Change of Measure for Bayesian Field Inversion with Hierarchical Hyperparameters Sampling

Nadège Polette, Olivier Le Maître, Pierre Sochala, Alexandrine Gesret

Change of Measure for Bayesian Field Inversion with Hierarchical Hyperparameters Sampling

Nadège Polette[a,b,*], Olivier Le Maître[c], Pierre Sochala[a], Alexandrine Gesret[b]

[a] *CEA, DAM, DIF, F-91297 Arpajon, France*
[b] *Mines Paris PSL, Geosciences center, 77300 Fontainebleau, France*
[c] *CMAP, CNRS, Inria, Ecole Polytechnique, IPP, 91120 Palaiseau, France*

Abstract

This paper proposes an effective treatment of hyperparameters in the Bayesian inference of a scalar field from indirect observations. Obtaining the joint posterior distribution of the field and its hyperparameters is challenging. The infinite dimensionality of the field requires a finite parametrization that usually involves hyperparameters to reflect the limited prior knowledge. In the present work, we consider a Karhunen-Loève (KL) decomposition for the random field and hyperparameters to account for the lack of prior knowledge of its autocovariance function. The hyperparameters must be inferred. To efficiently sample jointly the KL coordinates of the field and the autocovariance hyperparameters, we introduce a change of measure to reformulate the joint posterior distribution into a hierarchical Bayesian form. The likelihood depends only on the field's coordinates in a fixed KL basis, with a prior conditioned on the hyperparameters. We exploit this structure to derive an efficient Markov Chain Monte Carlo (MCMC) sampling scheme based on an adapted Metropolis–Hasting algorithm. We rely on surrogate models (Polynomial Chaos expansions) of the forward model predictions to further accelerate the MCMC sampling. A first application to a transient diffusion problem shows that our method is consistent with other approaches based on a change of coordinates (Sraj et al., 2016). A second application to a seismic traveltime tomography highlights the importance of inferring the hyperparameters. A third application to a 2D anisotropic groundwater flow problem illustrates the method on a more complex geometry.

Keywords: inverse problem, random fields, Karhunen–Loève decomposition, autocovariance function hyperparameters, polynomial chaos expansion

1. Introduction

Inverse problems arise in many situations whenever one searches for information about a physical system based on observations [1]. The Bayesian approach [2] is widely used to solve inverse problems in a probabilistic framework and has been applied in various geophysical applications [3, 4, 5, 6, 7]. The Bayesian approach is attractive because it requires a weak *a priori* on the unknown parameters and provides a full estimation of the parameter distributions. In practice, *Markov Chain Monte–Carlo* (MCMC) methods provide algorithms to sample from the posterior distribution [8]. MCMC methods require a considerable number of (generally expensive) forward model evaluations, and the convergence of the sampling might be difficult to reach, especially when dealing with a high-dimensional parameter space. The computational cost induced by the sampling can be reduced using surrogate models, where the forward model predictions are replaced by fast to evaluate approximations [9, 10, 4].

In this study, we are interested in the estimation of a scalar field through a set of indirect and noisy observations. Such inference problem is challenging because of the infinite dimensionality of the field. Previous

*Corresponding author
Email addresses: nadege.polette@minesparis.psl.eu (Nadège Polette), olivier.le-maitre@polytechnique.edu (Olivier Le Maître), pierre.sochala@cea.fr (Pierre Sochala), alexandrine.gesret@minesparis.psl.eu (Alexandrine Gesret)

