Laboratório
Nacional de
Computação
Científica

COMOD
Coordenação de Modelagem Computacional

Relatório de Atividades

Servidor: Marcio Borges

Petrópolis-RJ
September 26, 2023

1 Introdution

Natural reservoirs display a high degree of variability in their hydraulic properties in multiple length scales (**??**). Such variability plays a strong impact on flu-id flow patterns in subsurface formations (**?**). 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 **?**, 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 (**??**).

Despite its high computational cost that, in some cases, can be prohibitive, McMC methods are regarded as the gold standard technique for Bayesian inference (**?**). 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” (**?**)).

Within this context McMC methods have been widely used in the last decades in porous media flow problems (**???????????**). 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 (**?**). 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 **?** and **?** the permeability field is modeled as a log-normally distributed random space function

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

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

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

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

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

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 **?** and **?**. 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 (**??????**). 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 (**?**). 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 **?** 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 **?**.

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 (???). 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 (?).

Several methods have been developed in order to speed up the convergence: Adaptive Proposal (AP) (?); Adaptive Metropolis (AM) (?); Delayed Rejection (DR) (??); Delayed Rejection Adaptive Metropolis algorithm (DRAM) (?); Single Component Adaptive Metropolis (SCAM) (?); Differential Evolution Markov Chain (DE) (??); DiffeRential Evolution Adaptive Metropolis (DREAM) (?), 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.

? 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 ??, 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. $r_1 \neq r_2 \neq r_3$.

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 (?) 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 Differential Evolution Metropolis (DE)

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 $f(\mathsf{d}) = \sqrt{1 - 2\left(\frac{2.38}{\sqrt{\mathsf{d}}}\right)^2}$ are shown in Fig. 1. Again, this method is recommended for high-dimensional problems.

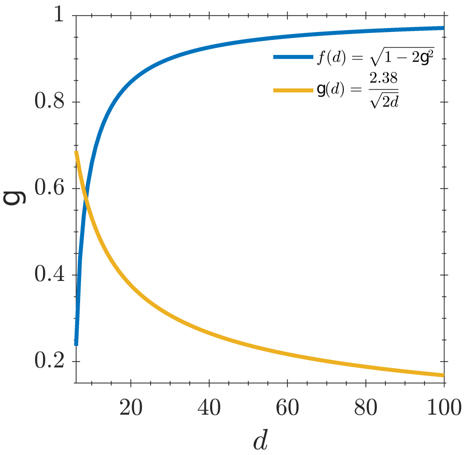


Figure 1: Coefficeents associated with autoregressive jump for autoregressive DE.

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 (**??????**). 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 **?** and **?**, 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 **?**. This is a generative technique whose central idea is the development of representations in a low-dimensional latent space that can be mapped back into a realistic-looking image.

? 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 ($\beta = 1$).

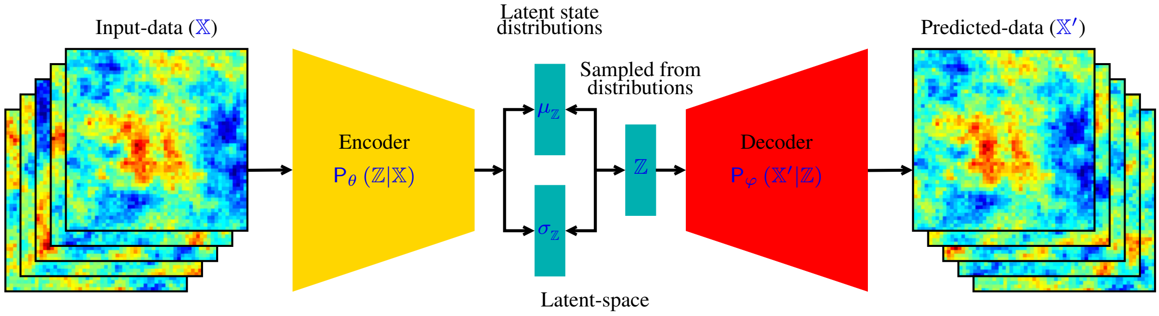
?????

4 ?

Variational auto-encoder (VAE) model is a generative network based on variational Bayes (VB) inference proposed by **?**.

? proposed a method to reconstruct porous media based on VAE and Fisher information with good quality and efficiency.

