Bayesian Inversion of Multi-Gaussian Log-Conductivity Fields With Uncertain Hyperparameters: An Extension of Preconditioned Crank-Nicolson Markov Chain Monte Carlo With Parallel Tempering

Sinan Xiao¹², Teng Xu¹³, Sebastian Reuschen¹, Wolfgang Nowak¹ and Harrie-Jan Hendricks Franssen²

¹Department of Stochastic Simulation and Safety Research for Hydrosystems, Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Stuttgart, Germany, ²Institute of Bio- and Geosciences (IBG-3): Agrosphere, Forschungszentrum Jülich GmbH, Jülich, Germany, ³State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China

Abstract In conventional Bayesian geostatistical inversion, specific values of hyperparameters characterizing the prior distribution of random fields are required. However, these hyperparameters are typically very uncertain in practice. Thus, it is more appropriate to consider the uncertainty of hyperparameters as well. The preconditioned Crank-Nicolson Markov chain Monte Carlo with parallel tempering (pCN-PT) has been used to efficiently solve the conventional Bayesian inversion of high-dimensional multi-Gaussian random fields. In this study, we extend pCN-PT to Bayesian inversion with uncertain hyperparameters of multi-Gaussian fields. To utilize the dimension robustness of the preconditioned Crank-Nicolson algorithm, we reconstruct the problem by decomposing the random field into hyperparameters and white noise. Then, we apply pCN-PT with a Gibbs split to this “new” problem to obtain the posterior samples of hyperparameters and white noise, and further recover the posterior samples of spatially distributed model parameters. Finally, we apply the extended pCN-PT method for estimating a finely resolved multi-Gaussian log-hydraulic conductivity field from direct data and from head data to show its effectiveness. Results indicate that the estimation of hyperparameters with hydraulic head data is very challenging and the posterior distributions of hyperparameters are only slightly narrower than the prior distributions. Direct measurements of hydraulic conductivity are needed to narrow more the posterior distribution of hyperparameters. To the best of our knowledge, this is a first accurate and fully linearization free solution to Bayesian multi-Gaussian geostatistical inversion with uncertain hyperparameters.

1. Introduction

The characterization of hydraulic properties of aquifers and soils is essential to better predict water flow in the subsurface and the transport of heat or solutes. Typically, not enough direct data (e.g., hydraulic conductivity) are available to characterize the heterogeneous subsurface. Thus, additional indirect data (e.g., hydraulic heads) are important for improving characterization and, in turn, predictions by subsurface flow and transport models. In the geostatistical context, the resulting inverse problem for subsurface problems is typically underdetermined. Therefore, approaches were developed which limited the number of independent parameters to be estimated, either by defining a limited number of zones with constant parameters (Carrera & Neuman, 1986) or parameterizing the spatially variable parameter field by a geostatistical function with a few unknown parameters (Kitanidis & Vomvoris, 1983). Later, methods were formulated to estimate a series of equally likely solutions to the groundwater inverse problem, either by an ensemble-based variational data assimilation approach (Gómez-Hernández et al., 1997) or by a sequential data assimilation approach for an ensemble of random parameter fields (Chen & Zhang, 2006). In summary, the combination of regularization and casting the problem in a stochastic framework, helped to tackle groundwater inverse problems.

Bayesian inversion has been widely used for model parameter inference (Bui-Thanh et al., 2013; Buland & Omre, 2003; Cotter et al., 2009; Mariethoz et al., 2010; Stuart, 2010). However, it is often not trivial to define the prior distribution, for example, there may not be enough information to confidently determine the prior mean and prior variance (Malinverno & Briggs, 2004). In addition, in subsurface hydrological problems, we are typically
dealing with heterogeneous parameter fields (e.g., hydraulic conductivity) to be estimated, and we need to define a spatial covariance function which characterizes this spatial variability. The hyperparameters that define the spatial covariance function (e.g., nugget, range, and sill) are typically very uncertain. This uncertainty roots back to the limited amount of direct measurements of the parameter of interest, to very incomplete geological information and to the absence of geophysical surveys that could provide indirect information on spatial correlation structures. Nevertheless, the large majority of geostatistical inversion studies does not consider hyperparameter uncertainty.

A seminal cornerstone to tackle this problem was the quasi-linear geostatistical approach by Kitanidis (1995), where parameters that govern the covariance function are jointly iterated with the spatial heterogeneous field of transmissivity. This approach actually includes the restricted maximum likelihood algorithm to update the hyperparameters for the spatial covariance model. It provides a first-order estimate of hyperparameters, together with a first-order variance to specify the remaining hyperparameter uncertainty. However, most of the published studies using the quasi-linear geostatistical approach did not implement the estimation of the hyperparameters, with only few exceptions (Malinverno & Briggs, 2004; Nagel & Sudret, 2016; Nowak & Cirpka, 2006; Woodbury & Ulrych, 2000). Another geostatistical approach, the successive linear estimator (Yeh et al., 1996; Zha et al., 2018), updates the spatial covariance using the posterior covariance, which is sort of similar to the update of hyperparameters, if the inverted data are strong enough to dominate over the prior. Recently, Zhao and Luo (2021a) also used the posterior covariance for the correction of biased prior hyperparameters.

Many hydrogeological practitioners may just want the best-estimate solution to an inverse problem, not its uncertainty. Yet, as the hyperparameters govern the structure of the best estimate, it is still important to estimate them well (i.e., linearization-free). This is important due to several reasons: (a) the hyperparameters control the regularization of the best-estimate inversion result by the prior covariance, and thus should be chosen well; (b) they are the only information available about the spatial variability unresolved by the best estimate, which appears both in the estimation variance and in the structure of conditional realizations; (c) both the resolved structure and the unresolved heterogeneity are highly important for predicting scale-sensitive processes such as contaminant transport. This third reason becomes apparent in classical studies on macrodispersion (Dagan, 1988), on effective dispersion and contaminant dilution (Dentz et al., 2000), and in the fact that transport simulations on best-estimate fields clearly lack dispersion (Cirpka & Nowak, 2003; Nowak & Cirpka, 2006). One may argue whether one needs to know the values of hyperparameters as such. But one must know the structure, and the structure is governed by the hyperparameters.

Once acknowledging the relevance of inferring the structure (i.e., the hyperparameters), the joint estimation problem of hyperparameters and random field is an extended Bayesian inversion, which is also known as “hierarchical Bayes” (Gelman et al., 2013), where the hyperparameters are also uncertain. Now, the final result of Bayesian inversion will be the joint posterior distribution of model parameters and hyperparameters that govern the spatial covariance function.

When sufficient data are available, the joint posterior distribution will be dominated by the data. Several studies did synthetic test cases to show that enough information can override a “wrong fixed” choice of hyperparameters (if not too far off), especially EnKF-based studies (Li et al., 2012). However, in real-world applications, this situation will hardly appear: Research sites such as MADE (Boggs et al., 1992) or Borden Airforce Base (Sudicky, 1986) have a dense enough instrumentation, but practical sites (e.g., in remediation Trolldborg et al., 2012) do not.

In most cases, it is not possible to get an analytical expression of the posterior distribution in hierarchical Bayes, especially for a nonlinear inverse problem. The above studies all rest on linearization concepts and then utilize an analytical first-order expression in an iterative scheme. At large variances and for strongly nonlinear problems, these methods have limitations. A widely used method to approximate the inverse solution is the sampling method Markov chain Monte Carlo (MCMC). Within MCMC, the random field is discretized to enable numerical computation. However, a refinement of the discretization (or an increase of the problem size), resulting in more parameters to be estimated, will usually lead to slower convergence rates of plain MCMC methods, such as the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953). In the literature, different approaches have been presented to overcome this challenge. The pre-conditioned Crank Nicolson MCMC (pCN-MCMC) ( Cotter et al., 2013) is one of these approaches. Its main advantage is that the acceptance probability for proposed solutions only depends on the proposal's likelihood to match with the data, not on the proposal's position in the
prior distribution. However, when pCN-MCMC is applied to nonlinear inverse problems with multi-modal posterior distributions, it may still need a long chain and a large potential space may still be missed since the MCMC simulation proceeds by local jumps in the vicinity of the current state (Robert et al., 2018). To deal with this problem, parallel tempering (Altekar et al., 2004; Earl & Deem, 2005) is a good candidate, which runs multiple chains with different temperatures simultaneously. The hot chains can more easily explore the whole parameter space, while the cold chains perform precise sampling in high-likelihood regions of the parameter space.

In a previous study (Xu et al., 2020), pCN-MCMC was combined with parallel tempering (pCN-PT) for Bayesian inversion of multi-Gaussian fields with fixed hyperparameters. The focus was to gain efficiency by combining these two ideas, while still focusing on the simpler problem with fixed hyperparameters. In this study, we will extend pCN-PT to Bayesian inversion with uncertain hyperparameters of multi-Gaussian fields. That means, our focus now is to extend the applicability of the highly efficient pCN-PT to a harder, wider and more realistic problem. If we apply pCN-MCMC or pCN-PT to the extended Bayesian inversion directly, the acceptance probability would still depend on the prior (Malinverno, 2002), which destroys the efficiency trick of pCN. To let the acceptance probability only depend on the likelihood, we will reconstruct the problem, that is, we decompose the random field model parameters into hyperparameters and white noise. The latter, after coloring through the covariance function, represents the subsurface field of hydraulic properties. Then, we have a “new” problem with hyperparameters and white noise as the primary inversion parameters. Thanks to this reconstruction, the acceptance probability is now only depending on the likelihood when using pCN-MCMC or pCN-PT. This allows to approach the hyperparameter inversion and the white noise inversion with specialized sub-algorithms within a joint iteration. In addition, this will also allow us to assess the feasibility to estimate hyperparameters. As indicated, hyperparameters are associated with significant uncertainty, and constraining this uncertainty by inversion is important. Previous inversion studies relied on linearization and it remains unclear to which degree measurement data like hydraulic conductivity and hydraulic heads carry enough information to reduce uncertainty with respect to hyperparameters. This study will also give more insight on the value of hydraulic conductivity and hydraulic head data for constraining hyperparameters with possible implications for the design of monitoring network.

The rest of this study is organized as follows. Section 2 gives the definition of the extended Bayesian inversion problem and the extended pCN-PT method (specifically in Section 2.4). In Section 3, we introduce a model setup, algorithmic settings, test cases for demonstration and testing criteria. Section 4 shows the results. Finally, Section 5 gives the conclusion.

