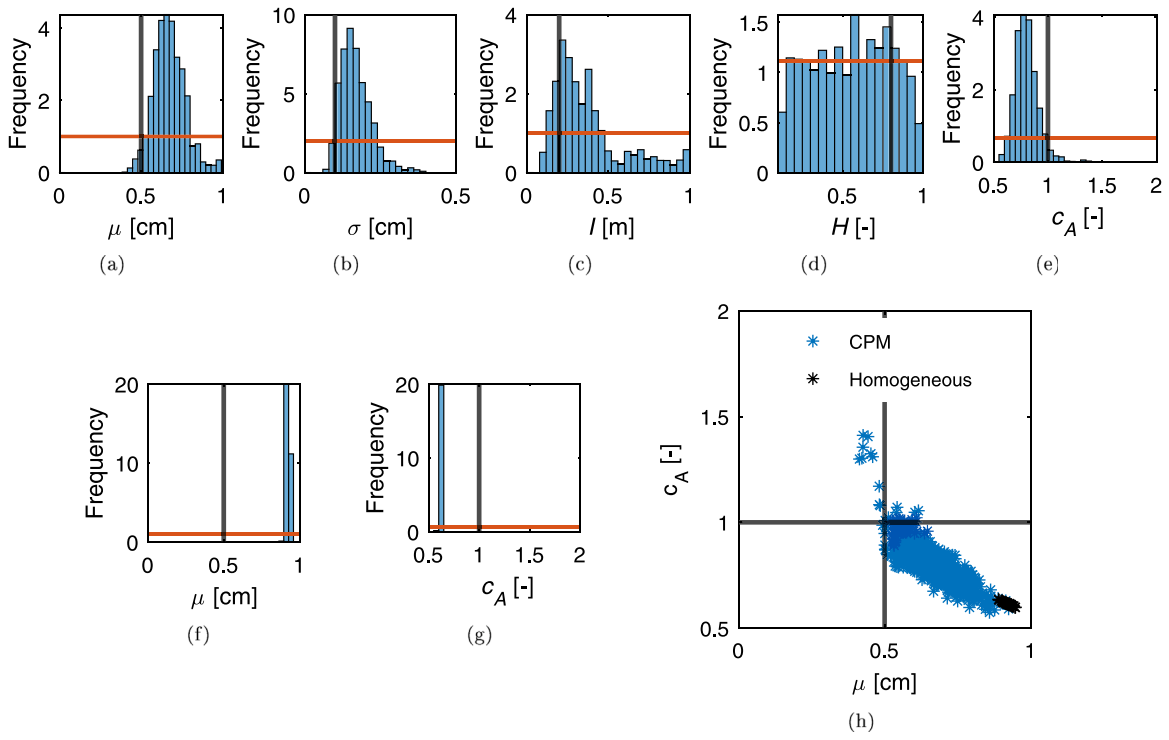


Fig. 9. Posterior samples obtained with CPM in the second test case for the (a) mean aperture μ, (b) standard deviation σ, (c) integral scale I, (d) Hurst exponent H and (e) amplitude factor c_A. Posterior samples obtained with an inversion assuming the aperture field to be homogeneous for the (f) mean aperture μ and (g) amplitude factor c_A. Scatter plot of the sampled pairs of mean aperture μ and amplitude factor c_A for the CPM method and the homogeneous inversion. The black lines indicate the true values and the red horizontal lines the prior PDFs.

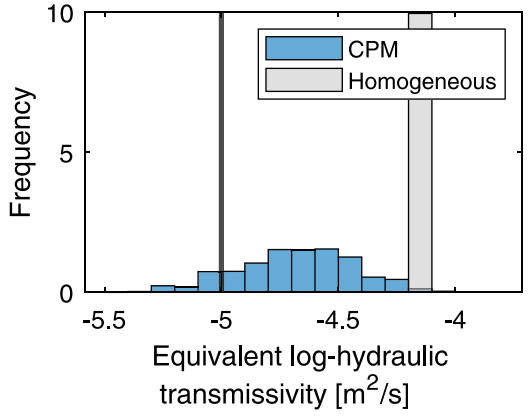


Fig. 10. Posterior PDF of the horizontal equivalent log-hydraulic transmissivity (log10-scale) for heterogeneous aperture fields (second test case) sampled with the inferred hyperparameters of the CPM method (blue) and homogeneous aperture fields generated with the inferred mean values of the inversion assuming the field to be homogeneous (light gray). The black vertical line indicates the value corresponding to the true aperture field.

infers the hyperparameters only, thereby, avoiding to infer the posterior PDF of the many thousands of latent variables. The two presented test cases cover one data-poor transmission problem governed by diffusion (e.g. electrical conduction, heat conduction, or groundwater flow) and one more data-rich reflection problem governed by wave-based physics (e.g., GPR, seismics, acoustics). The generality of these settings suggest that the CPM method has a wide applicability in hydrogeology and geophysics.

The first test case related to heterogeneous hydraulic conductivity fields concerns a very data-poor setting in which only the horizontal and vertical equivalent hydraulic conductivities are used as data points. To compare the performance of the CPM method with an inversion assuming ergodicity (referred to as simplified MH; Fig. 4), we consider a nearly ergodic and a non-ergodic data setting. In both settings, the geostatistical mean of the model domain can be inferred from the equivalent conductivities using the CPM method. In the ergodic setting, both the CPM method and the simplified MH lead to reasonable estimates of the geostatistical mean, with the posterior range of the CPM method being wider as its underlying assumptions are less restrictive (Figs. 5(a) and 5(c)). In the non-ergodic data setting, the simplified MH leads to important errors in the estimation of the geostatistical mean with a posterior range far from the true value (Fig. 5(f)). For the CPM method, the estimated posterior uncertainty is wider and the true value of the mean is included (Fig. 5(d)). Thereby, the logarithmic score is reduced from infinity to 1.57 when applying CPM compared with the simplified MH (Table 1). We conclude that even in this extremely data-poor setting, the use of simplified model assumptions leads to a substantial bias in the mean estimate and an overconfident posterior bound. For the other hyperparameters (standard deviation, integral scale and anisotropy factor), we conclude that only little information can be gained in this data-poor setting. Furthermore, we demonstrate that the CPM results are in agreement with those obtained by rejection sampling.

