L aboratório
N acional de
C omputação
C ientífica

COMOD
Coordenação de Modelagem Computacional

Relatório de Atividades

Servidor: Marcio Borges

Petrópolis-RJ
January 10, 2024

1 INTRODUTION

Natural reservoirs display a high degree of variability in their hydraulic properties in multiple length scales (Dagan, 1989; Gelhar, 1993). Such variability plays a strong impact on flu-id flow patterns in subsurface formations (Glimm et al., 1992). Due to operational difficulties and high cost, direct measurements of the properties of interest are scarce and, therefore, a deterministic description of the reservoirs cannot be achieved. Alternatively, a stochastic approach should be adopted to characterize uncertainties in reservoir parameters that, according to Efendiev et al. (2006), are the primary factor contributing to reservoir forecasting uncertainties. In other words, the predictability of computational models is severely limited by the lack of an adequate description of reservoir properties.

Data acquisition along with the use of high-performance computing has encouragedThe growing data acquisition along with the use of high-performance computing has encouraged the use of dynamic data directly in simulations to reduce the associated uncertainties and increase the predictability of the models. Well tests, production history, pressure variation on monitors, and tracer tests, among others, are direct measures of reservoir responses and can provide important information about the flow processes.

Matching field dynamic data with reservoir simulation results is a stochastic inverse problem that can be formalized in terms of Bayesian inference and Markov chain Monte Carlo methods (McMC). The Bayesian inference is convenient in quantifying the added value of information from several sources, while McMC methods provide a computational framework for sampling from *a posteriori* distribution (Robert and Casella, 2005; Liu et al., 2008).

Despite its high computational cost that, in some cases, can be prohibitive, McMC methods are regarded as the gold standard technique for Bayesian inference (Nemeth and Fearnhead, 2021). In 2000, Computing in Science & Engineering, a joint publication of the IEEE Computer Society and the American Institute of Physics, placed the Metropolis Algorithm on the list of the 10 algorithms with the greatest impact on the development of science and engineering practice in the 20th century (“ Top Ten Algorithms of the Century” (Madey et al., 2005)).

Within this context McMC methods have been widely used in the last decades in porous media flow problems (Liu and Oliver, 2003; Efendiev et al., 2005, 2006; Ma et al., 2008; Dostert et al., 2009; Das et al., 2010; Mondal et al., 2010; Ginting et al., 2011, 2012; Iglesias et al., 2013; Emerick and Reynolds, 2013). McMC algorithms typically require the design of proposal mechanisms to generate candidate hypotheses. Many existing machine learning algorithms can be adapted to become proposal mechanisms (de Freitas et al., 2013). This is often essential to obtain McMC algorithms that converge quickly.

1.1 Random permeability fields

Due to incomplete knowledge about the rock properties that show variability at multiple length scales, input parameters such as the permeability field, $\boldsymbol{\kappa}(\mathbf{x}, \omega)$, are treated as random space functions with statistics inferred from geostatistical models (here $\mathbf{x} = (x_1, x_2, x_3)^{\mathsf{T}} \in \mathbb{R}^3$

and ω is a random element in the probability space). In line with Dagan (1989) and Gelhar (1993) the permeability field is modeled as a log-normally distributed random space function

$$\boldsymbol{\kappa}(\mathbf{x},\omega) = \beta \exp\Big[\rho \mathsf{Y}(\mathbf{x},\omega)\Big], \tag{1}$$

where $\beta, \rho \in \mathbb{R}^+$ and $\mathsf{Y}(\mathbf{x},\omega) \sim \mathbb{N}(\boldsymbol{\mu}_\mathsf{Y}, \boldsymbol{\mathcal{C}}_\mathsf{Y})$ is a Gaussian random field characterized by its mean $\boldsymbol{\mu}_\mathsf{Y} = \langle \mathsf{Y}(\mathbf{x}) \rangle$ and two-point covariance function

$$\boldsymbol{\mathcal{C}}_\mathsf{Y}(\mathbf{x},\mathbf{y}) = \mathsf{Cov}\left(\mathsf{Y}(\mathbf{x}), \mathsf{Y}(\mathbf{y})\right) = \mathsf{E}\Big[\big(\mathsf{Y}(\mathbf{x}) - \langle \mathsf{Y}(\mathbf{x}) \rangle\big)\big(\mathsf{Y}(\mathbf{y}) - \langle \mathsf{Y}(\mathbf{y}) \rangle\big)\Big]. \tag{2}$$

Moreover, in this work, Y is a second-order stationary process (Gelhar, 1993), that is:

$$\begin{aligned} \langle \mathsf{Y}(\mathbf{x}) \rangle &= \mu_\mathsf{Y}, \quad \text{(constant)} \\ \boldsymbol{\mathcal{C}}_\mathsf{Y}(\mathbf{x},\mathbf{y}) &= \boldsymbol{\mathcal{C}}_\mathsf{Y}\left(\|\mathbf{x}-\mathbf{y}\|\right) = \boldsymbol{\mathcal{C}}_\mathsf{Y}(d)\,. \end{aligned} \tag{3}$$

1.2 Karhunen-Loève expansion

The Gaussian field Y can be represented as a series expansion involving a complete set of deterministic functions with correspondent random coefficients using the Karhunen-Loève (KL) expansion proposed independently by Karhunen (1946) and Loève (1955). It is based on the eigen-decomposition of the covariance function. Depending on how fast the eigenvalues decay one may be able to retain only a small number of terms in a truncated expansion and, consequently, this procedure may reduce the search to a smaller parameter space. In uncertainty quantification methods for porous media flows, the KL expansion has been widely used to reduce the number of parameters used to represent the permeability field (Efendiev et al., 2005, 2006; Das et al., 2010; Mondal et al., 2010; Ginting et al., 2011, 2012). Another advantage of KL expansion lies on the fact that it provides orthogonal deterministic basis functions and uncorrelated random coefficients, allowing for the optimal encapsulation of the information contained in the random process into a set of discrete uncorrelated random variables (Ghanem and Spanos, 1991). This remarkable feature can be used to simplify the Metropolis-Hastings McMC Algorithm in the sense of the search may be performed in the space of discrete uncorrelated random variables ($\boldsymbol{\theta}$), no longer in the space of permeabilities which have a more complex statistical structure.

In summary, we may have the reduction of the stochastic dimension and the sampling in the space of uncorrelated Gaussian variables.

2 Markov chain Monte Carlo method McMC

The Metropolis algorithm (MH) was originally introduced by Metropolis et al. (1953) for computing properties of substances composed of interacting individual molecules (when a symmetric proposal is used). This algorithm has been used extensively in statistical physics. A generalization to non-symmetric proposals was introduced by Hastings (1970).

2.1 Differential Evolution Metropolis (DE)

The Metropolis algorithm or one of the variants, are often used in situations where the *a priori* knowledge of the target distribution is quite limited.

The drawback of McMC is that the acceptance rate is generally very low due to the large dimensionality of the stochastic fields. Thus, McMC simulations usually require thousands of iterations before converging to a steady state. Each iteration involves the computation of the fine-scale flow problem which is very CPU demanding. This makes McMC simulations very expensive from the computational point of view. In order to overcome this difficulty, we have introduced multi-stage, multi-physics McMC methods (Ginting et al., 2011, 2012, 2013). The shape and size of the proposal distribution $\mathsf{q}(\cdot)$ is known to be very crucial for the convergence of the Markov chain corresponding the McMC algorithm (Haario et al., 1999).