2. Methodology

2.1. Bayesian Geostatistical Inversion With Uncertain Hyperparameters

Our forward problem can be formulated as follows:

\[ d = M(\theta) + e, \quad (1) \]

where \( M \) is a deterministic forward model (e.g., for groundwater flow), \( \theta \) denotes the uncertain (hydraulic) parameters for the (groundwater flow) model, \( d \) represents the measurements, and \( e \) contains measurement errors. Parameter \( \theta \) is a vector containing the discretized values of the random space function. The purpose of Bayesian geostatistical inversion is to infer the posterior distribution of uncertain parameters \( \theta \) given the data \( d \) and prior knowledge about \( \theta \). Then, we can make a better estimation of the uncertain parameter \( \theta \).

The posterior distribution of the uncertain parameters \( \theta \) can be obtained through Bayes’ rule as (Congdon, 2003)

\[ p(\theta | d) = \frac{p(\theta)p(d|\theta)}{p(d)} \propto p(\theta)p(d|\theta), \quad (2) \]

where \( p(\theta) \) is the prior distribution of the unknown (hydraulic) parameters, which describes the prior knowledge of \( \theta \) independently of the data \( d \); \( p(d|\theta) \) is the likelihood function, which quantifies how probable are the measurement data \( d \) for a given realization of uncertain parameters \( \theta \); and \( p(d) \) is the marginal likelihood.

The marginal likelihood \( p(d) \) is a normalizing factor for the posterior distribution. Since the marginal likelihood is not a function of parameters \( \theta \), it is often ignored when making inference of parameters. This allows writing
the posterior distribution as proportional to the product of prior and likelihood, as reflected in the last part of Equation 2 (Duijndam, 1988; Cary & Chapman, 1988).

The prior distribution of parameters $\theta$ is dependent on some hyperparameters $h$, such as the mean and parameters that govern a covariance function. In a conventional Bayesian geostatistical inversion, fixed values are given to $h$. Generally, we can also let the hyperparameters have their own uncertainty and assume a prior distribution for them to reflect our prior knowledge of hyperparameters (Kitanidis, 1995; Woodbury & Ulrych, 2000). Then, we can similarly update this prior distribution and get the corresponding posterior distribution. This is an extended Bayesian inversion (Malinverno & Briggs, 2004), and we can get the joint posterior distribution of model parameters $\theta$ and hyperparameters $h$ through Bayes' rule as

$$p(\theta, h | d) = \frac{p(\theta, h) p(d | \theta, h)}{p(d)} \propto p(\theta, h) p(d | \theta, h),$$

where $p(\theta, h)$ is the joint prior distribution of $\theta$ and $h$, $p(d | \theta, h)$ is the likelihood and $p(d)$ is the marginal likelihood.

Using the definition of conditional distribution, we can write the joint prior and posterior distributions as

$$p(\theta, h) = p(\theta | h) p(h) \tag{4}$$

$$p(\theta, h | d) = p(\theta | h, d) p(h | d) \tag{5}$$

where $p(h)$ and $p(h|d)$ are the prior and posterior distributions of $h$ separately, while $p(\theta | h)$ and $p(\theta | h, d)$ are the prior and posterior distributions of $\theta$ separately for a given choice of $h$. We can see that the distribution of model parameters $\theta$ depends on hyperparameters $h$. Based on the joint posterior distribution of $\theta$ and $h$, we can get the marginal prior and posterior distributions of $\theta$ and $h$ by marginalization. In general, the integral required for marginalization cannot be solved in closed form. This is the problem solved by Kitanidis (1995) by restricted maximum likelihood in successive linearization. Sampling methods, such as Markov chain Monte Carlo, can avoid this need for an explicit calculation of the integral, the need for restrictions and the need for linearization.

### 2.2. Markov Chain Monte Carlo

MCMC is an interesting method to get samples from a target probability distribution. It has been found to be well suited for Bayesian inversion problems (Besag et al., 1995; Gelman & Rubin, 1992; Grandis et al., 1999; Mosegaard & Tarantola, 1995; Schott et al., 1999; Tierney, 1994). Basically, MCMC algorithms construct a Markov chain, that is, a specific sequence of realizations of the solution where the next realization in the sequence only depends on the previous realization. For sampling, the equilibrium distribution of that chain is equal to the target distribution. To guarantee this property, the most important rule for the transition from the current realization to the next is that the so-called detailed balance condition has to be satisfied. After an initial “burn-in” period that is, influenced by the initialization of the chain (e.g., van Ravenzwaaij et al., 2018), the Markov chain samples the target distribution, that is, it provides a set of realizations following the target distribution. For Bayesian inversion problems, the posterior distribution is the target distribution. We can sample it with an MCMC and then we can easily estimate any desired properties of the posterior distribution, such as mean and variance or relevant probabilities. An important property of MCMC is its memory mechanism, which can make the Markov chain stay in the parameter space with high posterior probability (Malinverno, 2002). This property makes MCMC more efficient compared to other sampling approaches such as rejection sampling, but only if the MCMC is adapted well to the problem at hand.

At each stage with the current sample $\theta$, one needs to propose a candidate sample $\tilde{\theta}$ based on a proposal distribution $q(\theta | \tilde{\theta})$. Then, one needs to decide whether to accept the candidate as the next sample with the following acceptance probability (Hastings, 1970):
\[
\alpha = \min \left\{ \frac{\tilde{p}(\tilde{\theta} | d) q(\tilde{\theta} | \theta)}{\tilde{p}(\tilde{\theta} | d) q(\tilde{\theta} | \theta)} \right\} = \min \left\{ \frac{p(\tilde{\theta}) p(\tilde{\theta} | d) q(\tilde{\theta} | \theta)}{p(\tilde{\theta}) p(\tilde{\theta} | d) q(\tilde{\theta} | \theta)} \right\}. \tag{6}
\]

If the candidate sample \( \tilde{\theta} \) is accepted, the next sample of the chain will be changed into \( \tilde{\theta} \). Otherwise, the chain will stay at \( \theta \).

For the case of extended Bayesian inversion, we need to extend Equation 6 from \( \theta \) to \( (\theta, h) \). Then, the corresponding acceptance probability can be written as

\[
\alpha = \min \left\{ \frac{p(\tilde{\theta}, \tilde{h}) p(\tilde{h} | d, \tilde{h}) q(\tilde{\theta}, \tilde{h} | \theta, \tilde{h})}{p(\tilde{\theta}, \tilde{h}) p(\tilde{h} | d, \tilde{h}) q(\tilde{\theta}, \til{h} | \theta, \til{h})} \right\}. \tag{7}
\]

Since the model parameters \( \theta \) depend on the hyperparameters \( h \), it is intuitive to divide the proposal of a candidate into two steps: first propose a candidate hyperparameter set \( \tilde{h} \) based on the current hyperparameter set \( h \), then propose a candidate model parameter field \( \tilde{\theta} \) based on \( \tilde{h} \) and on the current model parameter field \( \theta \). Thus, the proposal distribution can be written as (Malinverno, 2002)

\[
q(\tilde{\theta}, \til{h} | \theta, h) = q(\til{h} | h) q(\til{\theta} | \til{h}, \theta). \tag{8}
\]

Then, the acceptance probability can be written as

\[
\alpha = \min \left\{ \frac{p(\tilde{\theta}, \til{h}) p(\til{h} | d, \til{h}) q(\til{\theta}, \til{h} | \theta, \til{h})}{p(\til{\theta}, \til{h}) p(\til{h} | d, \til{h}) q(\til{\theta}, \til{h} | \theta, \til{h})} \right\}. \tag{9}
\]

Generally, we can choose (almost) any proposal distribution \( q \). However, different proposal distributions will affect the convergence in a given problem. Therefore, it is important to choose a proper \( q \) to have a fast convergence.

### 2.3. pCN-MCMC With Fixed Hyperparameters

pCN-MCMC (Cotter et al., 2013) combines the pCN proposal with the MCMC algorithm. It is an approach where the proposal automatically samples from the prior distribution if the prior is (multi-)Gaussian. A significant feature of the pCN-MCMC algorithm is its dimension robustness (Hairer et al., 2014), which makes it interesting for high-dimensional sampling problems, such as finely resolved multi-Gaussian random fields.

In pCN-MCMC, one assumes that the prior \( p(\theta) \) follows a multivariate normal distribution, that is, \( \theta \sim \mathcal{N}(\mu, \Sigma) \). For these priors, the candidate sample in pCN-MCMC can be written as

\[
\tilde{\theta} = \sqrt{1 - \beta^2} (\theta - \mu) + \beta \xi + \mu, \xi \sim \mathcal{N}(0, \Sigma). \tag{10}
\]

The parameter \( \beta \) is a jumping factor that can be chosen freely (or optimized for statistical efficiency), and it follows the constraint \( 0 < \beta < 1 \). The corresponding proposal can also be written as \( q(\tilde{\theta} | \theta) = \mathcal{N}(\sqrt{1 - \beta^2} (\theta - \mu) + \mu, \beta^2 \Sigma) \).

It fulfills the following condition

\[
\frac{q(\tilde{\theta} | \theta)}{q(\theta | \tilde{\theta})} = \frac{p(\tilde{\theta})}{p(\theta)}, \tag{11}
\]

which confirms that it samples from the prior per construction. Inserting Equation 11 into Equation 6, the acceptance probability \( \alpha \) can be simplified as

\[
\alpha = \min \left\{ \frac{p(d | \tilde{\theta})}{p(d | \theta)} \right\}. \tag{12}
\]
Compared to the acceptance probability in conventional MCMC (Equation 6), the acceptance probability $\alpha$ with the pCN proposal (Equation 12) is only dependent on the likelihood. This contributes to a relatively large increase of efficiency, because there will be no rejections due to the prior.