In the second test case concerning fracture aperture fields, we limit ourselves to a non-ergodic data setting and compare the results obtained with CPM with those of an inversion assuming the aperture field to be homogeneous (Fig. 8). We can consider this homogeneous inversion as either (1) an inversion inferring the geostatistical mean under simplified model assumptions or (2) an inversion targeting the equivalent GPR aperture. We show that the homogeneous assumption leads to posterior samples being located far from the true geostatistical mean value (Fig. 9(f)), demonstrating in accordance with Shakas and Linde (2017) (1) that the geostatistical mean of the aperture field can be very different than the equivalent GPR aperture and (2) that inferring the geostatistical mean based on a too simple model description leads

to biased estimates. Indeed, in such an inversion one appears to get increasingly certain about the wrong parameter values as more data are added or the data noise level is decreased (Brynjarsdóttir and O'Hagan, 2014). In contrast, the CPM method accounting for non-ergodicity and heterogeneity by inferring additionally for the standard deviation, integral scale and Hurst exponent leads to a wider posterior including the true value of the aperture mean (Fig. 9(a)). For this second example, employing the CPM method leads to a reduction of the logarithmic score from infinity to 0.60 for the posterior estimate of the mean in comparison with the homogeneous inversion (Table 2). Additionally, CPM enables to infer information about other hyperparameters (standard deviation and integral scale) of the field.

Probabilistic inference of hyperparameters offers the possibility to translate from one type of equivalent property to another. We demonstrate this by predicting the equivalent log-hydraulic transmissivity at the fracture scale using the fracture aperture fields obtained in the second test case (Fig. 10). The predicted values for the constant aperture field inversion are obtained by applying the equivalent GPR aperture in the cubic law. When deriving hydraulic properties from these constant fields, we assume that this equivalent GPR aperture is the same as the equivalent "cubic law aperture" (in the sense of Tsang, 1992), which is the equivalent parallel plate aperture with respect to hydraulic flow properties. These predictions are very different from those obtained from the true aperture field when applying the local cubic law (Fig. 10). This visualizes clearly that the equivalent aperture for one type of physics cannot be assumed to be the same when considering another type of physics. Actually, the equivalent aperture (in a cubic law sense) with respect to the hydraulic data of the true aperture field is 0.47 cm, a value considerably diverging from the one inferred from the GPR data when assuming homogeneity (about 0.9 cm). Using field realizations sampled with the posterior PDFs of the hyperparameters obtained by CPM lead to a wider and more accurate range of effective log-hydraulic transmissivity values (Fig. 10). While the logarithmic score for the transmissivity predictions obtained with the homogeneous inversion is infinity, the one obtained with CPM is 0.23. This suggests that while equivalent properties always refer to one specific kind of physics, the inference of hyperparameters enables a general description of the model domain. The CPM method is well suited to achieve this by targeting only the hyperparameters of interest, thereby, enabling probabilistic forecasts for different types of physics.

This study expands further the range of applications that the CPM method can address in geoscientific settings. While Friedli et al. (2022) used it to account for uncertainties in petrophysical relationships in the context of hydrogeophysics, we provide here a very different problem setting in which the CPM method is used to account for non-ergodicity and small-scale heterogeneities when inferring hyperparameters. In these examples, we only consider heterogeneities in two dimensions. In field applications, the data are of course affected by heterogeneities outside the 2-D plane of measurements (e.g., between boreholes) or by outer-space effects (Maurer and Friedel, 2006). To further improve the estimation and uncertainty quantification in such setups, the CPM method could be employed to integrate out heterogeneities in three dimensions (in the context of the present study), or in the third out-of-plane dimension in the setting considered by Friedli et al. (2022) or in general 2-D inversions to avoid over-confident (and possibly biased) estimates. For the presented test cases we used a pixel-based representation of the Gaussian latent random field. We stress that there exist many alternative ways to represent and generate a Gaussian random field as, for example, the fast circulant embedding technique using a spectral representation by Dietrich and Newsam (1997). While such an approach offers an increased efficiency in the generation of the random field realizations, careful consideration must be given on a case-by-case basis as to whether this could be integrated into a well-working importance sampling strategy. Moreover, in settings where the correlation length is of similar size as the model domain, the embedding has to be extended and the efficiency is reduced. We assume the latent random fields to be Gaussian, simplifying the derivation of the importance sampling density. An important topic for future research would be to develop and assess suitable importance distributions in non-Gaussian settings.