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


Input-data ($\mathbb{X}$)
Latent state distributions
Predicted-data ($\mathbb{X}'$)
Sampled from distributions
Encoder
$\mathsf{P}_\theta(\mathbb{Z}|\mathbb{X})$
$\mu_\mathbb{Z}$
$\mathbb{Z}$
$\sigma_\mathbb{Z}$
Decoder
$\mathsf{P}_\varphi(\mathbb{X}'|\mathbb{Z})$
Latent-space

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

The Kullback–Leibler divergence (also called relative entropy and I-divergence) is a measure of divergence between two distributions (**??**):

$$\mathcal{D}_{\mathsf{KL}}\left(f_{\mathbb{P}} || f_{\mathbb{Q}}\right) = \int_{-\infty}^{\infty} \tag{8}$$

? proposed a Fisher autoencoder

5 Fisher Information

Let $f\left(x;\theta\right)$ 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):

$$\mathsf{S}\left(x, \theta\right) = \frac{\partial}{\partial\theta} \log\left[f\left(x;\theta\right)\right]. \tag{9}$$

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

$$\begin{aligned} \mathcal{I}\left(\theta\right) = \mathsf{E}[\mathsf{S}^2|\theta] &= \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\log\left[f\left(x;\theta\right)\right]\right)^2 \middle|\theta\right] \\ &= \int_{\mathbb{R}} \left(\frac{\partial}{\partial\theta}\log\left[f\left(x;\theta\right)\right]\right)^2 f\left(x;\theta\right)\mathrm{d}x \end{aligned} . \tag{10}$$

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

$$\mathcal{I}\left(\theta\right) = \mathsf{E}\left[-\frac{\partial^2}{\partial\theta^2}\log\left[f\left(x;\theta\right)\right] \middle|\theta\right]. \tag{11}$$

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

$$f\left(x;\mu,\sigma\right)=\frac{1}{\sigma\sqrt{2\pi}}\text{exp}\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}\right]. \tag{12}$$

Substituting Eq. (12) in Eq. (11) where $\theta=\mu$ or σ we can compute de Fisher information for a Gaussian variable as:

$$\begin{aligned}\mathcal{I}\left(\mu\right) &= \mathsf{E}\left[-\frac{\partial^{2}}{\partial\mu^{2}}\left[\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{1}{2}\left(\frac{\mathbb{X}-\mu}{\sigma}\right)^{2}\right]\bigg|\mu\right]\\ &= \mathsf{E}\left[-\frac{\partial}{\partial\mu}\left(-\frac{\mathbb{X}-\mu}{\sigma^{2}}\right)\bigg|\mu\right]\\ &= \mathsf{E}\left[\frac{1}{\sigma^{2}}\right]=\frac{1}{\sigma^{2}}\end{aligned} \tag{13}$$

$$\begin{aligned}\mathcal{I}\left(\sigma\right) &= \mathsf{E}\left[-\frac{\partial}{\partial\sigma}\left[-\frac{1}{\sigma}+\frac{\left(\mathbb{X}-\mu\right)^{2}}{\sigma^{3}}\right]\bigg|\sigma\right]\\ &= \mathsf{E}\left[-\frac{1}{\sigma^{2}}+\frac{3\left(\mathbb{X}-\mu\right)^{2}}{\sigma^{4}}\bigg|\sigma\right]\\ &= -\frac{1}{\sigma^{2}}+\frac{3\sigma^{2}}{\sigma^{4}}=\frac{2}{\sigma^{2}}\end{aligned} \tag{14}$$