Several methods have been developed in order to speed up the convergence: Adaptive Proposal (AP) (Haario et al., 1999); Adaptive Metropolis (AM) (Haario et al., 2001); Delayed Rejection (DR) (Tierney, 1994; Tierney and Mira, 1999); Delayed Rejection Adaptive Metropolis algorithm (DRAM) (Haario et al., 2006); Single Component Adaptive Metropolis (SCAM) (Haario et al., 2005); Differential Evolution Markov Chain (DE) (Vrugt et al., 2003; Ter Braak, 2006b); DiffeRential Evolution Adaptive Metropolis (DREAM) (Vrugt et al., 2008), among others.

2.1 Differential Evolution Metropolis (DE)

In this paper we choose the Differential Evolution Metropolis (DE) to perform our experiments. DE is a population McMC algorithm, in which multiple chains are run in parallel. The jumps are simply a fixed multiple of the differences of two random parameter vectors that are currently in the population.

Vrugt et al. (2003) developed a population-based McMC algorithm to enhance the efficiency of McMC sampling integrating Differential Evolution genetic algorithm ideas to Metropolis algoritm, resulting in the Differential Evolution Markov chain Monte Carlo method (DE) that allows for the exchange of information among multiple chains running in parallel. According to Vrugt et al. (2003); Ter Braak (2006a), this choice yields an appropriate scale and orientation for the jumping distribution.

Consider the Nc, d-dimensional, parameters $\boldsymbol{\theta}_i^t$, $i = 1, 2, \ldots, \mathsf{Nc}$, members of population $\mathbf{X}^{(t)}$ at state t. Thus, $\mathbf{X}^{(t)}$ is a $\mathsf{Nc} \times \mathsf{d}$ matrix. To draw the samples and to ensure that the whole parameter space can be reached we consider

$$\boldsymbol{\theta} = \boldsymbol{\theta}_{r_1}^t + \mathsf{g}\left(\boldsymbol{\theta}_{r_2}^t - \boldsymbol{\theta}_{r_3}^t\right) + \boldsymbol{e}, \tag{4}$$

where $\boldsymbol{e}$ is drawn from a symmetric distribution with a small variance compared to that of the target, but with unbounded support, e.g. $\boldsymbol{e} \sim \mathbb{N}\left(\mathbf{0}, b\mathbf{I}_{\mathsf{d}}\right)$ with b small. Here, $\boldsymbol{\theta}_{r_2}^t$ and $\boldsymbol{\theta}_{r_2}^t$ are randomly selected without replacement from the population current $\mathbf{X}^{(t)}$ (the population without $\boldsymbol{\theta}_{r_1}^t$), i.e. $r_1 \neq r_2 \neq r_3$.

2.1 Differential Evolution Metropolis (DE)

For large Nc and small b, the proposal (Eq. 4) thus looks like $\boldsymbol{\theta} = \boldsymbol{\theta}^t + \mathsf{g}\varepsilon$ with $\mathsf{E}[\varepsilon] = \mathbf{0}$ and $\mathsf{cov}(\varepsilon) = 2\boldsymbol{\Sigma}$, the covariance matrix of the target. In particular, if $\pi(\cdot)$ is multivariate normal, then $\mathsf{g}\varepsilon \sim \mathbb{N}\left(\mathbf{0}, 2\mathsf{g}^2\boldsymbol{\Sigma}\right)$ so that DE is expected to behave like RW. From the guidelines for $\mathsf{c_d}$ in (Roberts and Rosenthal, 2001) the optimal choice of g is then $2.38/\sqrt{2\mathsf{d}}$. This choice of g is expected to give an acceptance probability of 0.44 for $\mathsf{d} = 1$, 0.28 for $\mathsf{d} = 5$ and 0.23 for large d. If the initial population is drawn from the prior, DE translates the "prior population" to the "posterior population".

2.1.1 Autoregressive Differential Evolution Metropolis

Consider a set of high dimensional ($\mathsf{d} \gg 1$) random vectors $\boldsymbol{\theta}_i^t$, $i = 1, \ldots, \mathsf{Nc}$ independent and identically distributed. More specifically, $\boldsymbol{\theta}_i^t \sim \mathbb{N}\left(\mathbf{0}, \sigma^2_{\boldsymbol{\theta}^t}\mathbf{I}_\mathsf{d}\right), \quad \forall i$. In other words, $\boldsymbol{\theta}_i^t$ are independent standard Gaussian random variables. Using (Eq. 4) to yield the new generation (proposals) and considering that $\sigma^2_{\boldsymbol{e}}$ is negligible, the variance of $\boldsymbol{\theta}$ is given by:

$$\begin{aligned} \sigma^2_{\boldsymbol{\theta}} &= \sigma^2_{\boldsymbol{\theta}^t_{r_1}} + \mathsf{g}^2\left(\sigma^2_{\boldsymbol{\theta}^t_{r_2}} + \sigma^2_{\boldsymbol{\theta}^t_{r_3}}\right) \\ &= \sigma^2_{\boldsymbol{\theta}^t}\left(1 + 2\mathsf{g}^2\right). \end{aligned} \tag{5}$$

The previous result of Eq. (5) shows that as the iterations t evolve, there is a progressive increase in the variance of the new proposals, which can cause problems in the convergence of the method. To overcome this problem, we propose a modification of DE sampler (Eq. 4) inspired by the autoregressive version of RW (Eq. (**??**)):

$$\boldsymbol{\theta} = \left(\sqrt{1 - 2\mathsf{g}^2}\right)\boldsymbol{\theta}^t_{r_1} + \mathsf{g}\left(\boldsymbol{\theta}^t_{r_2} - \boldsymbol{\theta}^t_{r_3}\right) + \boldsymbol{e}, \tag{6}$$

with the following restriction: $\mathsf{g} = \min\left(2.38/\sqrt{2\mathsf{d}}, \sqrt{1/2}\right)$. In order words, the minimal dimension is 6 ($\mathsf{d} \geqslant 6$). The proposal Eq. (6) ensures that $\sigma^2_{\boldsymbol{\theta}} = \sigma^2_{\boldsymbol{\theta}^t}, \quad \forall i$. The variation of $\mathsf{g}(\mathsf{d})$ and $f(\mathsf{d}) = \sqrt{1 - 2\left(\dfrac{2.38}{\sqrt{\mathsf{d}}}\right)^2}$ are shown in Fig. 1. Again, this method is recommended for high-dimensional problems.

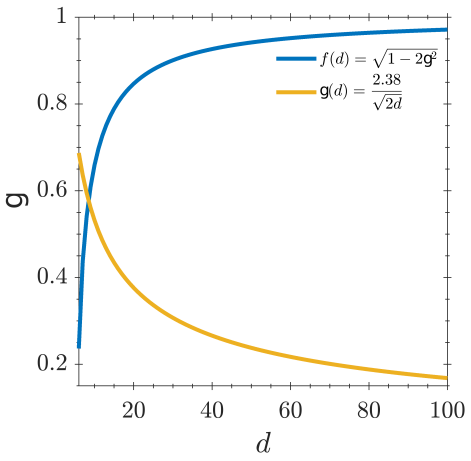


Figure 1: Coefficeents associated with autoregressive jump for autoregressive DE.

2.2 Dimensionality reduction

Another technique that has been widely used to speed up the convergence in the McMC framework is the reduction of the stochastic dimension using the Karhunen-Loève (KL) expansion (Efendiev et al., 2005, 2006; Das et al., 2010; Mondal et al., 2010; Ginting et al., 2011, 2012). The subsurface stochastic field (permeability, porosity, Young's modulus, etc.) may be represented as a series expansion involving a complete set of deterministic functions with correspondent random coefficients using the KL expansion, proposed independently by Karhunen (1946) and Loève (1955), which is based on the eigen-decomposition of the covariance function. If the eigenvalues decay very fast for a specific covariance function, only a small number of terms need to be retained in the truncated expansion. This procedure allows us to perform the search in a smaller parameter space.