The efficiency of the CPM method depends strongly on the variance of the log-likelihood ratio estimator. Especially in settings with a high number of observations with a low signal-to-noise ratio, one needs a well-working importance density when sampling the latent variables. The relevance of a well-tuned importance sampling strategy becomes clear when comparing the number of samples needed to control the variance in the first and second test cases (Figs. 3 and 7(b)). For the first test case, the IS density is of only moderate quality and many samples ($N = 50$) are needed even for this data-poor setting. For the second more data-rich test case with a well-defined IS density, only a few samples ($N = 5$) are sufficient. If the determination of a well-working IS distribution is not feasible, this can be detrimental to the applicability of the CPM method. In such a scenario there is also the risk of poor exploration of the latent space, namely if the likelihood estimator depends mainly on one or two latent variable samples with a particularly beneficial small-scale structure. One solution in such a scenario is to infer some additional main features of the latent field together with the hyperparameters and then to apply the CPM method to sample out the remaining randomness of the field. This could be done using the main components of a dimensionality reduction approach and should reduce the importance of a well-tuned IS density. We leave this idea for future research. Recently, Wang et al. (2022) proposed an hierarchical Bayesian inversion approach targeting first so-called global variables (such as hyperparameters but also physical variables) and then estimating the posterior of the whole field. For the estimation of the global variable's posterior in a non-linear setting, Wang et al. (2022) apply a machine-learning based approach and train a neural network to output the global variables given a data realization followed by kernel density estimation of the results. Such a method relies on the ability to estimate the hyperparameters by brute-force prior sampling and subsequent comparison of the resulting data with the true measurements. In strongly non-ergodic settings, this can be computationally challenging as an unrealistically high number of prior samples would be needed to obtain reasonable estimates. To illustrate this, Fig. 11 shows the 100 highest log-likelihood values sampled from 5000 prior samples of the aperture field in the second test case (Section 4). We note that no sample was generated with a likelihood close to the true one (black horizontal line) implying that an unfeasible large amount of samples would be needed to guarantee accurate hyperparameter estimates. Indeed, even the highest sampled likelihood has a likelihood that is still 10^{44} times smaller than the true likelihood. In contrast, our CPM method using three chains need 10'000 iterations per chain for convergence.

6. Conclusions

We consider Bayesian MCMC inversions inferring hyperparameters (e.g., mean, standard deviation and integral scales) from hydrogeological or geophysical data. To achieve this is particularly challenging in the non-ergodic setting, in which the data depend on the actual geostatistical field realization under consideration and not only on the hyperparameters. To prevent errors arising when assuming homogeneity or ergodicity, we rely on the correlated pseudo-marginal method targeting the hyperparameters while integrating out the random effects of actual field realizations in the likelihood estimation. This approach has the advantage of ensuring accurate posterior estimates of hyperparameters without having to infer thousands or more parameters as needed if the whole random field would be inferred. To ensure efficiency, the correlated pseudo-marginal method employs importance sampling and correlation of the latent draws used in the proposed and current steps of the MCMC chain. We assess the performance of this method through two synthetic test cases involving

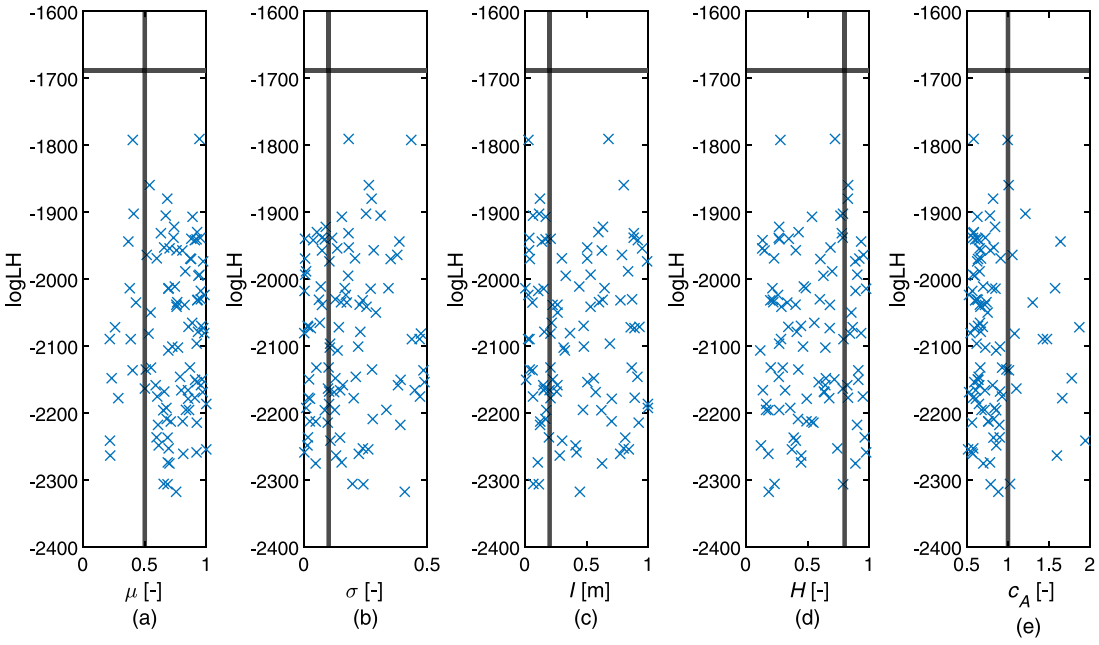


Fig. 11. Log-likelihood values obtained with brute-force sampling of the latent aperture field $\boldsymbol{X}$ in Test case 2 (Section 4) when drawing prior samples from the hyperparameters; for readability, we only show the samples with the 100 highest log-likelihood values (out of 5000 samples). The horizontal lines depict the true log-likelihood and the vertical lines the true values of the hyperparameters.

(1) diffusion-based physics in a data-poor setting targeting hydraulic properties using equivalent hydraulic conductivity data and (2) wave-based physics in a more data-rich example targeting a fracture aperture field using single-hole ground-penetrating radar (GPR) reflection data. By using these two examples that are representative of a broad range of geophysical and hydrogeological problems, we demonstrate that the correlated pseudo-marginal method provides accurate estimation of the geostatistical mean in both ergodic and non-ergodic settings. Furthermore, for all considered hyperparameters, we show that the correlated pseudo-marginal method avoids over-confident and biased posterior PDF estimates that plague inversion results obtained when assuming ergodicity or homogeneity. Estimating hyperparameters allows for a general description of property fields which is independent of the physics under consideration, thereby, allowing ultimately to use the estimated posterior PDFs to make predictions for other types of physics or experimental set-ups. This is demonstrated by transforming the fracture properties inferred by GPR data into predictions of equivalent hydraulic transmissivity at the fracture scale.