6 KLE

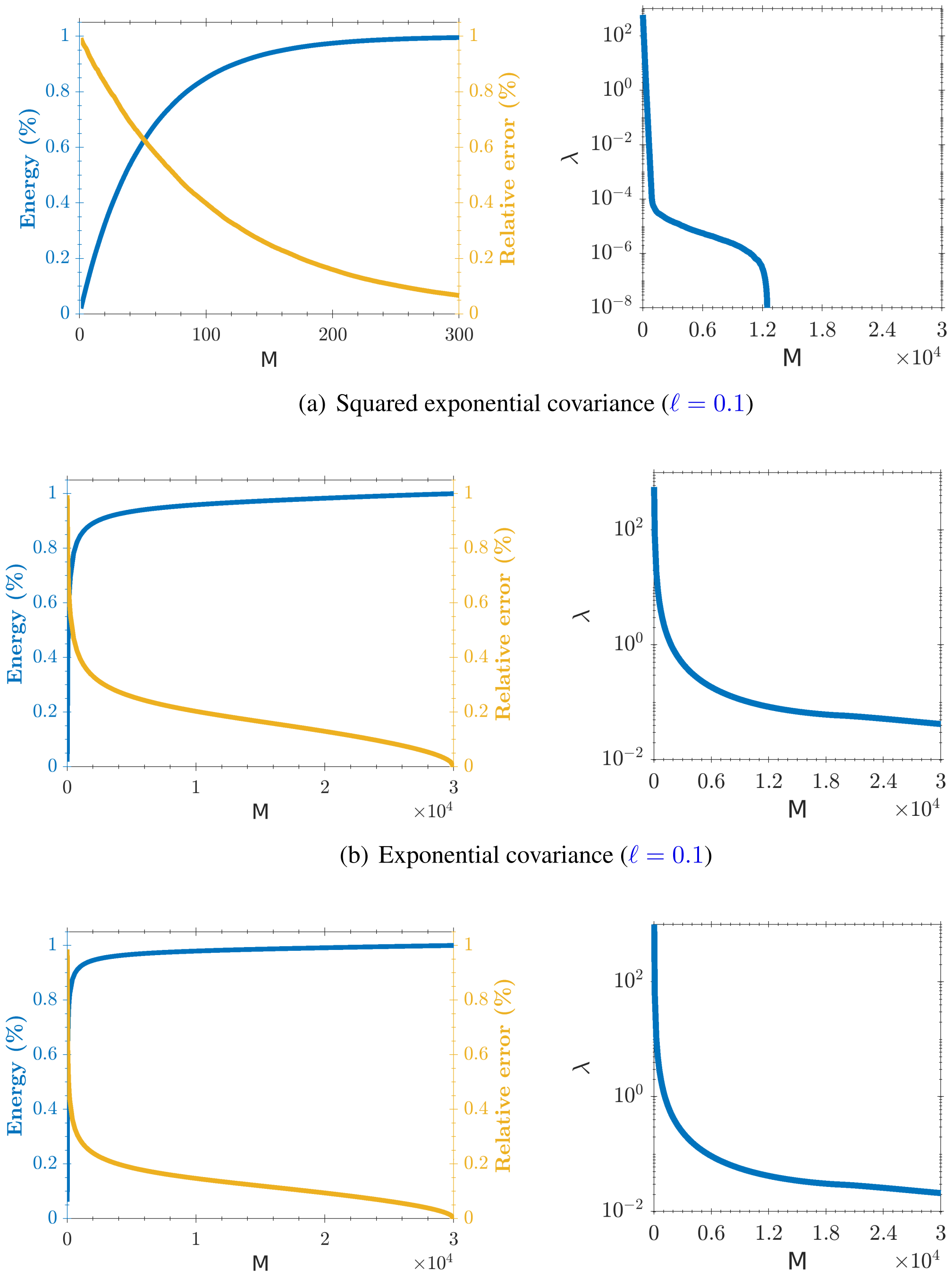
The eigenvalues associated with the eigenfunctions provide a measure of the energy contained in the respective mode. To quantify the energy involved in a truncated KL expansion (with M modes) we define the total relative energy as

$$\mathcal{E}_{\mathsf{M}}=\frac{\sum\limits_{j=1}^{\mathsf{M}}\lambda_{j}}{\sum\limits_{i=1}^{\infty}\lambda_{i}}, \tag{15}$$

in which the denominator will be approximated using $30,000$ eigenvalues ($\sum_{i=1}^{30,000}\lambda_{i}$).

6 KLE

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.