### 2.4. Extending pCN-MCMC for Uncertain Hyperparameters

Now, we extend, as our key contribution, the pCN-MCMC to the case of uncertain hyperparameters while maintaining the highly efficient acceptance probability as in Equation 12. For the case of extended Bayesian inversion, we can also use the pCN proposal for $q(\tilde{h}|h)$ and $q(\tilde{\theta}|\tilde{h}, \theta)$. The pCN proposal of hyperparameters can be written as

$$q(\tilde{h}|h) = N\left(\sqrt{1 - \beta_h^2} (h - \mu_h) + \mu_h, \beta_h^2 \Sigma_h \right),$$

where $\mu_h$ is the mean of $h$, $\Sigma_h$ is the covariance matrix of $h$ and $0 < \beta_h < 1$ is the jumping factor. For Equation 13, we still have the following condition

$$\frac{q(\tilde{h}|h)}{q(h|h)} = \frac{p(\tilde{h})}{p(h)}.$$

The pCN proposal for the model parameters can be written as

$$q(\tilde{\theta}|\tilde{h}, \theta) = N\left(\sqrt{1 - \beta_\theta^2} (\theta - \tilde{\theta}) + \tilde{\theta}, \beta_\theta^2 \tilde{\Sigma}_\theta \right),$$

where $\tilde{\theta}$ denotes the mean components in the hyperparameters $\tilde{h}$, $\tilde{\Sigma}_\theta$ denotes covariance components in the hyperparameters $\tilde{h}$ and $0 < \beta_\theta < 1$ is the jumping factor. However, due to the difference between $h$ and $\tilde{h}$, we have

$$\frac{q(\tilde{\theta}|\tilde{h}, \theta)}{q(\tilde{\theta}|\tilde{h}, \theta)} \neq \frac{p(\tilde{\theta}|\tilde{h})}{p(\tilde{\theta}|\tilde{h})}.$$

Therefore, the acceptance probability will still depend on the prior. One can find more details in Malinverno (2002). To let the acceptance probability stay independent of the prior just as in Equation 12, we will do the following reconstruction.

For the model parameters $\theta$ following a conditional multivariate Gaussian distribution (for given values of the hyperparameter $h$), that is, $\theta \sim N(\mu, \Sigma)$ ($\mu$ and $\Sigma$ are the hyperparameters), we can rewrite (Alabert, 1987; Davis, 1987)

$$\theta = \mu + Aw,$$

where $\Sigma = AA^T$, $A = A(h)$ is a matrix square root of $\Sigma = \Sigma(h)$, and $w$ is a standard normal random vector ($w \sim N(0, I)$, $I$ is the identity matrix). We will denote $w$ as white noise from here onwards. We can see that the model parameters are decomposed into their hyperparameters $h$ (i.e., to define $\mu$ and $\Sigma$ (or $A$)) and white noise $w$. Then, the forward problem from Equation 1 can be rewritten as

$$d = M(\theta(h, w)) + \epsilon.$$

Now, we can argue that $h$ and $w$ are the “fundamental model parameters” in Equation 18. Then, we can infer the posterior distribution of hyperparameters $h$ and white noise $w$ at first, and then reconstruct the posterior distribution of the original model parameters $\theta$ based on $h$ and $w$ through Equation 17. In this fashion, the extended Bayesian inversion problem can also be rewritten as

$$p(h, w|d) \propto p(h, w)p(d|h, w).$$

Since the hyperparameters $h$ and white noise $w$ are independent, the joint prior distribution $p(h, w)$ can be written as
\[ p(h, w) = p(h)p(w), \]  

(20)

and the proposal can also be written as

\[ q(\tilde{h}, \tilde{w}|h, w) = q(\tilde{h}|h)q(\tilde{w}|w). \]

(21)

Therefore, the acceptance probability can be written as

\[
\alpha = \min \left\{ \frac{p(\tilde{h})p(\tilde{w})p(d|\tilde{h}, \tilde{w})q(h|\tilde{h})q(w|\tilde{w})}{p(h)p(w)p(d|h, w)q(h|h)q(w|w)}, 1 \right\},
\]

(22)

Now, we can use the pCN proposal for \( h \) and \( w \) separately. The pCN proposal for \( h \) is the same as in Equation 13. The pCN proposal for \( w \) can be written as

\[ q(\tilde{w}|w) = \mathcal{N}(\sqrt{1 - \beta_w^2} w, \beta_w^2 I), \]

(23)

Now, we will have

\[
\frac{p(\tilde{h})}{p(h)} = \frac{q(\tilde{h}|h)}{q(h|h)}, \quad \frac{p(\tilde{w})}{p(w)} = \frac{q(\tilde{w}|w)}{q(w|w)},
\]

(24)

and then, the acceptance probability can again be simplified as

\[
\alpha = \min \left\{ \frac{p(d|\tilde{h}, \tilde{w})}{p(d|h, w)}, 1 \right\}.
\]

(25)

In Equations 14 and 24, we assumed that the prior distributions of hyperparameters and white noise follow multivariate Gaussian distributions. This is true for white noise since it is i.i.d. standard normal (the simplest case of multivariate Gaussian). For the hyperparameters \( h = (h_1, h_2, \ldots, h_k)^T \), the prior distribution may not be a Gaussian. Then, assuming that the hyperparameters are independent from each other, we can do an isoprobabilistic transformation at first, that is, \( \nu_i = \Phi^{-1}(F_i(h_j)), i = 1, 2, \ldots, k \). Here, \( F_i \) is the cumulative distribution function of \( h_i \) and \( \Phi \) is the cumulative distribution function of the standard normal distribution. Then, \( \nu_i \) will be a random variable following the standard normal distribution, and we can just perform the pCN procedure on \( \nu = (\nu_1, \nu_2, \ldots, \nu_k)^T \), which is also i.i.d. standard normal. The final posterior samples of hyperparameters \( h \) can be obtained based on the samples of \( \nu \) through the isoprobabilistic back-transformation, that is, \( h_i = F_i^{-1}(\Phi(\nu_i)) \). In the following, we assume that hyperparameters follow standard normal distribution just for simple expression.

2.5. Parallel Tempering

Parallel tempering (Altekar et al., 2004; Earl & Deem, 2005; Hukushima & Nemoto, 1996), also called Metropolis coupled MCMC, is a method for improving traditional MCMC algorithms for multi-modal distributions. In parallel tempering, multiple Markov chains are simulated at different temperatures simultaneously. The lower-temperature chains perform precise sampling in high-density regions of the parameter space, but they could easily get stuck in local modality. The higher-temperature chains can more easily explore the whole parameter space due to their flatter and broader likelihood. These different chains will regularly swap their members in a way that preserves the detailed balance condition. Therefore, the hotter chains will make sure the coldest chain with unit temperature (target chain) can access the desired regions of the parameter space. Several studies have also shown the superiority of parallel tempering compared to simple Monte Carlo, simulated annealing and simple MCMC (Moreno et al., 2003; Xu et al., 2020).

To implement parallel tempering, a temperature ladder \( T_1 < T_2 < \cdots < T_n \), with \( T_1 = 1 \), needs to be designed at first. Then, we can get a series of tempered posterior distributions corresponding to different temperatures:

\[
\frac{1}{p_i(h, w|d) \propto p(h, w)p(d|h, w)^{\frac{1}{T_i}}}, i = 1, \ldots, n.
\]

(26)
Apparently, the likelihood is taken to the power of the inverse temperature. Therefore, temperature $T_i = 1$ corresponds to the original target posterior distribution. With increasing the temperature, the corresponding likelihood will become flatter, with a fully flat likelihood (i.e., just the prior) at the limit of $T_i \to \infty$.

Then, for the $i$th chain with temperature $T_i$, if we use the pCN proposal, the acceptance probability in Equation 25 would turn into

$$
\alpha = \min \left\{ \frac{p(d|h, w)}{p(d|h, w)}^{1/T_i}, 1 \right\}.
$$

(27)

For two different chains with temperatures $T_i$ and $T_j$, there would be a potential swap between the states of them, that is,

$$
\{(h_i, w_i, T_i), (h_j, w_j, T_j)\} \to \{(h_j, w_j, T_j), (h_i, w_i, T_i)\},
$$

(28)

where $(h_i, w_i)$ and $(h_j, w_j)$ are the states of two chains corresponding to the temperatures $T_i$ and $T_j$ before the potential swap. The acceptance probability of this potential swap can be written as

$$
\alpha_s = \min \left\{ \frac{p(d|h_j, w_j)}{p(d|h_i, w_i)}^{1/T_i} \cdot 1 \right\}.
$$

(29)

Two candidate chains need to be determined at first to implement the potential swap. A common choice is restricting them to the neighboring chains (Earl & Deem, 2005), which will be used in this work.

### 2.6. Final Algorithm: Extended pCN-PT With Gibbs Split

In a recent study (Xu et al., 2020), pCN-PT (pCN-MCMC and parallel tempering) was used for the conventional Bayesian inversion with fixed values of hyperparameters. In this work, we will extend it to the extended Bayesian inversion problem with uncertain hyperparameters in Equation 19.

Since hyperparameters and white noise play different roles for generating a Gaussian random field and are a priori mutually independent (see Equation 20), we can update them separately during the MCMC process. This corresponds to a classical Gibbs sampler (Casella & George, 1992), blocked together to split between $h$ and $w$. It allows fine-tuning the corresponding MCMC parameters separately (here: $\beta_h$ and $\beta_w$) for improved efficiency. Considering the independence between hyperparameters $h$ and white noise $w$, the pCN-PT algorithm can be summarized as follows:

1. Initialize a temperature ladder $T_1 < T_2 < \cdots < T_n$ with $T_1 = 1$, two pCN jumping factor ladders $\beta_{h,1}, \beta_{h,2}, \ldots, \beta_{h,n}$, and $\beta_{w,1}, \beta_{w,2}, \ldots, \beta_{w,n}$ corresponding to hyperparameters and white noise separately. Then, set initial realizations $h_i^{(0)}$ and $w_i^{(0)}$, where $i \in \{1, 2, \ldots, n\}$ is the number of the chain in the parallel tempering.
2. Generate a pCN proposal $\tilde{h}_i^{(k)}$ of hyperparameters at the $k$th sampling iteration for each chain $i = 1, 2, \ldots, n$,

$$
\tilde{h}_i^{(k)} = \sqrt{1 - \beta_{h,i}^{(k)}} h_i^{(k)} + \beta_{h,i}^{(k)} y_h^{(k)} \sim \mathcal{N}(0, I).
$$

(30)

3. For each chain $i$, accept or reject $\tilde{h}_i^{(k)}$:

$$
\begin{cases} 
\tilde{h}_i^{(k+1)} = \tilde{h}_i^{(k)} & \text{with probability } \alpha(h_i^{(k)}, \tilde{h}_i^{(k)}) \\
\tilde{h}_i^{(k+1)} = \tilde{h}_i^{(k)} & \text{otherwise} 
\end{cases}
$$

(31)

where
4. Generate a pCN proposal \( \tilde{w}_i^{(k)} \) of white noise at the \( k \)th sampling iteration for each chain \( i = 1,2,\ldots,n \),
\[
\tilde{w}_i^{(k)} = \sqrt{1 - \beta_w^2} \tilde{w}_i^{(k)} + \beta_w \xi_i, \xi_i \sim \mathcal{N}(0, I).
\]

5. For each chain \( i \), accept or reject \( w_i^{(k)} \):
\[
w_{i}^{(k+1)} = \begin{cases} 
\tilde{w}_i^{(k)} & \text{with probability } \alpha(w_i^{(k)}, \tilde{w}_i^{(k)}) \\
 w_i^{(k)} & \text{otherwise} 
\end{cases}
\]
where
\[
\alpha(w_i^{(k)}, \tilde{w}_i^{(k)}) = \min \left\{ \frac{p(d|\tilde{h}_i^{(k)}, w_i^{(k)})}{p(d|h_i^{(k)}, w_i^{(k)})}, 1 \right\}.
\]

6. For the neighboring chains \( i \) and \( j \), swap values between them \( (h_i^{(k)}, w_i^{(k)}) \leftrightarrow (h_j^{(k)}, w_j^{(k)}) \) with swap acceptance probability
\[
\alpha_s = \min \left\{ \frac{p(d|h_i^{(k)}, w_j^{(k)})}{p(d|h_j^{(k)}, w_i^{(k)})}, 1 \right\}.
\]

7. \( k \to k + 1 \) and restart step 2.