CRediT authorship contribution statement

Lea Friedli: Conceptualization, Methodology, Software, Visualization, Formal analysis, Validation, Investigation, Writing – original draft. **Niklas Linde:** Supervision, Conceptualization, Methodology, Formal analysis, Funding acquisition, Writing – review & editing. **David Ginsbourger:** Supervision, Methodology, Writing – review & editing. **Alejandro Fernandez Visentini:** Software, Writing – review & editing. **Arnaud Doucet:** Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Lea Friedli reports financial support was provided by Swiss National Science Foundation.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported by the Swiss National Science Foundation (project number: 184574). No new experimental data were generated or analyzed in support of this research.

Appendix A. Importance sampling for CPM

As emphasized in Section 2.3.2, it is essential that the variance of the log-likelihood ratio estimator is low to ensure efficient PM or CPM performance. Especially in settings with large data sets with high signal-to-noise ratios, the integrand $g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})$ in Eq. (7) is likely to have a peak in a region of $\boldsymbol{X}$ with very small probability under its prior PDF $\boldsymbol{x} \mapsto f_{\boldsymbol{\theta}}(\boldsymbol{x})$. Sampling the latent variables using the prior distribution then leads to an inefficient algorithm. One remedy to this is importance sampling, where instead of the prior distribution, a so-called importance distribution given by the PDF $\boldsymbol{x} \mapsto m_{\boldsymbol{\theta}}(\boldsymbol{x})$ is employed. The importance distribution is chosen such that it preferentially generates samples with high $g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})f_{\boldsymbol{\theta}}(\boldsymbol{x})$ while guaranteeing that all values $\boldsymbol{x}$, for which $g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})f_{\boldsymbol{\theta}}(\boldsymbol{x}) > 0$, are included in its support (Owen and Zhou, 2000). It holds,

$$\int g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})f_{\boldsymbol{\theta}}(\boldsymbol{x})d\boldsymbol{x} = \int \frac{g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})f_{\boldsymbol{\theta}}(\boldsymbol{x})}{m_{\boldsymbol{\theta}}(\boldsymbol{x})} m_{\boldsymbol{\theta}}(\boldsymbol{x})d\boldsymbol{x}, \tag{A.1}$$

from which the unbiased importance sampling estimate of the likelihood function in Eq. (10) is derived. To minimize the variance of the estimator, $\boldsymbol{x} \mapsto m_{\boldsymbol{\theta}}(\boldsymbol{x})$ should be nearly proportional to $\boldsymbol{x} \mapsto g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})f_{\boldsymbol{\theta}}(\boldsymbol{x})$ as presented in Owen and Zhou (2000) referring to the results of Kahn and Marshall (1953). Since it holds that $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \propto g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})f_{\boldsymbol{\theta}}(\boldsymbol{x})$, we use for the importance density an approximation of $\boldsymbol{x} \mapsto p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$.

In Section 2.1, we specify $\boldsymbol{Y} = \mathcal{G}(\boldsymbol{X}) + \varepsilon_{\boldsymbol{\Theta}}$ with $\mathcal{G} : \mathbb{R}^{D^2} \to \mathbb{R}^T$ being the (physical) forward solver and $\varepsilon_{\boldsymbol{\Theta}}$ the observational noise. If the forward solver is non-linear, there is no exact expression for $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$. For this reason, we approximate $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ by expressing the map $\boldsymbol{x} \mapsto \mathcal{G}(\boldsymbol{x})$ based on either an upscaling formula assuming an anisotropic ergodic setting (test case 1) or a linearization of the forward solver (test case 2). Following Friedli et al. (2022), we use Gaussian distributions and a lemma about marginal and conditional Gaussians; see for example Bishop (2006).