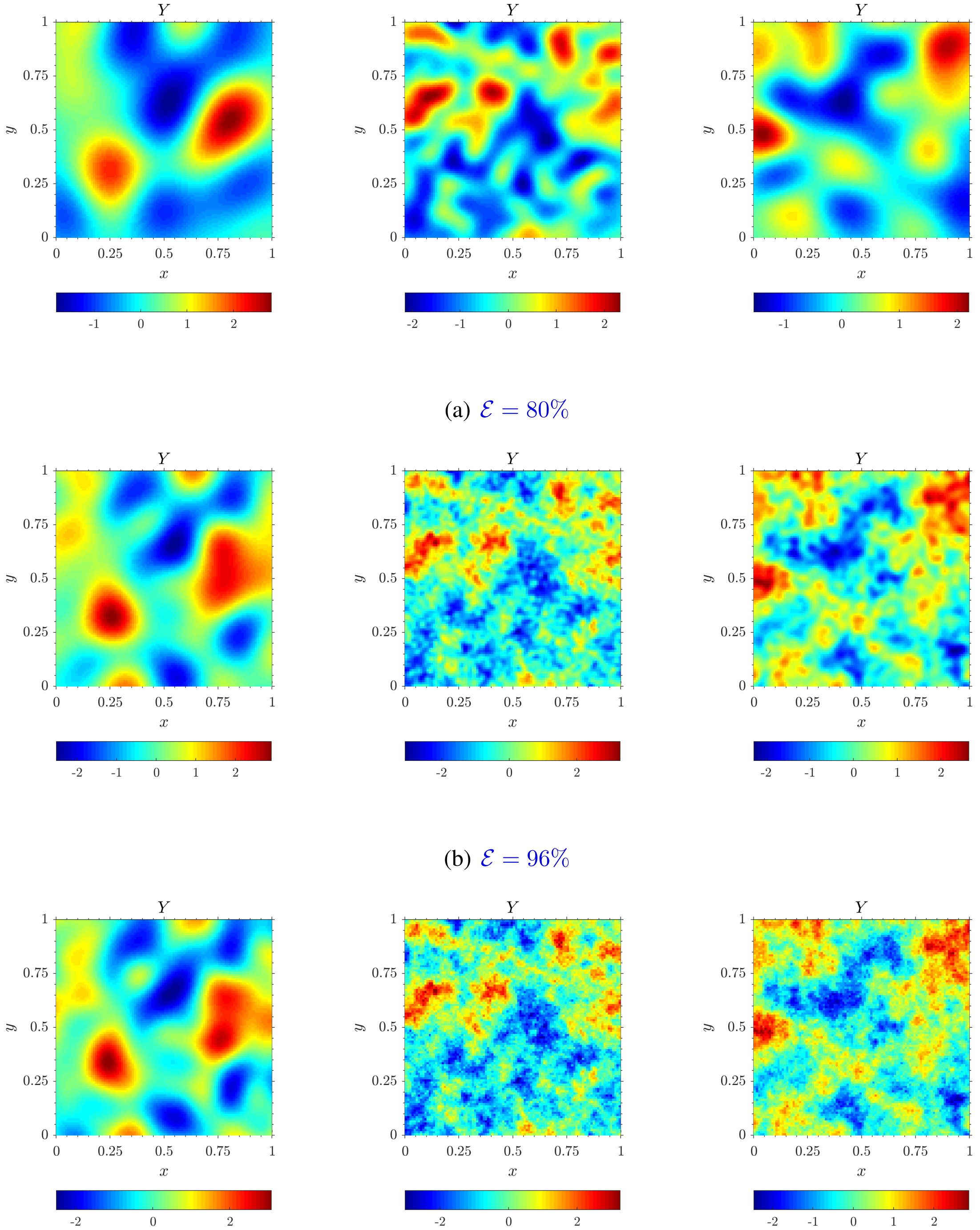


(a) Squared exponential covariance ($\ell = 0.1$)

(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.