The tuning parameters of the pCN-PT algorithm are the number of chains, the temperature ladder and the two jumping factor ladders (the latter for both \( h \) and \( w \), separately). Details about dealing with these tuning parameters are shown in Section 3.2.

As intuitive illustration, step 2 uses a change in covariance to morph the given field \( \theta \), step 3 tests the morphed field, step 4 tries out a new field with the same covariance, and step 5 tests the new field. The overall chain stores a new realization after an attempted morph and innovation. This also provides an option to extend other MCMC algorithms for the extended Bayesian inversion.

3. Application

3.1. Model Setup

We consider fully saturated, steady-state groundwater flow as test case, which is an extension of the test case of Xu et al. (2020). The flow equation can be written as
\[
\nabla \cdot (K \nabla H) + S = 0,
\]
where \( \nabla \cdot \) denotes the divergence operator, \( K \) is the hydraulic conductivity (m/day), \( \nabla \) is the Nabla operator, \( H \) denotes hydraulic head (m), and \( S \) denotes the source/sink term as volumetric injection flow rate per unit volume of aquifer (1/day). The flow equation is solved with the groundwater flow simulator MODFLOW (McDonald & Harbaugh, 1988).

We consider a synthetic confined, two-dimensional aquifer in a 5,000 m \( \times \) 5,000 m domain with 50 m thickness. The domain is discretized into 100 \( \times \) 100 \( \times \) 1 cells with cell sizes of 50 m \( \times \) 50 m \( \times \) 50 m. The west and east boundaries are specified head boundaries with heads fixed to 20 and 0 m, respectively. The north and the south
boundaries are impermeable. There are four pumping wells in the domain. The corresponding locations and pumping rates are listed in Table 1. A reference ln-conductivity field is generated based on a multi-Gaussian random field with the parameters listed in Table 2 (Xu et al., 2020). The reference ln-conductivity field and hydraulic head solution (obtained by MODFLOW) are shown in Figure 1. We will consider the ln-conductivity as the model parameter $E \theta$, which has a prior multivariate Gaussian distribution (conditionally on $E h$) with exponential covariance function. The following hyperparameters $E h$ will also be considered as uncertain: (a) mean $\mu$; (b) standard deviation $\sigma$; (c) correlation length $\lambda_1$; (d) correlation length $\lambda_2$. The angle of anisotropy is fixed as in Table 2. This extends the test cases already used by Xu et al. (2020) toward uncertain hyperparameters, while using the same reference field.

3.2. Algorithmic Settings and Adaptivity

Previous studies have shown that the optimal acceptance rate for MCMC-based methods is close to 23.4% and that acceptance rates between 10% and 40% still perform close to optimal (Gelman et al., 1996; Roberts & Rosenthal, 2001; Roberts et al., 1997). The parameter $\beta$ in the pCN proposal affects the acceptance rate of each chain. Basically, high (low) values of $\beta$ will lead to low (high) acceptance rate. Therefore, to obtain an acceptance rate close to the optimal one, we will use an adaptive way to adjust the values of $\beta$ during the “burn-in” period. First, we set an initial value for each $\beta$. Then, we estimate the acceptance rate by running the Markov chain for $N_a$ steps without changing $\beta$. If the acceptance rate is too high (low), we will try to increase (decrease) the value of $\beta$ a little. Thus, we can set an initial value, for example, 0.5, for all jumping factors. Then, through repeating this procedure several times, we can get an acceptance rate close to the optimal one. Here, the $\beta$ values for hyperparameters and white noise are adjusted based on their own acceptance rates separately. To adjust $\beta$ values adaptively, we set $N_a = 10^3$. The first $10^3$ steps in each Markov chain are used for adjusting, which means the adjusting procedure above is repeated 100 times.

In addition, the acceptance swap rate is also an important feature for parallel tempering. The optimal acceptance swap rate differs with respect to the specific applications (Laloy et al., 2016; Predescu et al., 2005; Rathore et al., 2005). In this study, we will control the acceptance swap rate between 10% and 30% just as in the previous work (Xu et al., 2020). The temperature ladder will affect the swap acceptance rate. A smaller (larger) distance between the neighboring temperatures will lead to a higher (lower) swap acceptance rate. Therefore, we will also adjust the temperature ladder adaptively similar to that for adjusting $\beta$ during the “burn-in” period. First, we set an initial temperature ladder and run the Markov chain for $N_a$ steps. Then we can estimate the current acceptance rate. If it is too high (low), we will try to increase (decrease) the distance between the neighboring temperatures a little. After repeating this procedure several times, we can obtain an acceptance swap rate between 10% and 30%. Similarly, we set $N_a = 10^3$ and use the first $10^3$ steps in each Markov chain for adjusting.
For each test case presented in the upcoming section, we will run 20 parallel chains. Each chain runs independently on a dedicated computing node, and data commutation only occurs for between-chain swaps.

For each pair of hyperparameters $h$ and white noise $w$, we need to combine them to get the original model parameters (Equation 17), and then we can calculate the model response and likelihood. A traditional way to achieve this purpose is the Cholesky decomposition of the covariance matrix (Albert, 1987; Davis, 1987), which is simple and can handle any grid structure. However, it is restricted to moderate numbers of discrete cells for the random field. For problems with finer resolution (100 × 100 in this work), the Cholesky decomposition will be very slow. For a case with $n$ discrete cells, the best algorithm for Cholesky decomposition has a complexity of $O(n^3)$ (Dietrich & Newsam, 1997).

In this study, we will use the circulant embedding approach (Dietrich & Newsam, 1993, 1997) as implemented in Fritz et al. (2009). Although this method is only suitable for regularly meshed grids, it is more computationally efficient compared to the Cholesky decomposition. The computational complexity of the circulant embedding approach is only $O(n \log n)$. An example of the computational cost of these two approaches is shown in Table 3, which is performed with MATLAB 2018a in a computer with Intel(R) Core(TM) i7-7700 CPU @ 3.60 GHz and 32 GB RAM. We can see that the circulant embedding approach is more than 30 times faster than the Cholesky decomposition approach in our setup. For larger fields, this advantage will grow due to the almost-linear complexity of the FFT. A recent study (Nowak & Litvinenko, 2013) has even upgraded FFT-based geostatistics to safely handle up to 1 million voxels per direction in 3D space (i.e., $10^{10}$ cells).

Options for irregular grids include, for example, the Karhunen-Loeve decomposition (which relies on a numerical eigendecomposition of the auto-covariance matrix without the circulant embedding and hence without FFT assistance). While this sounds much slower, the Karhunen-Loeve decomposition is usually truncated, such as in the dimension-reduced approach by Zhao and Luo (2021b), and it could be truncated early if desired. Yet another option is to re-parameterize, on a coarser grid, the random field through pilot points (Doherty et al., 2010), followed by interpolation or conditional simulation in between (Keller et al., 2021).

### 3.3. Test Cases

In this section, we will consider four different test cases. For all test cases, we use the following likelihood

$$p(d|\theta(h,w)) = \frac{\exp\left(-\frac{1}{2}(d - M(\theta(h,w)))^T \Sigma^{-1}_e (d - M(\theta(h,w))) \right)}{\sqrt{(2\pi)^d \det(\Sigma_e)}}. \tag{38}$$

where $\Sigma_e$ is the covariance matrix of the error term $e$, and $n_d$ is the number of measurements. In this study, the measurement standard deviation is set to $\sigma_e = 0.05$, that is, $\Sigma_e$ is a diagonal matrix with constant diagonal elements $\sigma_e^2$. All test cases use the exponential model for the covariance function with hyperparameters $\mu$ (for the mean), $\sigma^2$ (for the variance), and $\lambda_1, \lambda_2$ (for the length scales). Further comments on the problem of covariance-model choice are provided in out last section.

In test cases 1, 2, and 3, we will directly use ln-conductivity data, which will lead to a high-dimensional linear estimation problem. For this situation, we can analytically get the marginal likelihood function $p(d|h)$ for the direct inference of hyperparameters. Thus, we can easily get posterior samples of the hyperparameters with MCMC at first. Then, we can use kriging to interpolate the ln-conductivity field for each given posterior sample of hyperparameters and finally obtain the average result quasi-analytically. These results can be used for very precise comparison. The details can be found in Appendix A. For test Case 4, we will use hydraulic head data, which will lead to a high-dimensional nonlinear inverse estimation problem closer to reality.

The data is taken from the reference ln-conductivity field and hydraulic head solution in Figure 1 plus an artificial error (random value from a normal

---

| Well number | Position x (m) | Position y (m) | Pumping rate (m³/day) |
|-------------|----------------|----------------|-----------------------|
| #1          | 500            | 2,350          | 120                   |
| #2          | 3,500          | 2,350          | 70                    |
| #3          | 2,000          | 3,550          | 90                    |
| #4          | 2,000          | 1,050          | 90                    |

---

**Table 2**

| Variogram type | $\mu$ (m/day) | $\sigma$ (m/day) | $\lambda_1$ (m) | $\lambda_2$ (m) | Angle (deg) |
|----------------|---------------|-----------------|-----------------|-----------------|-------------|
| Exponential    | −2.5          | 2               | 2,000           | 1,500           | 135         |

*Note. $\mu$ is the mean, $\sigma$ is the standard deviation, $\lambda_1$ and $\lambda_2$ are the correlation lengths in the x and y directions.*
distribution with mean zero and standard deviation 0.05) at the predefined locations. A summary of the settings for each test case can be found in Table 4.