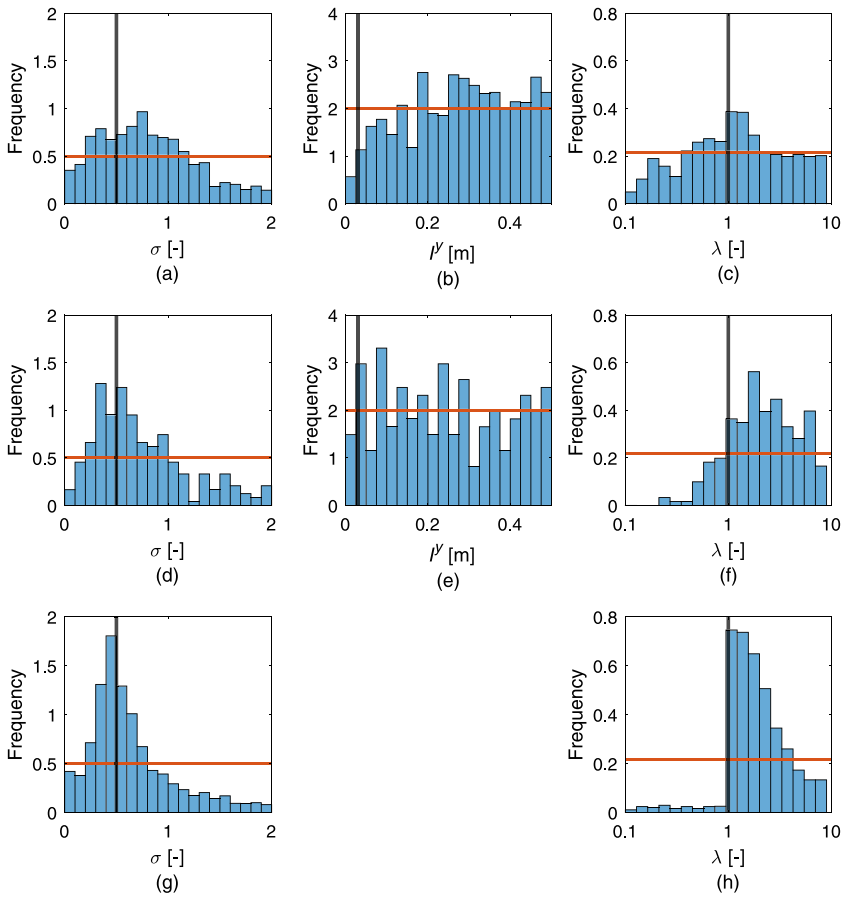


Fig. B.1. Posterior samples for the remaining hyperparameters for the first test case (log-hydraulic conductivity) using the data generated with the ergodic setting; (a) standard deviation σ, (b) integral scale I^y and (c) anisotropy factor λ with CPM. (d) Standard deviation σ, (e) integral scale I^y and (f) anisotropy factor λ with RS. (g) Standard deviation σ and (h) anisotropy factor λ with simplified MH. The solid vertical lines indicate the true hyperparameter values and the red horizontal lines the prior PDFs. Note that for the anisotropy factor λ we employ a logarithmic scale on the x-axis.

Lemma 1. *Marginal and Conditional Gaussians*

Assume a marginal Gaussian distribution for $\mathbf{X} \in \mathbb{R}^{D^2}$ *and a conditional Gaussian distribution for* $\mathbf{Y} \in \mathbb{R}^T$ *given* $\mathbf{X}$ *in the form*

$$p(\mathbf{x}) = \varphi_T(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}), \tag{A.2}$$

$$p(\mathbf{y}|\mathbf{x}) = \varphi_T(\mathbf{y}; \mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}), \tag{A.3}$$

with $\varphi_T(\cdot; \boldsymbol{\mu}, \boldsymbol{K})$ *denoting the PDF of the* T*-variate normal distribution with mean* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{K}$*. Then, the marginal distribution of* $\mathbf{Y}$ *and the conditional distribution of* $\mathbf{X}$ *given* $\mathbf{Y}$ *are given by*

$$p(\mathbf{y}) = \varphi_T(\mathbf{y}; \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathbf{T}}), \tag{A.4}$$

$$p(\mathbf{x}|\mathbf{y}) = \varphi_L(\mathbf{x}; \boldsymbol{\Sigma}\left(\mathbf{A}^{\mathbf{T}}\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\right), \boldsymbol{\Sigma}), \tag{A.5}$$

where

$$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{A}^{\mathbf{T}}\mathbf{L}\mathbf{A})^{-1}.$$

Test case 1: Hydraulic conductivity fields

We are concerned with the latent log-conductivity field $\boldsymbol{X} \sim f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \varphi_L(\boldsymbol{x}; \boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ and, to express the measurements as a linear function of the latent field, we use the log-transformed Eqs. (18) and (19) out of Sanchez-Vila et al. (2006),

$$log(K^H) = \overline{\boldsymbol{X}} + log\left(1 + \sigma^2(\frac{1}{2} - \frac{1}{1+\lambda})\right) \tag{A.6}$$

$$log(K^V) = \overline{\boldsymbol{X}} + log\left(1 + \sigma^2(\frac{1}{2} - \frac{\lambda}{1+\lambda})\right). \tag{A.7}$$

Then, we denote $\tilde{\boldsymbol{y}} = log(\boldsymbol{y})$ and write $\tilde{\boldsymbol{y}} \approx \boldsymbol{J}\boldsymbol{X} + \boldsymbol{b}$ with $\boldsymbol{J}$ being a $(2, D^2)$-dimensional matrix with constant entries of $\frac{1}{D^2}$ and $\boldsymbol{b} = (b_1, b_2)$, where $b_1 = log\left(1 + \sigma^2(\frac{1}{2} - \frac{1}{1+\lambda})\right)$ and $b_2 = log\left(1 + \sigma^2(\frac{1}{2} - \frac{\lambda}{1+\lambda})\right)$. Subsequently, we approximate $g_{\boldsymbol{\theta}}(\tilde{\boldsymbol{y}}|\boldsymbol{x})$ with $\tilde{g}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{y}}|\boldsymbol{x}) = \varphi_T(\tilde{\boldsymbol{y}}; \boldsymbol{J}\boldsymbol{X} + \boldsymbol{b}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{Y}})$, where $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{Y}} = I_2 * 0.1^2$, with I_2 denoting the two by two identity matrix. With this choice of $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}$, we transfer the observational error to the log-space and artificially increase the uncertainty to account for the errors resulting from the ergodic assumption made to derive the IS distribution. To finally derive an approximation for $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$, we use Eq. (A.5):

$$\tilde{p}(\boldsymbol{x}|\boldsymbol{\theta}, \tilde{\boldsymbol{y}}) = \varphi_L(\boldsymbol{x}; \boldsymbol{\mu}_{\boldsymbol{IS}}, \boldsymbol{\Sigma}_{\boldsymbol{IS}}), \text{ with} \tag{A.8}$$

$$\boldsymbol{\mu}_{\boldsymbol{IS}} = \boldsymbol{\Sigma}_{\boldsymbol{IS}}\left(\boldsymbol{J}^T \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}{}^{-1}(\tilde{\boldsymbol{y}} - \boldsymbol{b}) + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\theta}}\right),$$

$$\boldsymbol{\Sigma}_{\boldsymbol{IS}} = (\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} + \boldsymbol{J}^T \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}{}^{-1}\boldsymbol{J})^{-1}.$$