3 Introdution

Variational auto-encoder (VAE) model is a stochastic inference and learning algorithm based on variational Bayes (VB) inference proposed by Kingma and Welling (2014). This is a generative that enforces a *prior* on the low-dimensional latent space that can be mapped back into a realistic-looking image. Therefore, the most important characteristic of VAEs, in the context of Monte Carlo methods with Markov chains, is their ability to represent high-dimensional parametric spaces in a low-dimensional latent space.

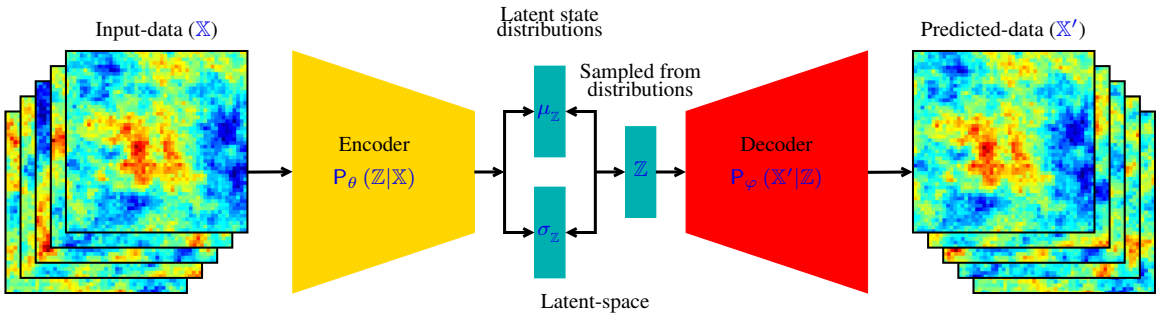
Higgins et al. (2016) introduced the β-VAE, a modification of the original VAE, that introduces an adjustable hyperparameter β to balance latent channel capacity and independence constraints with reconstruction accuracy. They demonstrate that with tuned values of β ($\beta > 1$) the β-VAE outperforms VAE ($\beta = 1$).

Makhzani et al. (2016); Louizos et al. (2017); Burda et al. (2016); Zheng et al. (2019); Vahdat and Kautz (2020)

4 ZHANG ET AL. (2022)

Zhang et al. (2022) proposed a method to reconstruct porous media based on VAE and Fisher information with good quality and efficiency.

Consider the input data set $\mathbb{X} = \left\{\boldsymbol{x}^{(i)}\right\}_{i=1}^{\mathsf{N}}$ ($\boldsymbol{x}^{(i)} \in \mathbb{R}^{\mathsf{N}_x}$) consisting of N *i*ndependent and *i*dentically *d*istributed (*i.i.d.*) samples of the continuous (or discrete) variable drawn from the *prior* distribution $p(\boldsymbol{x})$.



$$\mathbb{Z} = \mu_{\mathbb{Z}} + \sigma_{\mathbb{Z}} \cdot \varepsilon, \quad \text{where} \quad \varepsilon \sim \mathbb{N}(0,1) \tag{7}$$

The reconstruction loss is used to ensures that input image is reconstructes at the output one and, here, is given by the mean squared error (MSE):

$$\mathcal{L}_{\mathsf{MSE}}(\theta, \phi, \boldsymbol{x}) = \frac{1}{\mathsf{N}_b} \sum_{i=1}^{\mathsf{N}_b} \left[\boldsymbol{x}^{(i)} - \mathsf{D}_\phi\left(\mathsf{E}_\theta\left(\boldsymbol{x}^{(i)}\right)\right)\right]^2, \tag{8}$$

where E and D represent the encoder and decoder and θ, ϕ are their parameters, respectively.

In VAE, we assume that both *prior* distribution $p(\mathbb{Z}) \simeq \mathbb{N}(0,1)$ and *posterior* approximation of the latent space follow a standard Gaussian distribution, i.e., $q(\mathbb{Z}|\boldsymbol{x}) \simeq \mathbb{N}(0,1)$.

To keep the encoder outputs $\mathbb{Z}$ close to a standard normal distribution and sufficiently diverse we use the Kullback–Leibler divergence ($\mathcal{D}_{\mathsf{KL}}$, also called relative entropy and I-divergence). $\mathcal{D}_{\mathsf{KL}}$ is a measure of divergence between two distributions (Kullback and Leibler, 1951; Csiszar, 1975):

$$\mathcal{D}_{\mathsf{KL}}\left(p(\mathbb{Z})||\mathbb{N}(0,1)\right) = -\frac{1}{2} \sum_{i=1}^{\mathsf{N}_z} \left[1 + \log\left(\sigma_{\mathbb{Z}_i}^2\right) - {\mu_{\mathbb{Z}_i}}^2 - \sigma_{\mathbb{Z}_i}^2\right] \tag{9}$$

Zheng et al. (2019) proposed a Fisher autoencoder

5 Fisher Information

Let $f(x;\theta)$ be the probability density function of the random variable $\mathbb{X}$ conditioned on the parameter θ. The Fisher information measures the amount of information that an observation of $\mathbb{X}$ carries about the unknown parameter θ. The partial derivative of the natural logarithm of the likelihood function is called **score** (S):

$$\mathsf{S}(x,\theta) = \frac{\partial}{\partial\theta}\log\left[f(x;\theta)\right]. \tag{10}$$

Fisher information is defined as the variance of the score S:

$$\begin{aligned}\mathcal{I}(\theta) = \mathsf{E}[\mathsf{S}^2|\theta] &= \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\log\left[f(x;\theta)\right]\right)^2\middle|\theta\right] \\ &= \int_{\mathbb{R}}\left(\frac{\partial}{\partial\theta}\log\left[f(x;\theta)\right]\right)^2 f(x;\theta)\,\mathsf{d}x\end{aligned}. \tag{11}$$

If $\log\left[f(x;\theta)\right]$ is twice differentiable with respect to θ and under certain regularuty conditions, Fisher information can be written as

$$\mathcal{I}(\theta) = \mathsf{E}\left[-\frac{\partial^2}{\partial\theta^2}\log\left[f(x;\theta)\right]\middle|\theta\right]. \tag{12}$$

Let $\mathbb{X}$ be a scalar Gaussian random variable *i.e.* $\mathbb{X}\sim\mathbb{N}(\mu,\sigma^2)$. Then the probability density function is parameterized by the parameters μ and σ:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]. \tag{13}$$

Substituting Eq. (13) in Eq. (12) where $\theta=\mu$ or σ we can compute de Fisher information for a Gaussian variable as:

$$\begin{aligned}
\mathcal{I}(\mu) &= \mathsf{E}\left[-\frac{\partial^2}{\partial\mu^2}\left[\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{1}{2}\left(\frac{\mathbb{X}-\mu}{\sigma}\right)^2\right]\Big|\mu\right] \\
&= \mathsf{E}\left[-\frac{\partial}{\partial\mu}\left(-\frac{\mathbb{X}-\mu}{\sigma^2}\right)\Big|\mu\right] \\
&= \mathsf{E}\left[\frac{1}{\sigma^2}\right]=\frac{1}{\sigma^2}
\end{aligned} \tag{14}$$

$$\begin{aligned}
\mathcal{I}(\sigma) &= \mathsf{E}\left[-\frac{\partial}{\partial\sigma}\left[-\frac{1}{\sigma}+\frac{(\mathbb{X}-\mu)^2}{\sigma^3}\right]\Big|\sigma\right] \\
&= \mathsf{E}\left[-\frac{1}{\sigma^2}+\frac{3(\mathbb{X}-\mu)^2}{\sigma^4}\Big|\sigma\right] \\
&= -\frac{1}{\sigma^2}+\frac{3\sigma^2}{\sigma^4}=\frac{2}{\sigma^2}
\end{aligned} \tag{15}$$