works [9, 11] have already addressed this issue by means of *Karhunen–Loève* (KL) decomposition [12, 13, 14] to obtain a finite dimensional parametrization of the field. The KL expansion represents a random field using the eigenelements of its autocovariance function as a decomposition basis. The inference then consists in identifying the KL coordinates in the basis of the dominant eigenmodes. In practice, the autocovariance function depends on hyperparameters $\boldsymbol{q}$ that can be determined by optimizing a criterion such as the likelihood estimation [15], the leave-one-out error or by parametric estimation [16, 17]. Nevertheless, fixing a value of the hyperparameters can lead to overconfident results. This motivated the development of methods that jointly explore the parameters and hyperparameters spaces during the inverse problem solving. For stationnary fields, it is straightforward to deal with variable variances, as it reduces to a scaling of the KL coordinates [11, 18]. On the contrary, the other hyperparameters, like correlation length and exponent, are more delicate to infer since they affect directly the KL basis, raising significant computational challenges [19]. Several methods [20, 21, 22] have been proposed to deal with the computational cost induced by the eigenmodes basis $\boldsymbol{q}$-dependency. In particular, [21] introduced a method based on a fixed reference basis over the hyperparameters space. The reference basis is made of the dominant modes of the $\boldsymbol{q}$-averaged autocovariance function. The dependency on the hyperparameters is transferred to the coordinates through a linear transformation called *change of coordinates* (CoC). This construction minimizes the $\boldsymbol{q}$-averaged field representation error. Despite the benefit of the reference basis, the CoC method is limited to cases where the change of coordinates matrix is $\boldsymbol{q}$-continuous [21, 23]. Indeed, the computation of this matrix is computationally expensive since it requires solving an eigenvalue problem at each MCMC step. In addition, multiplicities in the eigenvalues and crossing of the eigenbranches with changes of the hyperparameters makes it difficult to ensure a smooth CoC as required by the MCMC samplers.

The aim of this paper is to introduce a new method to alleviate the difficulties of the CoC method by transferring the $\boldsymbol{q}$-dependency to the prior distribution of the field coordinates. In this approach, called *change of measure* (CoM), the coordinates of the field follow a Gaussian distribution whose covariance matrix depends smoothly on the hyperparameters and does not require to decompose the $\boldsymbol{q}$-dependent autocovariance function at each MCMC step. Another advantage of our formulation is that the posterior distribution only depends on $\boldsymbol{q}$ through the prior distribution. To accelerate the MCMC sampling, we rely on *polynomial chaos* (PC) surrogate models [24, 25, 26] for both model predictions and CoM derived quantities.

The paper is structured as follows. Section 2 presents the Bayesian framework and the field parametrization. Section 3 focuses on the derivation of the coordinates distribution in the CoM method, especially its covariance matrix expression and the associated sampling procedure. Section 4 describes the PC surrogate model used to replace the forward model, as well as the CoM derived quantities. The CoM method is compared with the CoC method on a transient diffusion case in Section 5 and applied to a traveltime seismic tomography problem in Section 6. Section 7 deals with the inference of a 2D anisotropic random field in a groundwater flow application. Conclusions and perspectives are drawn in Section 8.

2. Bayesian field inference

We detail the Bayesian framework in Section 2.1, and then we present the dimension reduction technique with the KL decomposition depending on hyperparameters in Section 2.2.

2.1. Bayes' formulation

We are interested in the inference of a scalar field $g \in \mathbb{G} \subset L^2(\Omega)$ defined over a compact spatial domain $\Omega \subset \mathbb{R}^d$, where $d \in \{1, 2, 3\}$ is the dimension of the physical space. Denoting $(\Theta, \mathfrak{S}, P)$ an abstract probability space and $\theta \in \Theta$ a particular random event, the field g is assumed to be a particular realization of a random process G, that is $g(\boldsymbol{x}) = G(\boldsymbol{x}, \theta)$.