Test case 2: Fracture aperture fields

We target the fracture aperture field $\boldsymbol{X} \sim f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \varphi_L(\boldsymbol{x}; \boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ and locally approximate $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ by expressing the map $\boldsymbol{x} \mapsto \mathcal{G}(\boldsymbol{x})$ based on a first-order expansion around $\boldsymbol{x}_{lin}$ (as Friedli et al., 2022),

$$\mathcal{G}(\boldsymbol{x}) = \mathcal{G}(\boldsymbol{x}_{lin} + \boldsymbol{x} - \boldsymbol{x}_{lin}) \approx \mathcal{G}(\boldsymbol{x}_{lin}) + \boldsymbol{J}_{\boldsymbol{x}_{lin}}(\boldsymbol{x} - \boldsymbol{x}_{lin}). \tag{A.9}$$

Here, $\boldsymbol{J}_{\boldsymbol{x}_{lin}}$ refers to the sensitivity (Jacobian) matrix of the forward solver corresponding to $\boldsymbol{x}_{lin}$, which is a homogeneous field with the currently proposed geostatistical mean ($\boldsymbol{\mu}_{\boldsymbol{\theta}} = \theta_1$). Subsequently, we approximate $g_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})$ with $\tilde{g}_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}) = \varphi_T(\boldsymbol{y}; \mathcal{G}(\boldsymbol{x}_{lin}) + \boldsymbol{J}_{\boldsymbol{x}_{lin}}(\boldsymbol{x} - \boldsymbol{x}_{lin}), \boldsymbol{\Sigma}_{\boldsymbol{Y}})$. Again, we derive an approximation for $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ by using Eq. (A.5):

$$\tilde{p}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) = \varphi_L(\boldsymbol{x}; \boldsymbol{\mu}_{\boldsymbol{IS}}, \boldsymbol{\Sigma}_{\boldsymbol{IS}}), \text{ with} \tag{A.10}$$

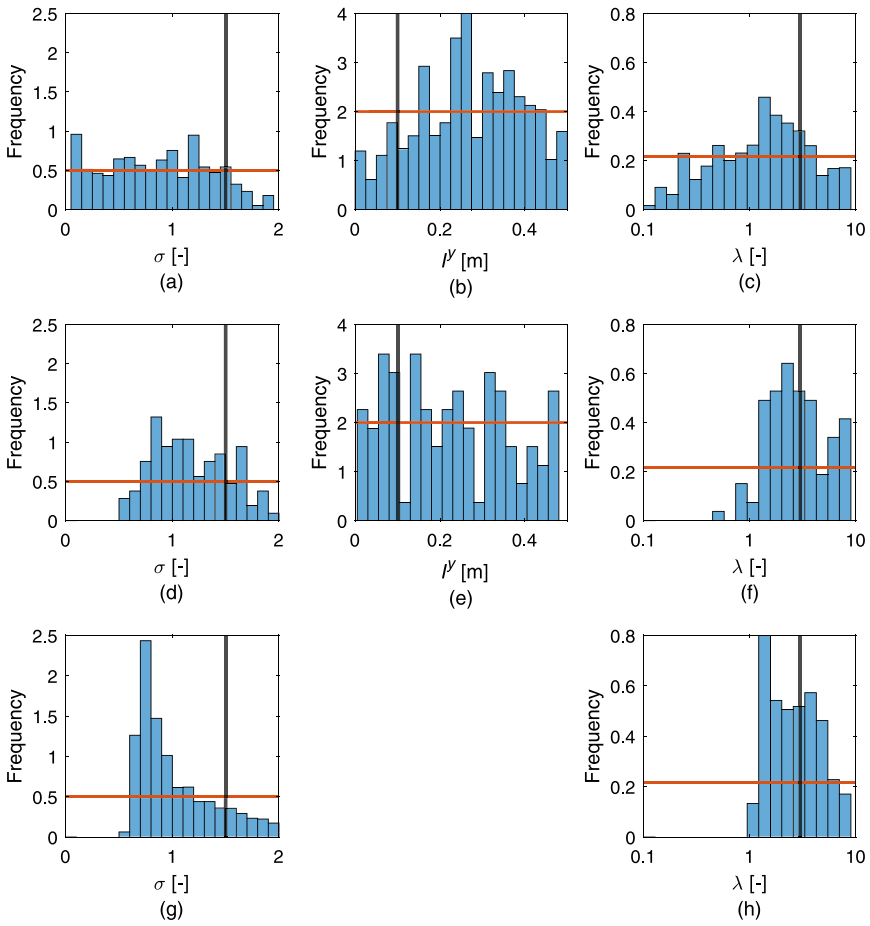


Fig. B.2. Posterior samples for the remaining hyperparameters for the first test case (log-hydraulic conductivity) using the data generated with the non-ergodic setting; (a) standard deviation σ, (b) integral scale I^y and (c) anisotropy factor λ with CPM. (d) Standard deviation σ, (e) integral scale I^y and (f) anisotropy factor λ with RS. (g) Standard deviation σ and (h) anisotropy factor λ with simplified MH. The solid vertical lines indicate the true hyperparameter values and the red horizontal lines the prior PDFs. Note that for the anisotropy factor λ we employ a logarithmic scale on the x-axis.

$$\boldsymbol{\mu}_{IS} = \boldsymbol{\Sigma}_{IS}\left(\boldsymbol{J}_{\boldsymbol{x}_{lin}}^T \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\left(\boldsymbol{y} - (\mathcal{G}(\boldsymbol{x}_{lin}) - \boldsymbol{J}_{\boldsymbol{x}_{lin}}\boldsymbol{x}_{lin})\right) + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\theta}}\right),$$
$$\boldsymbol{\Sigma}_{IS} = (\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} + \boldsymbol{J}_{\boldsymbol{x}_{lin}}^T \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\boldsymbol{J}_{\boldsymbol{x}_{lin}})^{-1}.$$

Since this expression is approximate due to the linearization step, we multiply $\boldsymbol{\Sigma}_{\boldsymbol{Y}}$ with a factor. Following Friedli et al. (2022), we use 1.2 as this choice led to a satisfactory performance.