6 KARHUNEN-LOÈVE EXPANSION

Due to incomplete knowledge about the rock properties that show variability at multiple length scales, input parameters such as the permeability field, $\boldsymbol{\kappa}(\mathbf{x},\omega)$, are treated as random space functions with statistics inferred from geostatistical models (here $\mathbf{x}=(x_1,x_2,x_3)^{\mathsf{T}}\in\mathbb{R}^3$ and ω is a random element in the probability space). In line with Dagan (1989) and Gelhar (1993) the permeability field is modeled as a log-normally distributed random space function

$$\boldsymbol{\kappa}(\mathbf{x},\omega)=\beta\exp\Big[\rho\mathsf{Y}(\mathbf{x},\omega)\Big], \tag{16}$$

where $\beta,\rho\in\mathbb{R}^+$ and $\mathsf{Y}(\mathbf{x},\omega)\sim\mathbb{N}(\mu_\mathsf{Y},\boldsymbol{\mathcal{C}}_\mathsf{Y})$ is a Gaussian random field characterized by its mean $\mu_\mathsf{Y}=\langle\mathsf{Y}\rangle$ and two-point covariance function

$$\boldsymbol{\mathcal{C}}_\mathsf{Y}(\mathbf{x},\mathbf{y})=\mathsf{Cov}\left(\mathsf{Y}(\mathbf{x}),\mathsf{Y}(\mathbf{y})\right)=\mathsf{E}\Big[\big(\mathsf{Y}(\mathbf{x})-\langle\mathsf{Y}(\mathbf{x})\rangle\big)\big(\mathsf{Y}(\mathbf{y})-\langle\mathsf{Y}(\mathbf{y})\rangle\big)\Big]. \tag{17}$$

Moreover, in this work, Y is a second-order stationary process (Gelhar, 1993), that is:

$$\begin{aligned}
\langle\mathsf{Y}(\mathbf{x})\rangle &= \mu_\mathsf{Y}, \quad \text{(constant)} \\
\boldsymbol{\mathcal{C}}_\mathsf{Y}(\mathbf{x},\mathbf{y}) &= \boldsymbol{\mathcal{C}}_\mathsf{Y}\left(\|\mathbf{x}-\mathbf{y}\|\right)=\boldsymbol{\mathcal{C}}_\mathsf{Y}(d)\,.
\end{aligned} \tag{18}$$

6 Karhunen-Loève expansion

The Gaussian field Y can be represented as a series expansion involving a complete set of deterministic functions with correspondent random coefficients using the Karhunen-Loève (KL) expansion proposed independently by Karhunen (1946) and Loève (1955). It is based on the eigen-decomposition of the covariance function. Depending on how fast the eigenvalues decay one may be able to retain only a small number of terms in a truncated expansion and, consequently, this procedure may reduce the search to a smaller parameter space. In uncertainty quantification methods for porous media flows, the KL expansion has been widely used to reduce the number of parameters used to represent the permeability field (Efendiev et al., 2005, 2006; Das et al., 2010; Mondal et al., 2010; Ginting et al., 2011, 2012). Another advantage of KL expansion lies on the fact that it provides orthogonal deterministic basis functions and uncorrelated random coefficients, allowing for the optimal encapsulation of the information contained in the random process into a set of discrete uncorrelated random variables (Ghanem and Spanos, 1991). This remarkable feature can be used to simplify the Metropolis-Hastings McMC Algorithm in the sense of the search may be performed in the space of discrete uncorrelated random variables ($\boldsymbol{\theta}$), no longer in the space of permeabilities which have a more complex statistical structure.

Here we recall the basic facts about the Karhunen-Loève expansion. Consider a random field $\mathsf{Y}(\mathbf{x},\omega)$ defined on a probability space $(\Omega,\mathcal{A},\mathcal{P})$ composed by the sample space, the ensemble of events and a probability measure, respectively, and indexed on a bounded domain $\mathcal{D}\in\mathbb{R}^3$. The process Y can be expressed as

$$\mathsf{Y}(\mathbf{x},\omega)=\langle\mathsf{Y}(\mathbf{x})\rangle+\sum_{i=1}^{\infty}\sqrt{\lambda_i}\phi_i(\mathbf{x})\theta_i(\omega), \tag{19}$$

where λ_i and ϕ_i are the eigenvalues and eigenfunctions of the covariance function $\boldsymbol{\mathcal{C}}_{\mathsf{Y}}(\mathbf{x},\mathbf{y})$, respectively. By definition, $\boldsymbol{\mathcal{C}}_{\mathsf{Y}}(\mathbf{x},\mathbf{y})$ is bounded, symmetric and positive definite and has the following eigen-decomposition:

$$\boldsymbol{\mathcal{C}}_{\mathsf{Y}}(\mathbf{x},\mathbf{y})=\sum_{i=1}^{\infty}\lambda_i\phi_i(\mathbf{x})\phi_i(\mathbf{y}).$$

The eigenvalues and eigenfunctions of Eq. (19) are the solution of the homogeneous Fredholm integral equation of second kind given by

$$\int_{\mathcal{D}}\boldsymbol{\mathcal{C}}_{\mathsf{Y}}(\mathbf{x},\mathbf{y})\phi(\mathbf{x})d\mathbf{x}=\lambda\phi(\mathbf{y}). \tag{20}$$

The solution of Eq. (20) forms a complete set of a square-integrable orthogonal eigenfunctions that satisfy the equation

$$\int_{\mathcal{D}}\phi_i(\mathbf{x})\phi_j(\mathbf{x})=\delta_{ij},$$

in which δ_{ij} is the Kronecker-delta function. $\theta_i(\omega)$ is a set of independent random variables which can be expressed as

$$\theta_i(\omega)=\frac{1}{\sqrt{\lambda_i}}\int_{\mathcal{D}}\tilde{\mathsf{Y}}\phi_i(\mathbf{x})d\mathbf{x},$$

where $\tilde{\mathsf{Y}} = \mathsf{Y} - \langle \mathsf{Y} \rangle$ is the fluctuation. For practical implementations of the KL expansion the eigenvalues are arranged from the largest to smallest and the series is approximated by a finite number of terms, say the first m, giving

$$\mathsf{Y}(\mathbf{x}, \omega) \approx \langle \mathsf{Y}(\mathbf{x}) \rangle + \sum_{i=1}^{\mathsf{m}} \sqrt{\lambda_i} \phi_i(\mathbf{x}) \theta_i(\omega). \tag{21}$$

The corresponding covariance function is given by

$$\mathcal{C}_{\hat{\mathsf{Y}}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{\mathsf{m}} \lambda_i \phi_i(\mathbf{x}) \phi_i(\mathbf{y}).$$

The factors affecting the convergence of the Karhunen-Loève series are the ratio of the length of the process over correlation parameter, the form of the covariance function, and the solution method for the eigensolutions of the covariance function (see Huang et al. (2001)). Next we discuss the field conditioning using the KL expansion.