3.3.1. Test Case 1

In the first test case, we consider 25 direct measurements of ln-conductivity (no head data) at the positions marked with red circles in Figure 2. Among these 25 measurements, 16 measurements are uniformly distributed and the remaining nine measurements are randomly distributed. For the hyperparameters, we assume the following prior distributions

\[
\begin{align*}
\mu & \sim U(-10,10), \\
\sigma & \sim U(0,20), \\
\lambda_1 & \sim U(0,5000), \\
\lambda_2 & \sim U(0,5000).
\end{align*}
\]

A constraint for the anisotropic ratio between the two correlation lengths \(\lambda_1\) and \(\lambda_2\) is also considered as \(0.1 < \lambda_1 / \lambda_2 < 10\). In this test case, we use uniform distributions for the hyperparameters over a broad interval to assume that we only have weak prior information about the hyperparameters. This tests the exploration capabilities of our proposed algorithm, while allowing to compare to a quasi-analytical solution.

3.3.2. Test Case 2

In the second test case, we still consider the same measurements of ln-conductivity as in test Case 1 so that we can again compare to quasi-analytical results. But for the hyperparameters, we now assume the following prior distributions

\[
\begin{align*}
\mu & \sim N(-4,1), \\
\log_2 \sigma & \sim N(2,0.5), \\
\log_{10} \lambda_1 & \sim N(3,0.5), \\
\log_{10} \lambda_2 & \sim N(3,0.5).
\end{align*}
\]

The same constraint \(0.1 < \lambda_1 / \lambda_2 < 10\) for the anisotropic ratio is adopted. In this test case, we use (log)normal distributions for the hyperparameters to assume that we have more prior information about the hyperparameters. This allows testing the exploitation capability of our proposed algorithm in a much more restricted prior setting.

3.3.3. Test Case 3

In the third test case, we consider 500 direct measurements of ln-conductivity at the positions marked with red circles in Figure 3. Among these 500 measurements, 400 measurements are uniformly distributed and the remaining 100 measurements are randomly distributed. The prior distributions of the hyperparameters are the same as those in test Case 2 with the same constraint for the anisotropic ratio. The large number of measurement data enhances test Case 2 by enforcing a highly restrictive likelihood function. This is the last test case against quasi-analytical results.

### Table 3

| Method                  | Cholesky decomposition | Circulant embedding |
|-------------------------|------------------------|---------------------|
| Time (s)                | 4.542                  | 0.128               |

**Table 3**

*Computational Time to Generate One Realization of a Gaussian Random Field (Zero Mean, Unit Variance, Exponential Covariance Function) on a Domain Discretized Into 100 \times 100 Cells*

### Table 4

| Test case | Data type    | Iterations | Iterations for parameter tuning | Methods                  |
|-----------|--------------|------------|---------------------------------|--------------------------|
| T1        | 25 ln-conductivity | 1,200,000  | 100,000                         | pCN-PT & AML-Kriging     |
| T2        | 25 ln-conductivity | 1,200,000  | 100,000                         | pCN-PT & AML-Kriging     |
| T3        | 500 ln-conductivity | 1,600,000  | 100,000                         | pCN-PT & AML-Kriging     |
| T4        | 25 head      | 1,200,000  | 100,000                         | pCN-PT                  |

**Table 4**

*Settings for Each Test Case*

Abbreviations: AML, analytical marginal likelihood; pCN-PT, preconditioned Crank-Nicolson-parallel tempering.
3.3.4. Test Case 4

In the fourth test case, we consider 25 head measurements as used in Xu et al. (2020). The positions (red circles) of these measurements are shown in Figure 4. The prior distributions of the hyperparameters are still (almost) the same as those in test Case 2 with the same constraint for the anisotropic ratio. The only difference is that we truncate the log-normal distributions of the correlation lengths \( \lambda_1 \) and \( \lambda_2 \) at \( \lambda_1 = 4500 \) m and \( \lambda_2 = 4500 \) m. The reason is that the log-normal distribution has no upper bound, head data are very weakly informative for scales due to their diffuse character, and numerical problems could arise within the FFT-based algorithm (Fritz et al., 2009) at correlation length values close to (or above) the domain size. As no quasi-analytical reference is available for test Case 4, we run two independent repetitions and assess convergence by comparison between these repetitions based on the potential scale reduction factor \( R \).

3.4. Testing Criteria

We compare the posterior samples or posterior mean to the synthetic reference. For the test cases 1, 2 and 3, the posterior mean and standard deviation of the ln-conductivity field will also be compared to the quasi-analytical kriging results, and the posterior distributions of hyperparameters will also be compared to the results obtained by MCMC with analytical marginal likelihood. We will also evaluate the difference between the posterior samples of the ln-conductivity field and the synthetic reference field through the root mean square error (RMSE)

\[
RMSE^{(i)} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (\theta_j^{(i)} - \theta_j^{ref})^2},
\]

where \( N \) is the number of cells in the entire domain, \( \theta_j^{(i)} \) is the value of posterior estimate of ln-conductivity at the \( j \)th cell in the \( i \)th iteration, \( \theta_j^{ref} \) is the value of ln-conductivity at the same grid cell for the reference field. Second, we will use ln-likelihood to measure the closeness of the posterior samples to the synthetic data, and we will just take \( \ln(p(d | \theta(h,w))) \) based on Equation 38.

Third, we will evaluate the convergence of the Markov chain by the potential scale reduction factor \( R \) introduced by Gelman and Rubin (1992). A \( R \) value less than 1.2 is considered as an acceptable convergence (Brooks & Gelman, 1998). The factor \( R \) is defined for a set of \( m \) Markov chains, each of which has \( n \) samples. The within-chain variance (for each pixel separately) is estimated as

\[
W = \frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{i=1}^{n} (\theta_j^{(i)} - \bar{\theta}_j)^2,
\]

where \( \theta_j^{(i)} \) is the \( i \)th sample of the \( j \)th chain and \( \bar{\theta}_j \) is the mean of the samples in the \( j \)th chain (for each pixel separately). The pixel-wise between-chain variance is estimated as

\[
B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{\theta}_j - \frac{1}{m} \sum_{j=1}^{m} \bar{\theta}_j)^2.
\]

Then, the estimated variance \( V \) is a weighted average of the within-chain variance and between-chain variance:

\[
V = \left( 1 - \frac{1}{n} \right) W + \frac{1}{n} B.
\]
Finally, the pixel-wise potential scale reduction factor $R$ is defined as

$$ R = \sqrt{\frac{V}{W}}. $$

We compute $R$ for ln-conductivity in each model cell and for all hyperparameters. For ln-conductivity, hence, we can show a map of $R$, and summarize the map by its mean and maximum.

### 3.5. Obtaining the (Reference) Solutions

For test cases 1, 2, and 3, when using the analytical marginal likelihood (AML) for direct inference of hyperparameters as reference solution, we only have a 4-dimensional small problem. Therefore, we generate 1 million posterior samples of hyperparameters by pCN-MCMC (running only on the Gaussian-transformed hyperparameters $h$) and assume these samples are enough to obtain an accurate estimation of the posterior distribution as reference solution. These posterior samples are saved every 10th iteration and the first half of them are discarded as a “burn-in” period. The second half of these samples are used to obtain the average (conditional) kriging results of the ln-conductivity field (denoted as AML-Kriging) and to plot the posterior distribution of hyperparameters (denoted as pCN-AML).

For the proposed pCN-PT method, 1.2 million posterior samples are generated for test cases 1, 2, and 4, 1.6 million posterior samples are generated for test Case 3. Similarly, these samples are saved every 10th iteration and the first half are discarded as a “burn-in” period. The second half of these samples are used to calculate the mean and standard deviation of the ln-conductivity field, plot the posterior distribution of hyperparameters, and estimate the potential scale reduction factor $R$.

### 4. Results

#### 4.1. Test Case 1 (T1)

Figure 5 shows the posterior mean and standard deviation of the ln-conductivity field for test Case 1 obtained by pCN-PT (top row) and AML-Kriging (bottom row). We can see that both results look very similar and the mean of the results captures the main features of the reference field. Additionally, the standard deviation around the measurement points is close to zero and the area with small standard deviation obtained by pCN-PT is only a little larger than that obtained by kriging. To quantitatively measure the closeness between these two results, we calculate the RMSE between them, see Table 5. The RMSE between these two results is 0.195 and 0.135 for mean and standard deviation of the ln-conductivity field. We also calculate the root mean square of the mean and standard deviation of the ln-conductivity field obtained by AML-Kriging, which is 2.607 and 2.185. Thus, we can see that the RMSE between the two results is small compared to the root mean square of the reference solution. These results show that pCN-PT is capable of estimating the hyperparameters and (conditionally) multi-Gaussian ln-conductivity field in a high-dimensional linear problem with limited prior information on hyperparameters.

Figure 6 shows the marginal posterior and prior distributions of the hyperparameters for test Case 1 obtained by pCN-PT and pCN-AML. We can see that the posterior distributions of hyperparameters $\mu$ and $\sigma$ obtained by pCN-PT are very close to the reference distributions obtained by pCN-AML. For hyperparameters $\lambda_1$ and $\lambda_2$, although there are still some differences between the posterior distributions obtained by pCN-PT and pCN-AML, their shapes are similar, that is, the posterior PDF increases significantly for small values of $\lambda_1$ ($\lambda_2$) and decreases gradually until the upper bound of $\lambda_1$ ($\lambda_2$). In addition, we also use the Kolmogorov-Smirnov statistic (maximum difference between the cumulative distribution functions in [0,1]) to quantitatively compare the marginal posterior or distributions of hyperparameters obtained by pCN-PT and pCN-AML, and the results are shown in Table 6. It also shows small differences between these two results. Based on these results, we can see that pCN-PT can obtain a reliable estimate of the posterior distribution of the hyperparameters in a high-dimensional linear problem when there is only little prior information.
Figure 7 shows the ln-likelihood, RMSE, and potential scale reduction factors $R$ of the ln-conductivity realizations and hyperparameters for test Case 1 obtained by pCN-PT along the MCMC chain. According to the results of ln-likelihood and RMSE, we can see that the ‘burn-in’ and convergence to the reference is very fast. Observing the potential scale reduction factors $R$, we can see that these factor values decrease as the length of the Markov chain increases and finally drop below the recommended value (Brooks & Gelman, 1998) of 1.2. We can also see that all the final pixel-wise potential scale reduction factor values on the map are close to 1. This shows a good convergence of the results obtained by pCN-PT.