Appendix B. Complementary figures concerning test case 1

Here we present the additional hyperparameter-plots of the posterior samples obtained for the first test case (Section 3). Fig. B.1 shows the results obtained for the ergodic data setting (Fig. 2(a)) and Fig. B.2 those for the non-ergodic setting (Fig. 2(b)). In the ergodic setting, the posteriors of the standard deviation σ obtained by CPM (Fig. B.1a), RS (Fig. B.1d) and the simplified MH (Fig. B.1g) are as wide as the prior PDFs with the mode of the distributions being located around the right value for all three approaches. Thereby, the posterior obtained with the simplified MH is better defined than the ones of RS and CPM. The integral scale is only inferred by CPM (Fig. B.1b) and RS (Fig. B.1e) with both methods generating posterior samples still distributed proportionally to the Uniform prior PDF. For the anisotropy factor λ, the simplified MH clearly favors values above 1 (Fig. B.1h) and the same holds true for RS (Fig. B.1f). Finally, CPM (Fig. B.1c) samples close to the prior PDF.

Employing the data generated with the non-ergodic setting, we obtain posteriors favoring correctly the horizontal layering of the field, whereby the estimates of RS (Fig. B.2f) and the simplified MH (Fig. B.2h) are better defined than the one of CPM (Fig. B.2c). For I^y, we again obtain estimates close to the prior for both RS (Fig. B.2e) and CPM (Fig. B.2b). While CPM also samples σ close to the prior (Fig. B.2a), the RS realizations show a tendency for higher values (Fig. B.2d). Finally, the simplified MH samples values of σ (Fig. B.2g) having a high concentration at incorrect values.

References

Andrieu, C., Roberts, G.O., 2009. The Pseudo-Marginal approach for efficient Monte Carlo computations. Ann. Statist. 37 (2), 697–725.
Beaumont, M.A., 2003. Estimation of population growth or decline in genetically monitored populations. Genetics 164 (3), 1139–1160.
Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.
Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. Hydrol. Process. 9 (3–4), 251–290.
Bohling, G.C., Liu, G., Dietrich, P., Butler, Jr., J.J., 2016. Reassessing the MADE direct-push hydraulic conductivity data using a revised calibration procedure. Water Resour. Res. 52 (11), 8970–8985.
Bosch, M., 1999. Lithologic tomography: From plural geophysical data to lithology estimation. J. Geophy. Res. Solid Earth 104 (B1), 749–766.
Bradford, J.H., Deeds, J.C., 2006. Ground penetrating radar theory and application of thin-bed offset-dependent reflectivity. Geophysics 71 (1), K47–K57.
Brynjarsdóttir, J., O'Hagan, A., 2014. Learning about physical parameters: The importance of model discrepancy. Inverse Problems 30 (11), 114007.
Butler, J.J., 2005. Hydrogeological methods for estimation of spatial variations in hydraulic conductivity. In: Rubin, Y., Hubbard, S.S. (Eds.), Hydrogeophysics, Vol. 2. Springer, pp. 3–58.
Chen, V., Dunlop, M.M., Papaspiliopoulos, O., Stuart, A.M., 2018. Dimension-robust MCMC in Bayesian inverse problems. arXiv preprint arXiv:1803.03344.
Chiles, J.P., Delfiner, P., 2009. Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons.