The definition of the inference problem relies on a set of N indirect observations $\boldsymbol{d}_{\text{obs}} \in \mathbb{R}^N$, an observation noise model, as well as a forward model $M : g \mapsto \boldsymbol{d}$ that predict $\boldsymbol{d}$ for a given field g, where M is for instance a system of PDEs. This relation is not invertible in general and cannot be used to obtain a closed-form solution for g, such as $g \coloneqq M^{-1}\boldsymbol{d}_{\text{obs}}$ composed with an observation operator. The inverse problem can

be solved in several ways [1, 27] and we choose here the Bayesian approach, which characterizes the full posterior distribution of g. According to the Bayes' formula, the posterior distribution of the field g reads

$$\pi_{\text{post}}(g|\boldsymbol{d}_{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}_{\text{obs}}|g)\pi_{\mathbb{G}}(g), \tag{1}$$

where $\mathcal{L}(\boldsymbol{d}_{\text{obs}}|g)$ denotes the likelihood of the observations given the field g and $\pi_{\mathbb{G}}(g)$ its prior distribution. The posterior can be sampled with MCMC methods. The two main difficulties to use the posterior distribution (1) in practice are the infinite dimension of the field and the likelihood computation cost, based on forward model evaluations at each MCMC step. This motivates the use of a dimension reduction technique to approximate the field in a low dimensional space. Besides, the computational cost can be decreased with surrogate models that replace the forward model predictions.

2.2. Dimension reduction of random fields

This section provides a parametrization of the field g along with its associated sample space $\mathbb{G}$ and its prior probability distribution $\pi_{\mathbb{G}}$ for the Bayesian inference. Several approaches are possible to obtain finite dimensional representation of the field. The nodal representation relies on interpolation between values defined at a finite set of points constituting a grid. This representation is particularly expensive when a fine mesh is required to accomodate small scale features. Transdimensional approaches [28, 29, 30] have been developed to mitigate this issue by iteratively optimizing the number of mesh cells. These approches are however incompatible with the use of surrogate models since the field parametrization changes along the MCMC steps. We choose instead a modal representation of the field [14, 31, 32, 33] which offers a low-dimensional parametrization that is fixed during the inference.

Consider a random field $G(\boldsymbol{x}, \theta)$ with mean $\mu : \Omega \mapsto \mathbb{R}$ and autocovariance function $k : \Omega^2 \mapsto \mathbb{R}$, defined as

$$\mu(\boldsymbol{x}) \coloneqq \mathbb{E}_\Theta\left(G(\boldsymbol{x}, \cdot)\right) \quad \text{and} \quad k(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \mathbb{E}_\Theta\left(\left(G(\boldsymbol{x}, \cdot) - \mu(\boldsymbol{x})\right)\left(G(\boldsymbol{y}, \cdot) - \mu(\boldsymbol{y})\right)\right), \tag{2}$$

where $\mathbb{E}_\Theta$ denotes the expectation. Without loss of generality, we set $\mu = 0$ to alleviate notations. The truncated KL expansion G^r of G writes

$$G^r(\boldsymbol{x}, \theta) \coloneqq \sum_{i=1}^{r} \lambda_i^{1/2} u_i(\boldsymbol{x})\eta_i(\theta), \quad \text{such that} \quad G(\boldsymbol{x}, \theta) = \lim_{r\to\infty} G^r(\boldsymbol{x}, \theta). \tag{3}$$

The couples $(u_i, \lambda_i)_{i\in\mathbb{N}^*}$ are the eigenfunctions and associated eigenvalues (sorted in descending order) of the autocovariance function k, obtained by solving the Fredholm equation of the second kind,

$$\forall \boldsymbol{x} \in \Omega,\ \forall i \in \mathbb{N}^*, \qquad \int_\Omega k(\boldsymbol{x}, \boldsymbol{y})u_i(\boldsymbol{y})d\boldsymbol{y} = \lambda_i u_i(\boldsymbol{x}). \tag{4}$$

The KL decomposition is bi-orthonormal in the sense that the coordinates $\eta_i = \lambda^{-1/2}\left\langle G(\theta), u_i\right\rangle$ are uncorrelated with unitary variance and the eigenfunctions u_i are orthonormal with respect to the inner product of $L^2(\Omega)$,

$$\mathbb{E}(\eta_i\eta_j) = \delta_{i,j} \quad \text{and} \quad \langle u_i, u_j\rangle_\Omega \coloneqq \int_\Omega u_i(\boldsymbol{x})u_j(\boldsymbol{x})d\boldsymbol{x} = \delta_{i,j}. \tag{5}$$