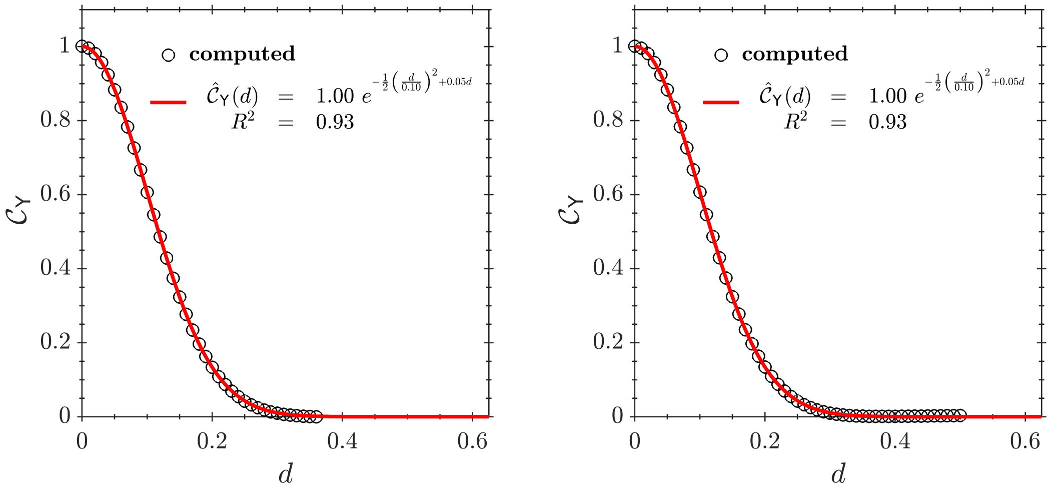
(a) $\mathcal{E} = 80\%$

(b) $\mathcal{E} = 96\%$

(c) $\mathcal{E} \sim 100\%$

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

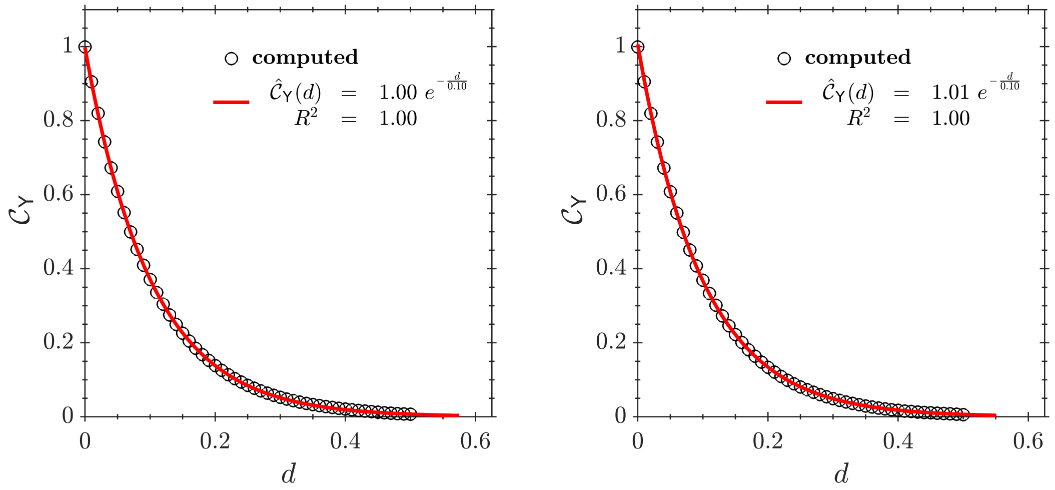
Energy	M			**Mean relative error**		
	Squared Exp.	Exponential	Exponential	Squared Exp.	Exponential	Exponential
(%)	($\ell = 0.1$)	($\ell = 0.1$)	($\ell = 0.2$)	($\ell = 0.1$)	($\ell = 0.1$)	($\ell = 0.2$)
80	84	592	154	0.46	0.45	0.47
90	122	2,323	612	0.32	0.32	0.33
94	150	5,760	1,655	0.25	0.25	0.25
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



(d) x direction

(e) y direction

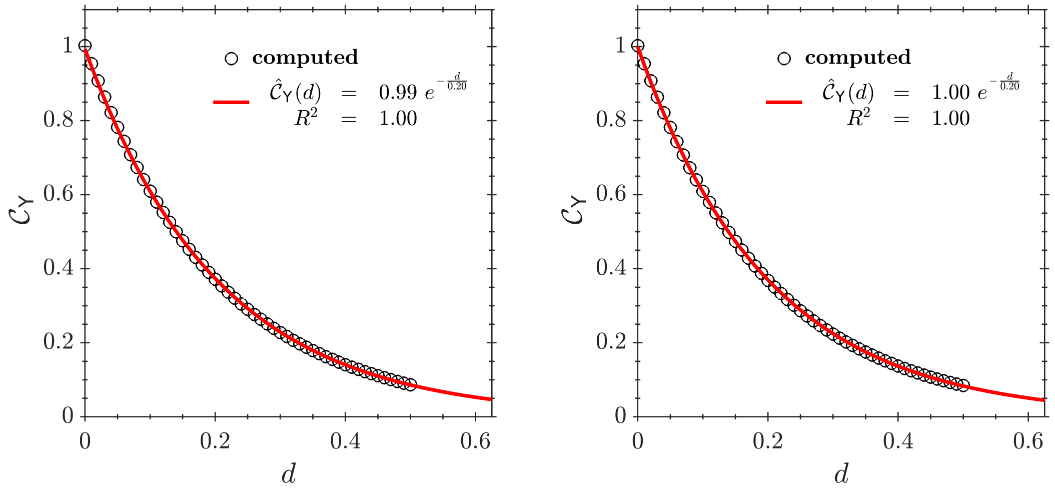
Figure 3: Estimated covariance in a sample of $5,000$ fields generate by KLE with $30,000$ terms. Squared exponential case with $\ell = 0.1$.



(a) x direction

(b) y direction

Figure 4: 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 5: Estimated covariance in a sample of $5,000$ fields generate by KLE with $30,000$ terms. Exponential case with $\ell = 0.2$.

REFERENCES

INDEX