### 4.2. Test Case 2 (T2)

Figure 8 shows the posterior mean and standard deviation of the ln-conductivity field for test Case 2 obtained by pCN-PT (top row) and AML-Kriging (bottom row). We can see that both solutions have a very similar appearance. The RMSE between the two results is 0.163 and 0.122, while the root mean squares of the results obtained by AML-Kriging (reference solution) are 2.605 and 2.315. Again, this indicates small RMSE and shows the similarity.
between these two results. Now, we see that pCN-PT is also able to estimate the multi-Gaussian ln-conductivity field in a high-dimensional linear problem with more prior information about the hyperparameters. In addition, these results are also very close to the results of test Case 1. This indicates that, although different prior knowledge of hyperparameters is assumed in test Case 1 and 2, we can still get similar results for the ln-conductivity field based on the same data set.

The marginal posterior and prior distributions of hyperparameters for test Case 2 obtained by pCN-PT and pCN-AML are shown in Figure 9. It confirms that pCN-PT provides similar results compared to the reference results obtained by pCN-AML also in this case. The Kolmogorov-Smirnov statistic in Table 6 again confirms the similarity between these two results. This indicates that pCN-PT can also estimate reliably the posterior distribution of the hyperparameters and conductivity field in problems with more prior information.

Figure 10 shows the ln-likelihood, RMSE, and pixel-wise potential scale reduction factors $\hat{R}$ of the ln-conductivity field and the hyperparameters for test Case 2 obtained by pCN-PT. We can see that the Markov chain converges.
very fast based on the results of ln-likelihood and RMSE. The potential scale reduction factor values also decrease as the length of Markov chain increases. We can see that all the final potential scale reduction factor values are very close to 1, which indicates a good convergence of the results obtained by pCN-PT also for test Case 2 with a more constraining prior.

4.3. Test Case 3 (T3)

Figure 11 shows the posterior mean and standard deviation of the ln-conductivity field for test Case 3 obtained by pCN-PT (top row) and AML-Kriging (bottom row). Compared to test cases 1 and 2, more (500) measurements...
are used in test Case 3. First, we can see that both results are very close to each other. The RMSE in Table 5 also confirms the similarity between these two results (the root mean square of the results obtained by kriging is 2.7757 and 0.8657). This indicates that pCN-PT still has a good performance with a large number of data (highly restrictive likelihood function). Second, we can see that the posterior mean is also very close to the reference ln-conductivity field. This follows the plain expectation that a fully Bayesian inversion executed via pCN-PT can recover the multi-Gaussian ln-conductivity field with a large number of (direct) measurements.

Figure 12 shows the marginal posterior and prior distributions of hyperparameters for test Case 3 obtained by pCN-PT and pCN-AML. We find again a large overlap between the posterior distributions obtained by pCN-PT and pCN-AML. The Kolmogorov-Smirnov statistic between these two results is also shown in Table 6. Compared to test cases 1 and 2, the Kolmogorov-Smirnov statistic is a little larger for test Case 3. One reason is that parallel tempering does not provide significant help in test Case 3. In this test case, the contribution of the likelihood function is much larger than for test cases 1 and 2 (see Figures 7, 10 and 13). To ensure enough swap between the neighboring chains, the distance between the neighboring temperatures has to be chosen very small. Therefore, with 20 chains, the hottest chain still does not have a very flat likelihood and cannot explore the whole parameter space very easily. Thus, more chains are needed to let the parallel tempering provide more help, which will also
need more computational resources. However, the current results of pCN-PT still look good, especially for the peak values of the marginal posterior distributions.

Figure 13 shows the ln-likelihood, RMSE, and pixel-wise potential scale reduction factors \( R \) of the ln-conductivity field and the hyperparameters for test Case 3 obtained by pCN-PT. We can see that the ln-likelihood and RMSE become stable very fast, which shows a short “burn-in” period and fast convergence of the Markov chain. The overall potential scale reduction factor \( R \) of the ln-conductivity field keeps decreasing as the length of the Markov chain increases and finally becomes less than 1.2 in most parts of the domain. The potential scale reduction factor \( R \) of hyperparameters stays around 1.2 and is finally close to 1. This also shows a good convergence of the results obtained by pCN-PT.

### 4.4. Test Case 4 (T4)

For test Case 4 with hydraulic head data, we only have the results obtained by pCN-PT. Figure 14 shows the posterior mean and standard deviation of the ln-conductivity field. Compared to the reference ln-conductivity field, we can see that the mean of the results captures the main features in the high conductivity areas. The marginal posterior and prior distributions of the hyperparameters for test Case 4 are shown in Figure 15. We can find that all the posterior distributions have similar shapes compared to the prior distributions and have a slight shift to the nominal values of the hyperparameters, which can also be seen from Table 7. This means the data helps to calibrate the hyperparameters a little, although head values are less informative for hyperparameters compared...
Figure 10. Testing criteria for test Case 2: Ln-likelihood of the posterior samples of the ln-conductivity field, root mean square error between the posterior samples of the ln-conductivity field and the synthetic reference field, the evolution of mean and max potential scale reduction factor of the ln-conductivity field, the evolution of the potential scale reduction factor of the hyperparameters, and the final potential scale reduction factor of the ln-conductivity field obtained by preconditioned Crank-Nicolson-parallel tempering. The dashed line in the figure of ln-likelihood corresponds to the ln-likelihood of the reference. The dashed lines in the evolution of the potential scale reduction factor denote the value of 1.2.

Figure 16 shows the testing criteria for test Case 4. The ln-likelihood once again shows a fast convergence of the Markov chain. Although the RMSE is a little larger for some realizations, most realizations in the second half still have a small RMSE. For the ln-conductivity field, the potential scale reduction factor \( R \) decreases as the length of Markov chain increases. Finally, most parts of the ln-conductivity field have a pixel-wise \( R \) value less than 1.2. The potential scale reduction factor \( R \) of the hyperparameters also decreases and finally is less than 1.2. Overall, pCN-PT still gets a good result for this test case with a high-dimensional nonlinear problem.
5. Conclusion, Final Discussion, and Outlook

In this work, we extend the highly efficient pCN-PT algorithm for geostatistical inversion and estimation of unknown spatially variable hydraulic conductivity fields to the extended Bayesian inversion problem with the estimation of uncertain hyperparameters of multi-Gaussian fields besides the hydraulic conductivities. This extended Bayesian inversion poses a harder, wider and more realistic problem since the values of hyperparameters are not fixed a priori, but one performs a formal joint Bayesian inference of hyperparameters together with the geostatistical field. To keep the high efficiency of the pCN-PT algorithm, we first reconstruct the original problem by decomposing the original model parameters (ln-conductivity) into hyperparameters (mean, standard deviation, correlation lengths) and white noise (a standard normal random vector). Then, we perform the Bayesian inversion with pCN-PT by considering hyperparameters and white noise as the primary inversion parameters. With this approach, the acceptance probability is still only dependent on the likelihood when using pCN-PT. Finally, the posterior samples of original model parameters are recovered by combining the posterior samples of hyperparameters and white noise.

In this extended pCN-PT algorithm, we update hyperparameters and white noise separately during the MCMC process since they can be considered as independent to each other. This leads to the classical Gibbs sampler and allows a better fine-tuning of the corresponding MCMC parameters. There are mainly two kinds of tuning
parameters, that is, a jumping factor $\beta$ in the pCN proposal and temperatures $T$ for the parallel chains. We adjust the jumping factor $\beta$ adaptively based on the recent acceptance rate in each chain. The temperature $T$ is adjusted based on the recent swap acceptance rate between neighboring chains. This adaptive parameter tuning can help us easily find proper values of the MCMC parameters.

Based on the results of the test cases, we see that the extended pCN-PT algorithm is applicable for the extended Bayesian inversion with uncertain hyperparameters of multi-Gaussian random fields in both high-dimensional linear problems and high-dimensional nonlinear problems. We also find that it is difficult to constrain hyperparameters, especially with a relative small set of hydraulic head data. The posterior distribution of the hyperparameters is a bit narrower than the prior distribution for such a case, and approaches the reference value, but considerable uncertainty is left. This points to the importance of direct data (hydraulic conductivity) and other data sources which give information on spatial structures of hydraulic conductivity, like data from geological maps and geophysical surveys. The use of these data sources in this inversion framework requires further research.

There are many other data types that could be included in the inverse problem. In this study, we just use head and conductivity data for demonstration on a well-known benchmark case. Whether other data types (e.g., hydraulic tomography data, borehole dilution or tracer data) would make hyperparameters almost certain is another
This question can be answered only after one has an accurate algorithm, as the one proposed in the current study. In fact, since our test case is not too large to be ergodic, some degree of uncertainty will always remain in the hyperparameters. However, we should still do our best to let the data speak about the structural assumptions of geostatistical models, and have the algorithms at hand to see the remaining structural uncertainty while doing the inversion.

Our test cases all used the exponential covariance function, and inferred its hyperparameters (i.e., mean, variance, and scale). However, smoothness and distribution-across-scales is also an important issue, which means the type

Figure 13. Testing criteria for test Case 3: Ln-likelihood of the posterior samples of the ln-conductivity field, root mean square error between the posterior samples of the ln-conductivity field and the synthetic reference field, the evolution of mean and max potential scale reduction factor of the ln-conductivity field, the evolution of the potential scale reduction factor of the hyperparameters, and the final potential scale reduction factor of the ln-conductivity field obtained by preconditioned Crank-Nicolson-parallel tempering. The dashed line in the figure of ln-likelihood corresponds to the ln-likelihood of the reference. The dashed lines in the evolution of the potential scale reduction factor denote the value of 1.2.
Figure 14. Posterior mean and standard deviation of the ln-conductivity field for test Case 4 obtained by preconditioned Crank-Nicolson-parallel tempering.

Figure 15. Marginal posterior and prior distributions of the hyperparameters for test Case 4 obtained by preconditioned Crank-Nicolson-parallel tempering. The vertical dot-dashed lines denote the nominal values of the hyperparameters.
Figure 16. Testing criteria for test Case 4: Ln-likelihood of the posterior samples of the ln-conductivity field, root mean square error between the posterior samples of the ln-conductivity field and the synthetic reference field, the evolution of mean and max potential scale reduction factor of the ln-conductivity field, the evolution of the potential scale reduction factor of the hyperparameters, and the final potential scale reduction factor of the ln-conductivity field obtained by preconditioned Crank-Nicolson-parallel tempering. The dashed line in the figure of ln-likelihood corresponds to the ln-likelihood of the reference. The dashed lines in the evolution of the potential scale reduction factor denote the value of 1.2.
of covariance function is known to be equally important. This could be included either as a Bayesian model selection problem or by using the Matérn covariance function (Handcock & Stein, 1993). The latter includes an additional shape parameter that controls the differentiability and scale-distribution of the covariance function. In fact, the exponential, Gaussian and power-law variograms are included as special cases and/or limit cases in the Matérn model. Therefore, it was successfully applied to parameterize the choice across covariance models (Leube et al., 2012; Nowak et al., 2010).