In addition, the truncated decomposition minimizes the representation error in the L^2-sense [14]: $\mathbb{E}_\theta\left(\left\|G - G^r\right\|_{L^2(\Omega)}\right) = \sum_{i>r} \lambda_i$. The inference problem (1) for G^r can be recast into an identification of the r-dimensional coordinates vector $\boldsymbol{\eta}$,

$$\pi_{\text{post}}(\boldsymbol{\eta}|\boldsymbol{d}_{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}_{\text{obs}}|\boldsymbol{\eta})\pi(\boldsymbol{\eta}). \tag{6}$$

In the following, a Gaussian random process for the field is considered, and therefore $\boldsymbol{\eta} \sim \mathcal{N}(0, \mathrm{I}_r)$.

The autocovariance function k may depend on some hyperparameters $\boldsymbol{q} \in \mathbb{H}$ that are poorly known [20, 22]. In case of an uncertain correlation length, one can choose a small value of it, use a KL truncature with

many modes and then apply model reduction techniques [34, 35]. Another solution consists in integrating the hyperparameters in the inference problem. This yields

$$\pi_{\text{post}}\left(\boldsymbol{\eta}, \boldsymbol{q} \,|\boldsymbol{d}_{\text{obs}}\right) \propto \mathcal{L}(\boldsymbol{d}_{\text{obs}}|\boldsymbol{\eta}, \boldsymbol{q})\pi(\boldsymbol{\eta}, \boldsymbol{q}), \tag{7}$$

where $\pi(\boldsymbol{\eta}, \boldsymbol{q})$ is the joint prior distribution of the KL coordinates and hyperparameters. The KL expansion basis and the coordinates of the field G depend on the hyperparameters,

$$G(\boldsymbol{x}, \theta) \simeq \sum_{i=1}^{r} \lambda_i(\boldsymbol{q})^{1/2} u_i(\boldsymbol{x}, \boldsymbol{q})\eta_i(\theta), \quad \text{with} \quad \eta_i(\theta) = \left\langle \lambda_i(\boldsymbol{q})^{-1/2} u_i(\cdot, \boldsymbol{q}), G(\cdot, \theta) \right\rangle_{\Omega}. \tag{8}$$

The estimation of the coordinates and hyperparameters in decomposition (8) faces several difficulties, at the sampling stage as well as to construct surrogate models. First, the eigenelements must be computed at each MCMC step. Further, the $\boldsymbol{q}$-continuity of the field is not guaranteed due to the non uniqueness of the eigenelements, particularly in the case of crossing eigenbranches [23]. To get rid of this dependency in the field representation, [21] introduces a fixed reference basis to expand the field. The method proposed in this work follows this idea by expressing the hyperparameter dependencies through the prior of the coordinates in the reference basis.

3. Change of measure method

Sections 3.1 and 3.2 briefly describe the reference basis and the change of coordinates method used to estimate the KL coordinates. Section 3.3 presents the novel approach, called CoM, that overcomes the change of coordinates limitations. Section 3.4 details the log-posterior expression and the sampling strategy for the CoM.

3.1. Reference basis

The motivation of using a reference basis is to eliminate the eigenfunctions $\boldsymbol{q}$-dependency. Such basis could be computed from the kernel $k(\cdot, \cdot, \overline{\boldsymbol{q}})$ associated with a particular value of $\boldsymbol{q}$, its mean value $\overline{\boldsymbol{q}} = \mathbb{E}_{\mathbb{H}}(\boldsymbol{q})$ or from the kernel averaged over the hyperparameters domain $\mathbb{H}$. We opt for the latter choice because it minimizes the representation error of the field in average over $\boldsymbol{q}$ [21]. We denote $\overline{k}$ the averaged autocovariance function,