Deligiannidis, G., Doucet, A., Pitt, M.K., 2018. The Correlated Pseudo-Marginal method. J. R. Stat. Soc. Ser. B Stat. Methodol. 80 (5), 839–870.
Dentz, M., Le Borgne, T., Englert, A., Bijeljic, B., 2011. Mixing, spreading and reaction in heterogeneous media: A brief review. J. Contam. Hydrol. 120, 1–17.
Deparis, J., Garambois, S., 2008. On the use of dispersive APVO GPR curves for thin-bed properties estimation: theory and application to fracture characterization. Geophysics 74 (1), J1—J12.
Dietrich, C.R., Newsam, G.N., 1997. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. SIAM J. Sci. Comput. 18 (4), 1088–1107.
Doucet, A., Pitt, M.K., Deligiannidis, G., Kohn, R., 2015. Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. Biometrika 102 (2), 295–313.
Friedli, L., Linde, N., Ginsbourger, D., Doucet, A., 2022. Lithological tomography with the correlated pseudo-marginal method. Geophys. J. Int. 228 (2), 839–856.
Gelhar, L.W., Axness, C.L., 1983. Three-dimensional stochastic analysis of macrodispersion in aquifers. Water Resour. Res. 19 (1), 161–180.
Gelman, A., Roberts, G.O., Gilks, W.R., 1996. Efficient Metropolis jumping rules. Bayesian Stat. 5 (42), 599–608.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statist. Sci. 7 (4), 457–472.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Amer. Statist. Assoc. 102 (477), 359–378.
Good, I.J., 1952. Rational decisions. J. R. Stat. Soc. Ser. B 14, 107—114.
Grégoire, C., Hollender, F., 2004. Discontinuity characterization by the inversion of the spectral content of ground penetrating radar (GPR) reflections – Application of the Jonscher model. Geophysics 69 (6), 1414–1424.
Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 7 (2), 223–242.
Hansen, T.M., Cordua, K.C., Looms, M.C., Mosegaard, K., 2013a. SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 1—Methodology. Comput. Geosci. 52, 470—480.
Hansen, T.M., Cordua, K.C., Looms, M.C., Mosegaard, K., 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.
Hansen, T.M., Cordua, K.C., Mosegard, K., 2012. Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling. Comput. Geosci. 16, 593—611.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57 (1), 97—109.
Hess, K.M., Wolf, S.H., Celia, M.A., 1992. Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 3. Hydraulic conductivity variability and calculated macrodispersivities. Water Resour. Res. 28 (8), 2011–2027.
Hill, R., 1963. Elastic properties of reinforced solids: some theoretical principles. J. Mech. Phys. Solids 11 (5), 357–372.
Hunziker, J., Laloy, E., Linde, N., 2017. Inference of multi-Gaussian relative permittivity fields by probabilistic inversion of crosshole ground-penetrating radar data. Geophysics 82 (5), H25–H40.
Jougnot, D., Jiménez-Martínez, J., Legendre, R., Le Borgne, T., Méheust, Y., Linde, N., 2018. Impact of small-scale saline tracer heterogeneity on electrical resistivity monitoring in fully and partially saturated porous media: Insights from geoelectrical milli-fluidic experiments. Adv. Water Resour. 113, 295–309.
Kahn, H., Marshall, A.W., 1953. Methods of reducing sample size in Monte Carlo computations. J. Oper. Res. Soc. Am. 1 (5), 263–278.
Kitanidis, P.K., 1995. Quasi-linear geostatistical theory for inversing. Water Resour. Res. 31 (10), 2411–2419.
Klemeš, V., 1983. Conceptualization and scale in hydrology. J. Hydrol. 65 (1–3), 1–23.
Koop, J.C., 1972. On the derivation of expected value and variance of ratios without the use of infinite series expansions. Metrika 19, 156—170.
Krüger, F., Lerch, S., Thorarinsdottir, T., Gneiting, T., 2021. Predictive inference based on Markov chain Monte Carlo output. Internat. Statist. Rev. 89 (2), 274–301.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.
Laloy, E., Linde, N., Diederik, J., Vrugt, J.A., 2015. Probabilistic inference of multi-Gaussian fields from indirect hydrological data using circulant embedding and dimensionality reduction. Water Resour. Res. 51, 4224–4243.
Maurer, H., Friedel, S., 2006. Outer-space sensitivities in geoelectrical tomography. Geophysics 71 (3), G93–G96.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21 (6), 1087–1092.
National Research Council (NRC), 1996. Committee on fracture characterization and fluid flow. In: Rock Fractures and Fluid Flow: Contemporary Understanding and Applications. National Academies Press.
Neuman, S.P., Di Federico, V., 2003. Multifaceted nature of hydrogeologic scaling and its interpretation. Rev. Geophys. 41 (3), 4.1–4.31.
Owen, A., Zhou, Y., 2000. Safe and effective importance sampling. J. Amer. Statist. Assoc. 95 (449), 135–143.
Rehfeldt, K.R., Boggs, J.M., Gelhar, L.W., 1992. Field study of dispersion in a heterogeneous aquifer: 3. Geostatistical analysis of hydraulic conductivity. Water Resour. Res. 28 (12), 3309–3324.
Renard, P., De Marsily, G., 1997. Calculating equivalent permeability: a review. Adv. Water Resour. 20 (5–6), 253–278.
Ripley, B.D., 1987. Stochastic Simulations. J. Wiley & Sons.
Robert, C., Casella, G., 2013. Monte Carlo Statistical Methods. Springer Science and Business Media..
Robert, C.P., Elvira, V., Tawn, N., Wu, C., 2018. Accelerating MCMC algorithms. Wiley Interdiscip. Rev. Comput. Stat. 10 (5), e1435.
Rubin, Y., Chen, X., Murakami, H., Hahn, M., 2010. A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields. Water Resour. Res. 46 (10).
Sanchez-Vila, X., Guadagnini, A., Carrera, J., 2006. Representative hydraulic conductivities in saturated groundwater flow. Rev. Geophys. 44 (3).
Shakas, A., Linde, N., 2015. Effective modeling of ground penetrating radar in fractured media using analytic solutions for propagation, thin-bed interaction and dipolar scattering. J. Appl. Geophys. 116, 206–214.
Shakas, A., Linde, N., 2017. Apparent apertures from ground penetrating radar data and their relation to heterogeneous aperture fields. Geophys. J. Int. 209 (3), 1418–1430.
Shakas, A., Linde, N., Le Borgne, T., Bour, O., 2018. Probabilistic inference of fracture-scale flow paths and aperture distribution from hydrogeophysically-monitored tracer tests. J. Hydrol. 567, 305–319.
Torquato, S., Haslach, Jr., H.W., 2002. Random heterogeneous materials: microstructure and macroscopic properties. Appl. Mech. Rev. 55 (4), B62–B63.
Tsang, Y.W., 1992. Usage of "equivalent apertures" for rock fractures as derived from hydraulic and tracer tests. Water Resour. Res. 28 (5), 1451–1455.
Tsoflias, G.P., Perll, C., Baker, M., Becker, M.W., 2015. Cross-polarized GPR imaging of fracture flow channeling. J. Earth Sci. 26 (6), 776–784.
Visentini, A.F., Linde, N., Le Borgne, T., Dentz, M., 2020. Inferring geostatistical properties of hydraulic conductivity fields from saline tracer tests and equivalent electrical conductivity time-series. Adv. Water Resour. 146, 103758.
Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. Environ. Model. Softw. 75, 273–316.
Wang, L., Kitanidis, P.K., Caers, J., 2022. Hierarchical Bayesian inversion of global variables and large-scale spatial fields. Water Resour. Res. 58 (5), e2021WR031610.
Xiao, S., Xu, T., Reuschen, S., Nowak, W., Hendricks Franssen, H.J., 2021. Bayesian inversion of multi-Gaussian log-conductivity fields with uncertain hyperparameters: An extension of preconditioned Crank–Nicolson Markov chain Monte Carlo with parallel tempering. Water Resour. Res. 57 (9), e2021WR030313.
Yu, X., Michael, H.A., 2021. Impacts of the scale of representation of heterogeneity on simulated salinity and saltwater circulation in coastal aquifers. Water Resour. Res. e2020WR029523.
Zhao, Y., Luo, J., 2021. Bayesian inverse modeling of large-scale spatial fields on iteratively corrected principal components. Adv. Water Resour. 151.