The number of terms used in the series can be chosen based on the energy represented by the sum of the eigenvalues. Then we define the relative energy for n terms as

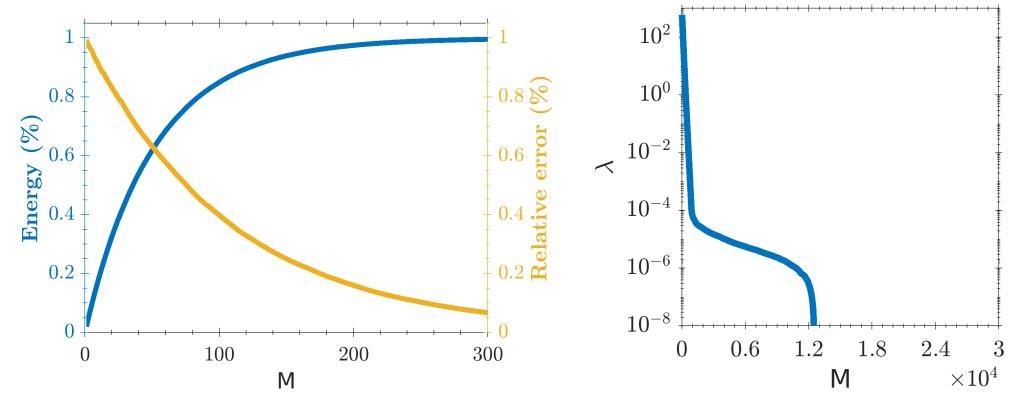
$$\mathsf{ER}_{\mathsf{n}} = \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{j=1}^{m \to \infty} \lambda_j}. \tag{22}$$

The eigenvalues associated with the eigenfunctions provide a measure of the energy contained in the respective mode. To quantify the energy involved in a truncated KL expansion (with M modes) we define the total relative energy as

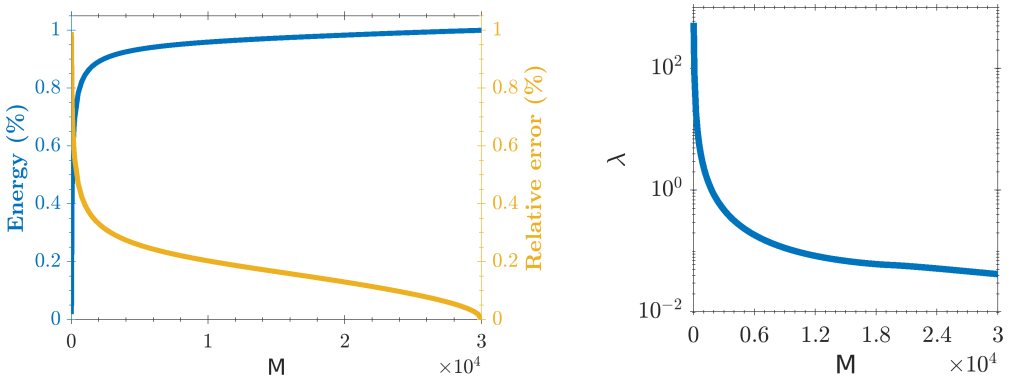
$$\mathcal{E}_{\mathsf{M}} = \frac{\sum_{j=1}^{\mathsf{M}} \lambda_j}{\sum_{i=1}^{m \to \infty} \lambda_i}, \tag{23}$$

in which the denominator will be approximated by $m = 30,000$ eigenvalues.

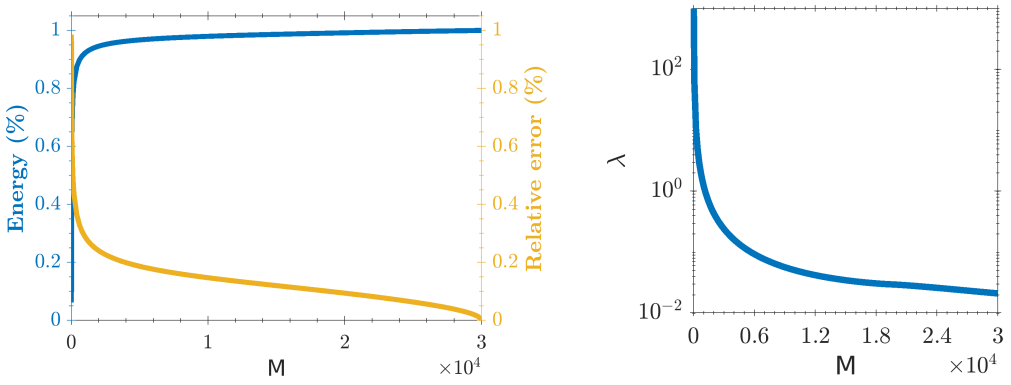
The Fig. 2 shows the behavior of the eigenvalues, the energy contained, and the average error as a function of the number of modes (M) for the three covariance functions considered in this study. The decay of the eigenvalues of the square exponential case is much faster than the exponential case. The latter requires a more significant number of modes to capture the same energy when compared to the first one.



(a) Squared exponential covariance ($\ell = 0.1$)

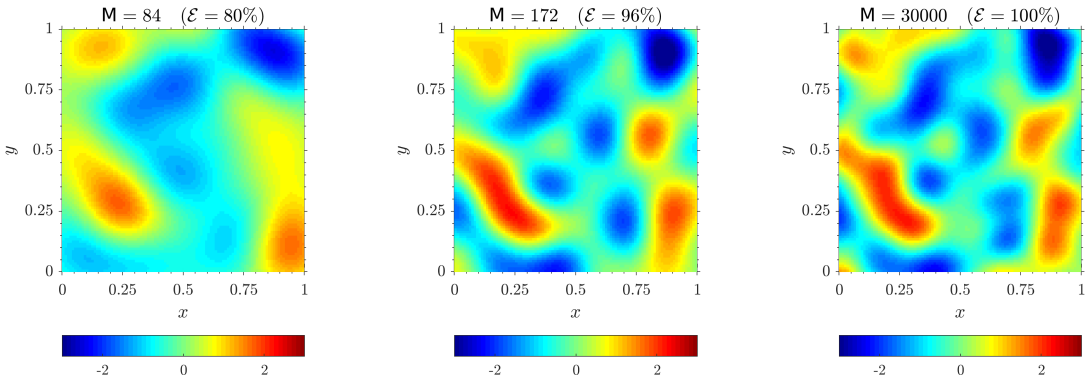


(b) Exponential covariance ($\ell = 0.1$)

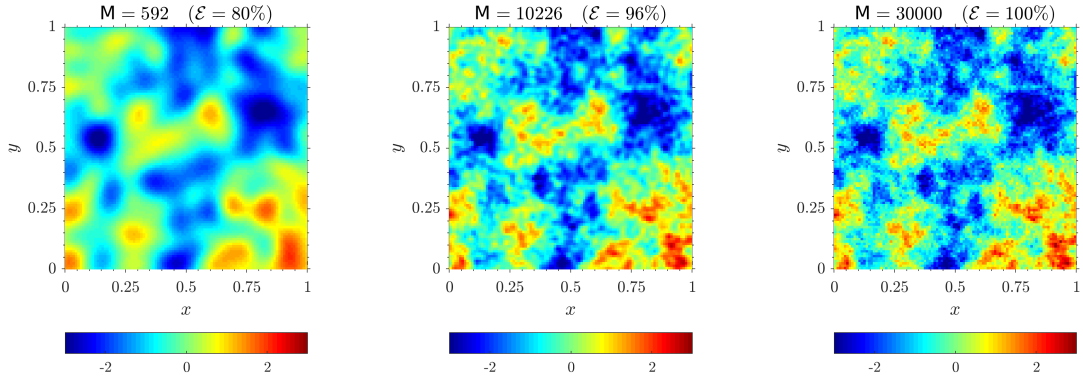


(c) Exponential covariance ($\ell = 0.2$)

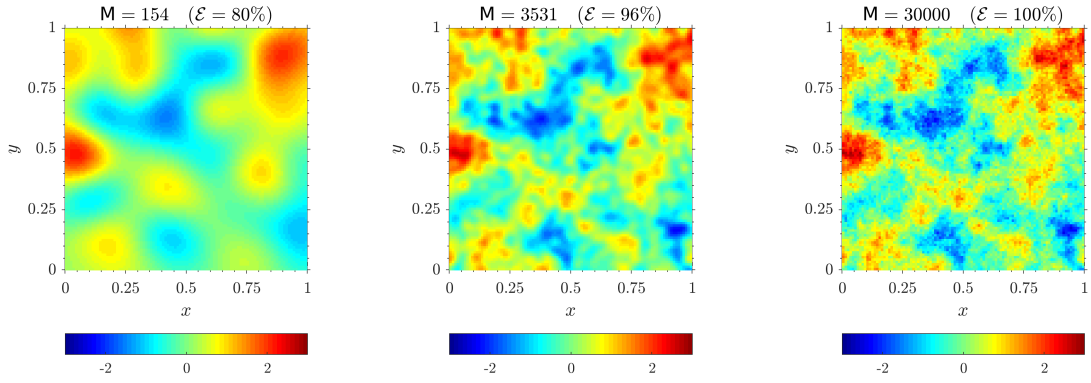
Figure 2: Decay of eigenvalues and contained energy as a function of the number of terms in the expansion.