Such extensions toward more complicated (nonlinear) data types and additional structural uncertainty will make the problem harder to solve. In the current work, we used 20 chains, and more chains may become necessary to optimally explore and exploit the solution space. Additionally, larger domains and transient 3D problems will consume more computing time. Then, there will be the ubiquitous trade-off between computational resources and computational accuracy.

The core contribution of our work, that is, the decomposition of the problem into white noise at the basis of our pCN-MCMC extension, can be used to extend other geostatistical MCMC algorithms for uncertain hyperparameters. For example, one could extend the blocking MCMC approach as used in Fu and Gómez-Hernández (2009). The idea would be to repeat a random simulation (via sequential simulation algorithms) with the same random seed decisions (white noise and sequential simulation path), but with a new covariance to do the morphing step. Then, one can use a conventional blocking MCMC step (which also uses sequential simulation) to achieve a change in the parameter field.

Appendix A

The posterior distribution \( p(h|d) \) of hyperparameters \( h \) can be obtained through marginalization based on the joint posterior distribution \( p(\theta, h|d) \) obtained through Equation 3. However, it can also be obtained directly through Bayes’ rule as (Malinverno & Briggs, 2004)

\[
p(h|d) \propto p(h)p(d|h),
\]

where the marginal likelihood function \( p(d|h) \) can be calculated by

\[
p(d|h) = \int_\Omega p(d|\theta, h)p(\theta|h)d\theta,
\]

where \( p(d|\theta, h) \) is the likelihood function in Equation 3.

For the test cases with ln-conductivity data only, the forward problem transforms into a simple linear problem according

\[
d = \theta + e.
\]

For this simple linear problem, the likelihood function \( p(d|\theta, h) \) can be written as

\[
p(d|\theta, h) = \frac{\exp\left(-\frac{1}{2}(d - \theta)^\top \Sigma_e^{-1}(d - \theta)\right)}{\sqrt{(2\pi)^d \det(\Sigma_e)}}.
\]

Here, \( \theta \) denotes the model parameters at the measurement locations, that is, the dimension of \( \theta \) is equal to the number of data points.

Since we assume a priori a conditionally multivariate Gaussian distribution for model parameters \( \theta \), \( p(\theta|h) \) can be written as

\[
p(\theta \mid h) = \frac{\exp\left(-\frac{1}{2}(\theta - \mu)^\top \Sigma^{-1}(\theta - \mu)\right)}{\sqrt{(2\pi)^d \det(\Sigma)}}.
\]
Here, \( \mu \) and \( \Sigma \) are the mean vector and covariance matrix of model parameters \( \theta \) at the measurement locations, with \( \mu = \mu(h) \) and \( \Sigma = \Sigma(h) \).

Utilizing the properties of Gaussian distributions, the marginal likelihood function \( p(d|h) \) in Equation A2 can be calculated analytically as (Tarantola, 2005, Section 6.21)

\[
p(d|h) = \frac{\exp\left(-\frac{1}{2}(d-\mu)^T(\Sigma + \Sigma_e)^{-1}(d-\mu)\right)}{\sqrt{(2\pi)^{nd} \det(\Sigma + \Sigma_e)}}.
\]

(A6)

In this study, we have four hyperparameters, that is, mean \( \mu \), standard deviation \( \sigma \) and two correlation lengths \( \lambda_1 \) and \( \lambda_2 \). In Equation A6, the mean vector \( \mu \) is constructed based on the hyperparameter mean \( \mu \), and the covariance matrix \( \Sigma \) is constructed based on the hyperparameters standard deviation \( \sigma \) and correlation lengths \( \lambda_1, \lambda_2 \). Then, we can easily sample the posterior distribution \( p(\theta|d) \) based on Equations A1 and A6 through MCMC (e.g., Metropolis-Hastings algorithm or pCN-MCMC).

For a given posterior sample of hyperparameters, we can use kriging to obtain a best linear unbiased estimator \( m_K \) of ln-conductivity over the entire domain and obtain the corresponding variance \( \nu_K \) of the estimator. Then, for all posterior samples of hyperparameters, we can get a set of kriging estimators \( m_K^{(i)} \) (\( i = 1, \ldots, n \), where \( n \) is the number of posterior samples of hyperparameters) and the corresponding variances \( \nu_K^{(i)} \) (\( i = 1, \ldots, n \)). The final statistical mean of the ln-conductivity field based on all realizations of hyperparameters is estimated by the average of all kriging estimators as

\[
m_K^f = \frac{1}{n} \sum_{i=1}^{n} m_K^{(i)}.\]

(A7)

The final statistical variance of the ln-conductivity field across all realizations of hyperparameters is estimated based on the law of total variance as

\[
\nu_K^f = \frac{1}{n} \sum_{i=1}^{n} \nu_K^{(i)} + \frac{1}{n-1} \sum_{i=1}^{n} (m_K^{(i)} - m_K^f)^2.
\]

(A8)

In this study, kriging is implemented with the FFT-based algorithm in Fritz et al. (2009).

**Data Availability Statement**

The code and related date are available from this site [https://data.mendeley.com/datasets/mtk879vsst/draft?a=a9a048b5-f7de-4a37-907b-27ecb1159d5f](https://data.mendeley.com/datasets/mtk879vsst/draft?a=a9a048b5-f7de-4a37-907b-27ecb1159d5f).

**References**

Alabert, F. (1987). The practice of fast conditional simulations through the LU decomposition of the covariance matrix. *Mathematical Geology*, 19(5), 369–386. [https://doi.org/10.1007/bf00897191](https://doi.org/10.1007/bf00897191)

Altekar, G., Dwarkadas, S., Hueskenbeck, J. P., & Ronquist, F. (2004). Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*, 20(3), 407–415. [https://doi.org/10.1093/bioinformatics/tbh427](https://doi.org/10.1093/bioinformatics/tbh427)

Besag, J., Green, P., Higdon, D., & Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, 10(1), 3–41. [https://doi.org/10.1214/ss/1177011730](https://doi.org/10.1214/ss/1177011730)

Boggs, J. M., Young, S. C., Beard, L. M., Gelhar, L. W., Rehfeldt, K. R., & Adams, E. E. (1992). Field study of dispersion in a heterogeneous aquifer: I. overview and site description. *Water Resources Research*, 28(12), 3281–3291. [https://doi.org/10.1029/92wr01756](https://doi.org/10.1029/92wr01756)

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational & Graphical Statistics*, 7(4), 434–455. [https://doi.org/10.1080/10618600.1998.10474787](https://doi.org/10.1080/10618600.1998.10474787)

Bui-Thanh, T., Ghattas, O., Martin, J., & Stadler, G. (2013). A computational framework for infinite-dimensional Bayesian inverse problems part I: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, 35(6), A2494–A2523. [https://doi.org/10.1137/12089586x](https://doi.org/10.1137/12089586x)

Buland, A., & Omre, H. (2003). Bayesian linearized AVO inversion. *Geophysics*, 68(1), 185–198. [https://doi.org/10.1190/1.1543206](https://doi.org/10.1190/1.1543206)

Carrera, J., & Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: I. Maximum likelihood method incorporating prior information. *Water Resources Research*, 22(2), 199–210. [https://doi.org/10.1029/WR022i002p00199](https://doi.org/10.1029/WR022i002p00199)

Cary, P. W., & Chapman, C. H. (1988). Automatic 1-D waveform inversion of marine seismic refraction data. *Geophysical Journal*, 93(3), 527–546. [https://doi.org/10.1111/j.1365-246x.1988.tb03879.x](https://doi.org/10.1111/j.1365-246x.1988.tb03879.x)

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3), 167–174. [https://doi.org/10.1080/00031305.1992.10475878](https://doi.org/10.1080/00031305.1992.10475878)
Chen, Y., & Zhang, D. (2006). Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources*, 29(8), 1107–1122. https://doi.org/10.1016/j.adwres.2005.09.007

Cirpka, O. A., & Nowak, W. (2003). Dispersion on kriged hydraulic conductivity fields. *Water Resources Research*, 39(2), https://doi.org/10.1029/2001WR000598

Congdon, P. (2003). *Applied Bayesian modelling*. John Wiley & Sons, Ltd.

Cotter, S. L., Doshi, M., Robinson, J. C., & Stuart, A. M. (2009). Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, 25(11), 115008. https://doi.org/10.1088/0266-5611/25/11/115008

Cotter, S. L., Roberts, G. O., Stuart, A. M., & White, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science*, 28(3), 424–446. https://doi.org/10.1214/13-sts421

Dagan, G. (1988). Time-dependent macroporosity for solute transport in anisotropic heterogeneous aquifers. *Water Resources Research*, 24(9), 1491–1500. https://doi.org/10.1029/WR024i009p01491

Davis, M. W. (1987). Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Mathematical Geology*, 19(2), 91–98. https://doi.org/10.1007/bf00897749

Dentz, M., Kinzelbach, H., Attinger, S., & Kinzelbach, W. (2000). Temporal behavior of a solute cloud in a heterogeneous porous medium: 1. Point-like injection. *Water Resources Research*, 36(12), 3591–3604. https://doi.org/10.1029/2000wr900162

Dietrich, C. R., & Newsam, G. N. (1993). A fast and exact method for multidimensional Gaussian stochastic simulations. *Water Resources Research*, 29(8), 2861–2869. https://doi.org/10.1029/93wr01070

Dietrich, C. R., & Newsam, G. N. (1997). Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(6), 1088–1107. https://doi.org/10.1137/s1064827592240555

Doherty, J. E., Fienen, M. N., & Hunt, R. J. (2010). Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions. *US Geological Survey scientific investigations report*, 5168, 36.