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \Omega, \qquad \overline{k}(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \mathbb{E}_{\mathbb{H}}(k(\boldsymbol{x}, \boldsymbol{y}, \cdot)) \coloneqq \int_{\mathbb{H}} k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q})\pi_{\mathbb{H}}(\boldsymbol{q})d\boldsymbol{q}, \tag{9}$$

where $\pi_{\mathbb{H}}$ denotes the hyperparameters prior. The reference eigenelements $\{\overline{u}_i, \overline{\lambda}_i\}_{i\in\mathbb{N}^*}$ are obtained by solving the eigenvalue problem for the reference kernel,

$$\forall \boldsymbol{x} \in \Omega, \ \forall i \in \mathbb{N}^*, \qquad \int_{\Omega} \overline{k}(\boldsymbol{x}, \boldsymbol{y})\overline{u}_i(\boldsymbol{y})d\boldsymbol{y} = \overline{\lambda}_i \overline{u}_i(\boldsymbol{x}). \tag{10}$$

The field approximation in the reference basis is

$$G(\boldsymbol{x}, \theta) \simeq \overline{G}^r(\boldsymbol{x}, \theta) \coloneqq \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x})\xi_i(\theta), \quad \text{with} \quad \xi_i(\theta) = \left\langle \overline{\lambda}_i^{-1/2} \overline{u}_i, G(\cdot, \theta) \right\rangle_{\Omega} \tag{11}$$

The following sections propose two methods to use this parametrization of G in the inference of the coordinates and hyperparameters joint distribution.

3.2. Change of coordinates

In [21], the coordinates $\boldsymbol{\xi}$ are obtained by a transformation of the coordinates in the $\boldsymbol{q}$-dependent basis defined previously in Eq. (8),

$$G^r(\boldsymbol{x},\theta) = \sum_{i=1}^{r} \lambda_i(\boldsymbol{q})^{1/2} u_i(\boldsymbol{x},\boldsymbol{q})\eta_i(\theta) \simeq \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x})\xi_i(\theta,\boldsymbol{q}). \tag{12}$$

The two sets of coordinates $\boldsymbol{\xi}(\theta,\boldsymbol{q})$ and $\boldsymbol{\eta}(\theta)$ are related by the change of coordinates matrix $B(\boldsymbol{q})$ defined by

$$B(\boldsymbol{q})_{ij} = \overline{\lambda}_i^{-1/2} \left\langle \lambda_j(\boldsymbol{q})^{1/2} u_j(\cdot,\boldsymbol{q}), \overline{u}_i \right\rangle_\Omega, \quad \boldsymbol{\xi}(\theta,\boldsymbol{q}) = B(\boldsymbol{q})\boldsymbol{\eta}(\theta). \tag{13}$$

The inference problem here writes

$$\pi_{\text{post}}\left(\boldsymbol{\eta},\boldsymbol{q}\,|\boldsymbol{d}_{\text{obs}}\right) \propto \mathcal{L}(\boldsymbol{d}_{\text{obs}}|\boldsymbol{\eta},\boldsymbol{q})\pi(\boldsymbol{\eta})\pi_{\mathbb{H}}(\boldsymbol{q}), \text{ with } \boldsymbol{\eta} \sim \mathcal{N}(0,\mathrm{I}_r). \tag{14}$$

In this formulation, for fixed $\boldsymbol{\eta}$, changing the hyperparameters modifies the coordinates in the reference basis and thus the shape of the field. Therefore, changing the hyperparameter modifies the realization of the process, which is not so natural. In addition, the posterior probability distribution of the coordinates $\boldsymbol{\eta}$ depends on the hyperparameters: for each value of $\boldsymbol{q}$, a different posterior distribution for $\boldsymbol{\eta}$ is sampled, which can complicate the inference process. Moreover, the dependency of B on the eigenelements $\{u_i(\cdot,\boldsymbol{q}),\lambda_i(\boldsymbol{q})\}_{1\leqslant i\leqslant r}$ requires to solve the eigenvalue problem (4) for each new value of $\boldsymbol{q}$. In [21, 23] a surrogate model is built for $B(\boldsymbol{q})$. However, this strategy suffers from the numerical ambiguity of $B(\boldsymbol{q})$ due to the choice of the eigenvectors orientation and indexation [23]. These drawbacks lead to reformulate the field parametrization thanks to a *change of measure* (CoM), where the hyperparameters values affects the prior of the proposed field and not the field values themselves.