(a) Squared exponential, $\ell = 0.1$



(b) Exponential, $\ell = 0.1$

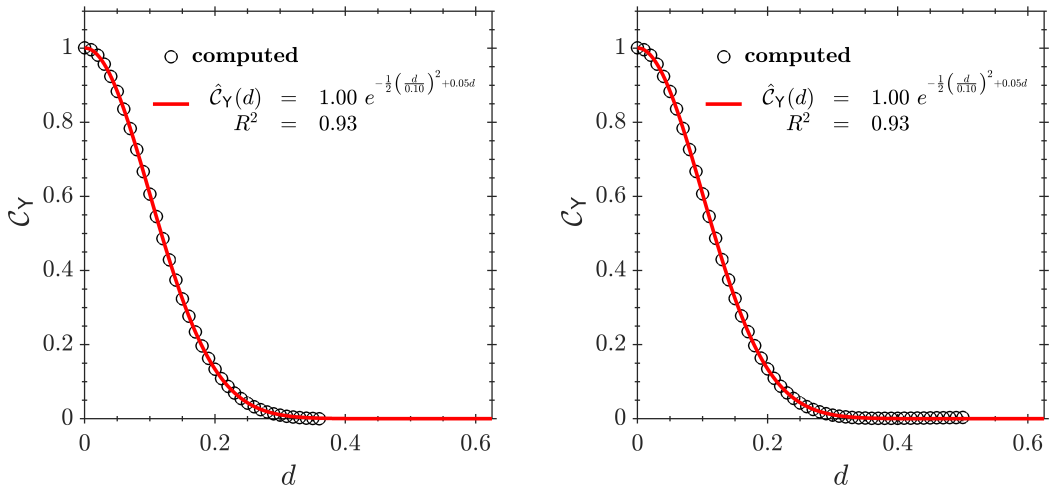


(c) Exponential, $\ell = 0.2$

Figure 3: Random fields generated with different numbers of modes M (or, equivalently, amount of energy).

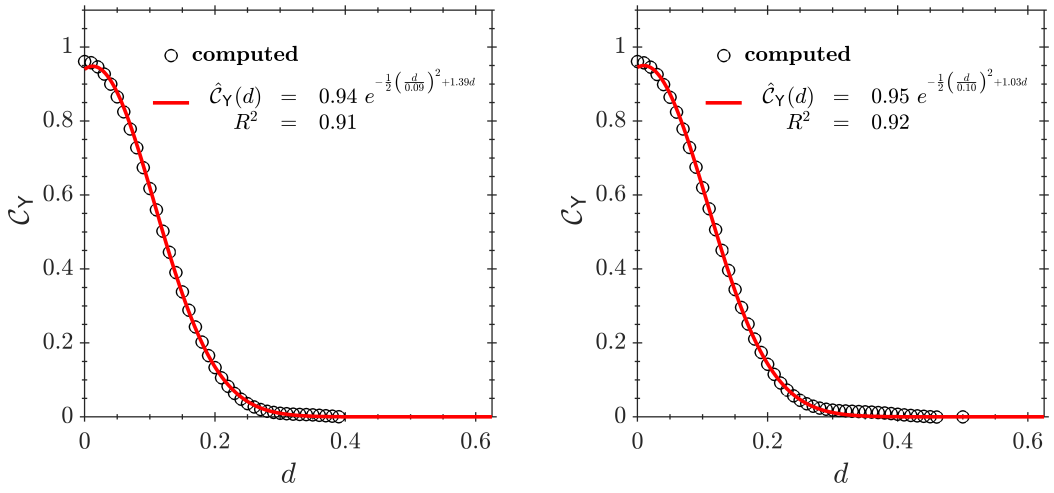
Table 1: Number of KL expansion terms needed to obtain an given energy level

Energy	M			**Mean relative error**		
(%)	Squared Exp. ($\ell = 0.1$)	Exponential ($\ell = 0.1$)	Exponential ($\ell = 0.2$)	Squared Exp. ($\ell = 0.1$)	Exponential ($\ell = 0.1$)	Exponential ($\ell = 0.2$)
80	84	592	154	0.46	0.45	0.47
90	122	2,323	612	0.32	0.32	0.33
94	150	5,760	1,655	0.25	0.25	0.25
96	172	10,226	3,531	0.21	0.20	0.21
98	211	18,327	10,239	0.14	0.14	0.15
100	∼30,000	∼30,000	∼30,000	0.00	0.00	0.00



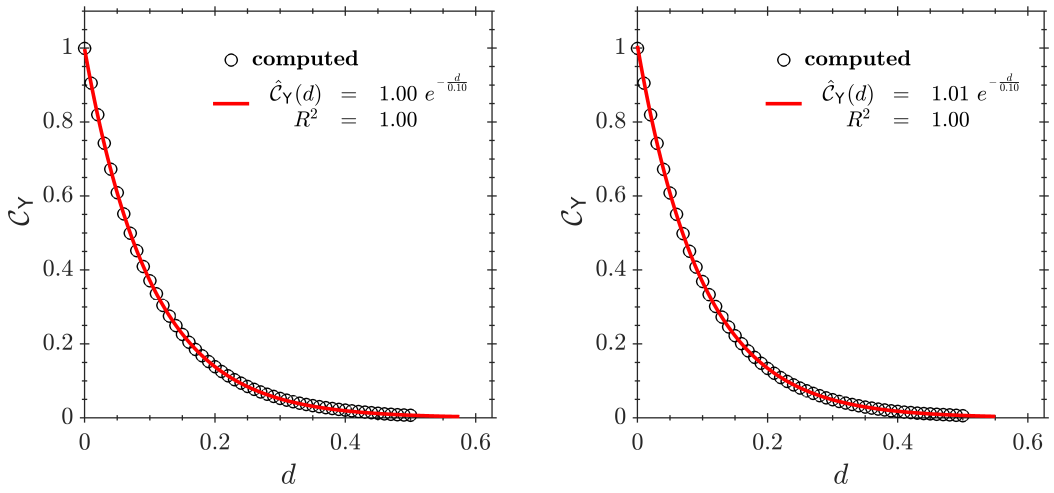
(a) x direction (b) y direction

Figure 4: Estimated covariance in a sample of $5,000$ fields generate by KLE with $30,000$ terms. Squared exponential case with $\ell = 0.1$.