Duijndam, A. J. W. (1988). Bayesian estimation in seismic inversion. Part I: Principles. *Geophysical Prospecting*, 36(8), 878–898. https://doi.org/10.1111/j.1365-2478.1988.tb01985.x

Earl, D. J., & Deem, M. W. (2005). Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics, 7*, 3910–3916. https://doi.org/10.1039/b509983h

Fritz, J., Neuweiler, I., & Nowak, W. (2009). Application of FFT-based algorithms for large-scale universal Kriging problems. *Mathematical Geosciences*, 41(5), 509–533. https://doi.org/10.1007/s11004-009-9220-x

Fu, J., & Gómez-Hernández, J. J. (2009). A blocking Markov chain Monte Carlo method for inverse stochastic hydrogeological modeling. *Mathematical Geosciences*, 41(2), 105–128. https://doi.org/10.1007/s11004-008-9206-0

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Gelman, A., Roberts, G., & Gilks, W. (1996). Efficient metropolis jumping rules. *Bayesian Statistics*, 5, 599–607.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gómez-Hernández, J. J., Sahuquillo, A., & Capilla, J. E. (1997). Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data—I. Theory. *Journal of Hydrology*, 203(1), 162–174.

Grandis, H., Menivelle, M., & Roussignol, M. (1999). Bayesian inversion with Markov chains—1. The magnetotelluric one-dimensional case. *Geophysical Journal International*, 138(3), 757–768. https://doi.org/10.1046/j.1365-246x.1999.00904.x

Hairer, M., Stuart, A. M., & Vollmer, S. I. (2014). Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Annals of Applied Probability*, 24(6), 2455–2490. https://doi.org/10.1214/14-AAP1098

Handcock, M. S., & Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics*, 35(4), 403–410. https://doi.org/10.1080/00401706.1993.10485354

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

Hukushima, K., & Nemoto, K. (1996). Exchange Monte Carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65(6), 1604–1608. https://doi.org/10.1143/jpsj.65.1604

Keller, J., Hendricks Franssen, H.-J., & Nowak, W. (2021). Investigating the pilot point ensemble Kalman filter for geostatistical inversion and data assimilation. *Advances in Water Resources*, 104010. https://doi.org/10.1016/j.adwres.2021.104010

Kitanidis, P. K. (1995). Quasi-linear geostatistical theory for inversion. *Water Resources Research*, 31(10), 2411–2419. https://doi.org/10.1029/94wr01945

Kitanidis, P. K., & Vomvoris, E. G. (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resources Research*, 19(3), 677–690. https://doi.org/10.1029/wr019i003p00677

Laloy, E., Linde, N., Jacques, D., & Mariethoz, G. (2016). Merging parallel tempering with sequential geostatistical resampling for improved posterior exploration of high-dimensional subsurface categorical fields. *Advances in Water Resources*, 90, 57–69. https://doi.org/10.1016/j.adwres.2016.02.008

Leube, P., Geiges, A., & Nowak, W. (2012). Bayesian assessment of the expected data impact on prediction confidence in optimal sampling design. *Water Resources Research*, 48, W02501. https://doi.org/10.1029/2010WR010137

Li, L., Zhou, H., Hendricks Franssen, H.-J., & Gómez-Hernández, J. J. (2012). Groundwater flow inverse modeling in non-Multigaussian media: Performance assessment of the normal-score ensemble Kalman filter. *Hydrology and Earth System Sciences*, 16(2), 573–590. https://doi.org/10.5194/hess-16-573-2012

Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, 151(3), 675–688. https://doi.org/10.1046/j.1365-246x.2002.01847.x

Malinverno, A., & Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, 69(4), 1005–1016. https://doi.org/10.1190/1.1778243

Mariethoz, G., Renard, P., & Caers, J. (2010). Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resources Research*, 46(11), W11530. https://doi.org/10.1029/2010wr009274

McDonald, M. G., & Harbaugh, A. W. (1988). A modular three-dimensional finite-difference ground-water flow model (Tech. Rep.). U.S. Geological Survey. Retrieved from http://pubs.er.usgs.gov/publication/wr9061

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087–1092. https://doi.org/10.1063/1.1699114

Moreno, J. J., Katzgraber, H. G., & Hartmann, A. K. (2003). Finding low-temperature states with parallel tempering, simulated annealing and simple Monte Carlo. *International Journal of Modern Physics C*, 14(03), 285–302. https://doi.org/10.1142/s0129183103004498
Zhao, Y., & Luo, J. (2021b). A quasi-newton reformulated geostatistical approach on reduced dimensions for large-dimensional inverse problems.

Zhao, Y., & Luo, J. (2021a). Bayesian inverse modeling of large-scale spatial fields on iteratively corrected principal components.

Zha, Y., Yeh, T.-C. J., Illman, W. A., Zeng, W., Zhang, Y., Sun, F., & Shi, L. (2018). A reduced-order successive linear estimator for geostatistical

Yeh, T.-C. J., Jin, M., & Hanna, S. (1996). An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields.

Xu, T., Reuschen, S., Nowak, W., & Hendricks Franssen, H.-J. (2020). Preconditioned Crank-Nicolson Markov chain Monte Carlo coupled with parallel tempering: An efficient method for Bayesian inversion of multi-Gaussian log-hydraulic conductivity fields. Water Resources Research, 56(8), e2020WR027110. doi:10.1029/2020wr027110

Yeh, T.-C. J., Jin, M., & Hanna, S. (1996). An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. Water Resources Research, 32(1), 85–92. doi:10.1029/95wr02869

Zha, Y., Yeh, T.-C. J., Illman, W. A., Zeng, W., Zhang, Y., Sun, F., & Shi, L. (2018). A reduced-order successive linear estimator for geostatistical inversion and its application in hydraulic tomography. Water Resources Research, 54(3), 1616–1632. doi:10.1002/2017wr021884

Zhao, Y., & Luo, J. (2021a). Bayesian inverse modeling of large-scale spatial fields on iteratively corrected principal components. Advances in Water Resources, 151, 103913. doi:10.1016/j.adwatres.2021.103913

Zhao, Y., & Luo, J. (2021b). A quasi-newton reformulated geostatistical approach on reduced dimensions for large-dimensional inverse problems. Water Resources Research, 57(1), e2020WR028399. doi:10.1029/2020wr028399

Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. Journal of Geophysical Research, 100(B7), 12431–12447. https://doi.org/10.1029/94jb03097

Nagel, J. B., & Sudret, B. (2016). A unified framework for multilevel uncertainty quantification in Bayesian inverse problems. Probabilistic Engineering Mechanics, 43, 68–84. https://doi.org/10.1016/j.probengmech.2015.09.007

Nowak, W., & Cirpka, O. A. (2006). Geostatistical inference of hydraulic conductivity and dispersivities from hydraulic heads and tracer data. Water Resources Research, 42(8), W08416. https://doi.org/10.1029/2005wr004832

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

Nowak, W., & Litvinenko, A. (2013). Kriging and spatial design accelerated by orders of magnitude: Combining low-rank covariance approximations with FFT-techniques. Mathematical Geosciences, 45(4), 411–435. https://doi.org/10.1007/s11004-013-9453-6

Predescu, C., Predescu, M., & Ciobanu, C. V. (2005). On the efficiency of exchange in parallel tempering Monte Carlo simulations. The Journal of Physical Chemistry B, 109(9), 4189–4196. https://doi.org/10.1021/jp045073+

Rathore, N., Chopra, M., & de Pablo, J. J. (2005). Optimal allocation of replicas in parallel tempering simulations. The Journal of Chemical Physics, 122(2), 024111. https://doi.org/10.1063/1.1831273

Robert, C. P., Elvira, V., Tawn, N., & Wu, C. (2018). Accelerating MCMC algorithms. WIREs Computational Statistics, 10(5), e1435. https://doi.org/10.1002/wics.1435

Roberts, G. O., Gelman, A., & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. Annals of Applied Probability, 7(1), 110–120.

Roberts, G. O., & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. Statistical Science, 16(4), 351–367. https://doi.org/10.1214/ss/1015346320

Schott, J.-J., Roussignol, M., Menville, M., & Nomenjahanary, F. R. (1999). Bayesian inversion with Markov chains—II. The one-dimensional DC multilayer case. Geophysical Journal International, 138(3), 769–783. https://doi.org/10.1046/j.1365-246x.1999.00905.x

Stuart, A. M. (2010). Inverse problems: A Bayesian perspective. Acta Numerica, 19, 451–559. https://doi.org/10.1017/s0962492910000061

Sułkicy, E. A. (1986). A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process. Water Resources Research, 22(33), 2069–2082. https://doi.org/10.1029/wr022i03p02069

Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics.

Tierney, L. (1994). Markov chains for exploring posterior distributions. Annals of Statistics, 22(4), 1701–1728. https://doi.org/10.1214/aoas/1176325755

Trolldborg, M., Nowak, W., Lange, I. V., Santos, M. C., Binning, P. J., & Bjerg, P. L. (2012). Application of Bayesian geostatistics for evaluation of mass discharge uncertainty at contaminated sites. Water Resources Research, 48(9), W09535. https://doi.org/10.1029/2011wr011785
van Ravenzwaaij, D., Cassev, P., & Brown, S. D. (2018). A simple introduction to Markov chain Monte-Carlo sampling. Psychonomic Bulletin & Review, 25(1), 143–154. https://doi.org/10.3758/s13423-016-1015-8

Woodbury, A. D., & Ulych, T. J. (2000). A full-Bayesian approach to the groundwater inverse problem for steady state flow. Water Resources Research, 36(8), 2081–2093. https://doi.org/10.1029/2000wr900806

Xu, T., Reuschen, S., Nowak, W., & Hendricks Franssen, H.-J. (2020). Preconditioned Crank-Nicolson Markov chain Monte Carlo coupled with parallel tempering: An efficient method for Bayesian inversion of multi-Gaussian log-hydraulic conductivity fields. Water Resources Research, 56(8), e2020WR027110. https://doi.org/10.1029/2020wr027110

Ye, T.-C. J., Jin, M., & Hanna, S. (1996). An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. Water Resources Research, 32(1), 85–92. https://doi.org/10.1029/95wr02869

Zha, Y., Yeh, T.-C. J., Illman, W. A., Zeng, W., Zhang, Y., Sun, F., & Shi, L. (2018). A reduced-order successive linear estimator for geostatistical inversion and its application in hydraulic tomography. Water Resources Research, 54(3), 1616–1632. https://doi.org/10.1002/2017wr021884

Zhao, Y., & Luo, J. (2021a). Bayesian inverse modeling of large-scale spatial fields on iteratively corrected principal components. Advances in Water Resources, 151, 103913. https://doi.org/10.1016/j.adwatres.2021.103913

Zhao, Y., & Luo, J. (2021b). A quasi-newton reformulated geostatistical approach on reduced dimensions for large-dimensional inverse problems. Water Resources Research, 57(1), e2020WR028399. https://doi.org/10.1029/2020wr028399