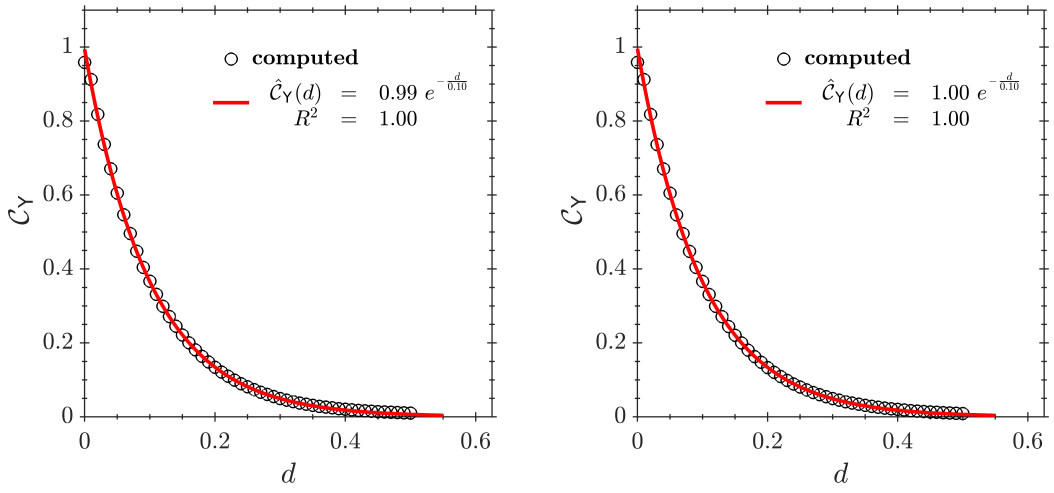
(a) x direction (b) y direction

Figure 5: Estimated covariance in a sample of $5,000$ fields generate by `KLE` with 172 terms. Squared exponential case with $\ell = 0.1$.



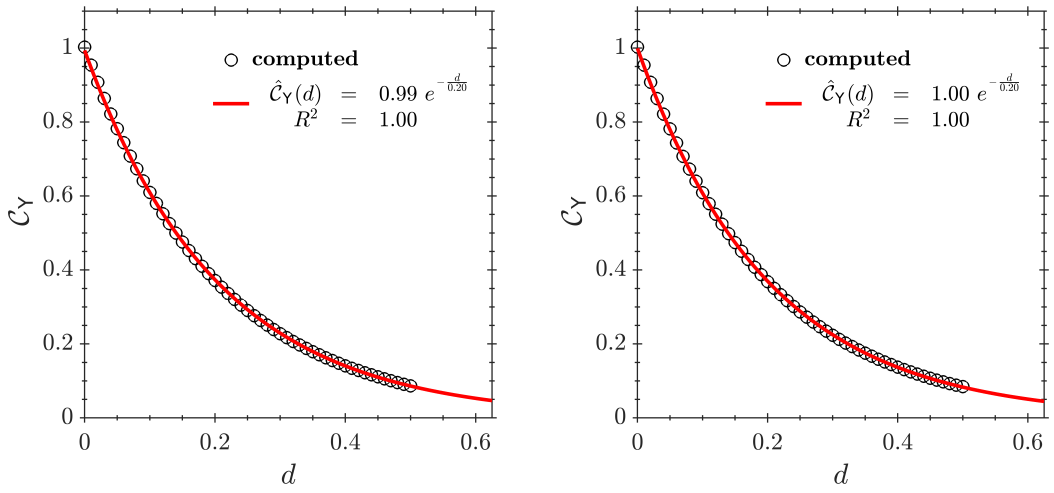
(a) x direction (b) y direction

Figure 6: Estimated covariance in a sample of $5,000$ fields generate by `KLE` with $30,000$ terms. Exponential case with $\ell = 0.1$.



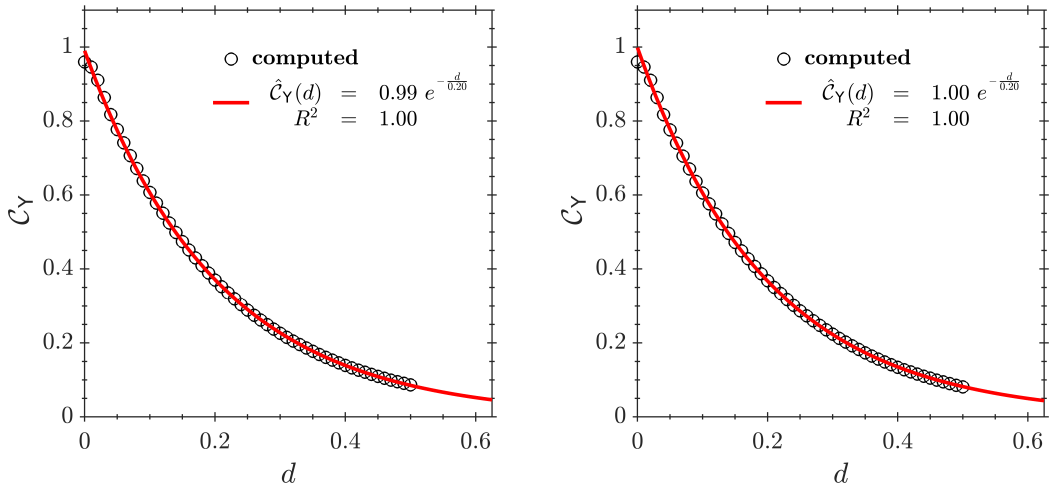
(a) x direction (b) y direction

Figure 7: Estimated covariance in a sample of $5,000$ fields generate by `KLE` with $10,226$ terms. Exponential case with $\ell = 0.1$.



(a) x direction (b) y direction

Figure 8: Estimated covariance in a sample of $5,000$ fields generate by `KLE` with $30,000$ terms. Exponential case with $\ell = 0.2$.



(a) x direction (b) y direction

Figure 9: Estimated covariance in a sample of $5,000$ fields generate by `KLE` with $3,531$ terms. Exponential case with $\ell = 0.2$.

???

? proposed a multiscale Bayesian inference approach based on a multiscale deep generative model (MDGM)

REFERENCES

Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. Importance weighted autoencoders, 2016.

I. Csiszar. I-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability*, 3(1):146 – 158, 1975. doi: 10.1214/aop/1176996454. URL https://doi.org/10.1214/aop/1176996454.

G. Dagan. *Flow and transport in porous formations*. Springer-Verlag, 1989.

N. N. Das, B. P. Mohanty, and Y. Efendiev. Characterization of effective saturated hydraulic conductivity in an agricultural field using Karhunen-Loève expansion with the Markov chain Monte Carlo technique. *Water Resources Research*, 46, W06521 2010. doi: 10.1029/2008WR007100.

Nando de Freitas, Pedro Hojen-Sorensen, Michael I. Jordan, and Stuart Russell. Variational mcmc, 2013.

P. Dostert, Y. Efendiev, and B. Mohanty. Efficient uncertainty quantification techniques in inverse problems for richards' equation using coarse-scale simulation models. *Advances in Water Resources*, 32(3):329–339, 2009. doi: DOI:10.1016/j.advwatres.2008.11.009.

Y. Efendiev, A. Datta-gupta, V. Ginting, X. Ma, and B. Mallick. An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. *Water Resources Research*, 41: 12423, 2005.

Y. Efendiev, T. Hou, and W. Luo. Preconditioning Markov Chain Monte Carlo Simulations Using Coarse-Scale Models. *SIAM J. Sci. Comput.*, 28:776–803, 2006.

A. A. Emerick and A. C. Reynolds. Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. *Computational Geosciences*, 17(2):325–350, 2013. ISSN 1420-0597. doi: 10.1007/s10596-012-9333-z. URL http://dx.doi.org/10.1007/s10596-012-9333-z.

L. W. Gelhar. *Stochastic subsurface hydrology*. Englewood Cliffs. Prentice-Hall, 1993.

R. Ghanem and P.D. Spanos. *Stochastic Finite Element: A Spectral Approach*. Springer, New York, 1991.

V. Ginting, F. Pereira, M. Presho, and S. Wo. Application of the two-stage Markov chain Monte Carlo method for characterization of fractured reservoirs using a surrogate flow model. *Computational Geosciences*, 15:691–707, 2011.

V. Ginting, F. Pereira, and A. Rahunanthan. Multiple Markov chains Monte Carlo approach for flow forecasting in porous media. *Procedia Computer Science*, 5:707–716, 2012.

REFERENCES

V. Ginting, F. Pereira, and A. Rahunanthan. A multi-stage bayesian prediction framework for subsurface flows. *International Journal for Uncertainty Quantification*, 2013. to appear.

J. Glimm, W. B. Lindquist, F. Pereira, and R. Peierls. The multi-fractal hypothesis and anomalous diffusion. *Mat. Aplic. Comp.*, 11(2):189–207, 1992.

H. Haario, E. Saksman, and J. Tamminen. Adaptive proposal distribution for random walk metropolis algorithm. *Computational Statistics*, 14(3):375–395, 1999. doi: 10.1007/s001800050022. URL http://dx.doi.org/10.1007/s001800050022.

H. Haario, E. Saksman, and J. Tamminen. Componentwise adaptation for high dimensional mcmc. *Computational Statistics*, 20(2):265–273, 2005. ISSN 1613-9658. doi: 10.1007/BF02789703. URL http://dx.doi.org/10.1007/BF02789703.

Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 04 2001. URL http://projecteuclid.org/euclid.bj/1080222083.

Heikki Haario, Marko Laine, Antonietta Mira, and Eero Saksman. Dram: Efficient adaptive mcmc. *Statistics and Computing*, 16(4):339–354, 2006. ISSN 1573-1375. doi: 10.1007/s11222-006-9438-0. URL http://dx.doi.org/10.1007/s11222-006-9438-0.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97, 1970.

Irina Higgins, Loïc Matthey, Arka Pal, Christopher P. Burgess, Xavier Glorot, Matthew M. Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2016.

S. P. Huang, S. T. Quek, and K. K. Phoon. Convergence study of the truncated Karhunen-Loève expansion for simulation of stochastic processes. *International Journal for Numerical Methods in Engineering*, 52(9):1029–1043, 2001.

M. A. Iglesias, K. J. H. Law, and A. M. Stuart. Evaluation of gaussian approximations for data assimilation in reservoir models. *Computational Geosciences*, pages 1–35, 2013. doi: 10.1007/s10596-013-9359-x.

K. Karhunen. Zur spektraltheorie stochastischer prozesse. *Ann. Acad. Sci. Fennicae*, 1946.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *International conference on learning representations*, pages 14–27, 2014.

S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79 – 86, 1951. doi: 10.1214/aoms/1177729694. URL https://doi.org/10.1214/aoms/1177729694.

REFERENCES

Gaisheng Liu, Yan Chen, and Dongxiao Zhang. Investigation of flow and transport processes at the made site using ensemble kalman filter. *Advances in Water Resources*, 31(7):975–986, 2008. doi: 10.1016/j.advwatres.2008.03.006.

N. Liu and D. S. Oliver. Evaluation of monte carlo methods for assessing uncertainty. *SPE Journal*, 8:188–195, 2003. doi: 10.2118/84936-PA.

M. M. Loève. *Probability Theory*. Princeton, N.J., 1955.

Christos Louizos, Kevin Swersky, Yujia Li, Max Welling, and Richard Zemel. The variational fair autoencoder, 2017.

X. Ma, M. Al-Harbi, A. Datta-Gupta, and Y. Efendiev. An efficient two-stage sampling method for uncertainty quantification in hystory matching geological models. *SPE Journal*, 2008.

G. Madey, X. Xiang, S. E. Cabaniss, and Y. Huang. Agent-based scientific simulation. *Computing in Science & Engineering*, 2(01):22–29, jan 2005. ISSN 1558-366X. doi: 10.1109/MCSE.2005.7.

Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, and Ian Goodfellow. Adversarial autoencoders. In *International Conference on Learning Representations*, 2016. URL http://arxiv.org/abs/1511.05644.

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: 10.1063/1.1699114. URL http://dx.doi.org/10.1063/1.1699114.

A. Mondal, Y. Efendiev, B. Mallick, and A. Datta-Gupta. Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump Markov chain Monte Carlo methods. *Advances in Water Resources*, 33(3):241 – 256, 2010.

Christopher Nemeth and Paul Fearnhead. Stochastic gradient markov chain monte carlo. *Journal of the American Statistical Association*, 116(533):433–450, 2021. doi: 10.1080/01621459.2020.1847120. URL https://doi.org/10.1080/01621459.2020.1847120.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling for various metropolis-hastings algorithms. *Statist. Sci.*, 16(4):351–367, 11 2001. doi: 10.1214/ss/1015346320. URL http://dx.doi.org/10.1214/ss/1015346320.

Cajo J. F. Ter Braak. A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3):239–249, 2006a. ISSN 1573-1375. doi: 10.1007/s11222-006-8769-1. URL http://dx.doi.org/10.1007/s11222-006-8769-1.

REFERENCES

Cajo J. F. Ter Braak. A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing*, 16 (3):239–249, Sep 2006b. ISSN 1573-1375. doi: 10.1007/s11222-006-8769-1. URL https://doi.org/10.1007/s11222-006-8769-1.

Luke Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22: 1701–1762, 1994.

Luke Tierney and Antonietta Mira. Some adaptive monte carlo methods for bayesian inference. *Statistics in Medicine*, 18(17-18):2507–2515, 1999. ISSN 1097-0258. doi: 10.1002/(SICI)1097-0258(19990915/30)18:17/18<2507::AID-SIM272>3.0.CO;2-J.

Arash Vahdat and Jan Kautz. Nvae: a deep hierarchical variational autoencoder. In *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, number 1650, page 19667–19679, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

J.A. Vrugt, H. V. Gupta, W. Bouten, and S. Sorooshian. A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8):1201–1217, 2003.

Jasper A. Vrugt, Cajo J. F. Ter Braak, Martyn P. Clark, James M. Hyman, and Bruce A. Robinson. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation. *Water Resources Research*, 44(12):n/a–n/a, 2008. ISSN 1944-7973. doi: 10.1029/2007WR006720. URL http://dx.doi.org/10.1029/2007WR006720. W00B09.

Yingzhi Xia and Nicholas Zabaras. Bayesian multiscale deep generative model for the solution of high-dimensional inverse problems. *Journal of Computational Physics*, 455:111008, 2022. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2022.111008. URL https://www.sciencedirect.com/science/article/pii/S0021999122000705.

Yingzhi Xia, Qifeng Liao, and Jinglai Li. Vi-dgp: A variational inference method with deep generative prior for solving high-dimensional inverse problems. *Journal of Scientific Computing*, 97(16), 2023.

Zhihang Xu, Yingzhi Xia, and Qifeng Liao. A domain-decomposed VAE method for bayesian inverse problems. *CoRR*, abs/2301.05708, 2023. doi: 10.48550/arXiv.2301.05708. URL https://doi.org/10.48550/arXiv.2301.05708.

Ting Zhang, Tu Hongyan, Xia Pengfei, and Du Yi. Reconstruction of porous media using an information variational auto-encoder. *Transport in Porous Media*, 143(2):271–295, 2022. doi: 10.1007/s11242-022-01769-5. URL https://doi.org/10.1007/s11242-022-01769-5.

REFERENCES

Huangjie Zheng, Jiangchao Yao, Ya Zhang, Ivor Tsang, and Jia Wang. Understanding vaes in fisher-shannon plane. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33: 5917–5924, 07 2019. doi: 10.1609/aaai.v33i01.33015917.

INDEX