3.3. Change of measure

We propose to use directly the decomposition on the reference basis regardless of the choice of hyperparameters; transferring the dependency on hyperparameters to the distribution of the coordinates. The inference problem has then the following general formulation,

$$\pi_{\text{post}}\left(\boldsymbol{\xi},\boldsymbol{q}\,|\boldsymbol{d}_{\text{obs}}\right) \propto \mathcal{L}(\boldsymbol{d}_{\text{obs}}|\boldsymbol{\xi})\pi(\boldsymbol{\xi}|\boldsymbol{q})\pi_{\mathbb{H}}(\boldsymbol{q}). \tag{15}$$

This method therefore generalizes the handling of the autocovariance scaling in [11] to the other hyperparameters. Note that, unlike the CoC method, the likelihood is independent of $\boldsymbol{q}$. The advantage is to work with the reference coordinates $\boldsymbol{\xi}$ only, without using the $\boldsymbol{q}$-dependent KL decomposition,

$$\overline{G}^r(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x})\xi_i, \text{ where } \boldsymbol{\xi} \sim \pi(\boldsymbol{\xi}|\boldsymbol{q}). \tag{16}$$

In this section, we derive the distribution for the $\boldsymbol{\xi} \coloneqq (\xi_i)_{1\leqslant i\leqslant r} \in \Xi \subset \mathbb{R}^r$ coordinates in order to best represent $G(\boldsymbol{x},\theta)$. The statistics of the new coordinates $\boldsymbol{\xi}$ are deduced by projecting the field on the reference basis. Assuming that $\overline{G}^r(\boldsymbol{x},\boldsymbol{\xi}) \simeq G(\boldsymbol{x},\theta)$, then $\left\langle \overline{G}^r(\cdot,\boldsymbol{\xi}), \overline{u}_i \right\rangle_\Omega = \langle G(\cdot,\theta), \overline{u}_i \rangle_\Omega$ and using Eq. (5) we obtain

$$\overline{\lambda}_i^{1/2}\xi_i = \sum_{k=1}^{+\infty} \lambda_k(\boldsymbol{q})^{1/2} \langle u_k(\cdot,\boldsymbol{q}), \overline{u}_i \rangle_\Omega \, \eta_k(\theta), \tag{17}$$

where the coordinates $(\eta_k)_{k\in\mathbb{N}^*}$ are standard normal variables. By construction, $\boldsymbol{\xi}$ is a multivariate Gaussian random vector. It is clear that the ξ_i have zero mean,

$$\mathbb{E}_\Theta(\xi_i) \simeq \overline{\lambda}_i^{-1/2} \sum_{k=1}^{+\infty} \lambda_k(\boldsymbol{q})^{1/2} \langle u_k(\cdot,\boldsymbol{q}), \overline{u}_i \rangle_\Omega \, \mathbb{E}_\Theta(\eta_k) = 0. \tag{18}$$

The covariance between ξ_i and ξ_j is

$$\begin{aligned}\mathbb{E}_\Theta(\xi_i\xi_j) &\simeq \mathbb{E}_\Theta\left(\left(\overline{\lambda}_i^{-1/2}\sum_{k=1}^{+\infty}\lambda_k(\boldsymbol{q})^{1/2}\left\langle u_k(\cdot,\boldsymbol{q}),\overline{u}_i\right\rangle_\Omega\eta_k\right)\left(\overline{\lambda}_j^{-1/2}\sum_{k=1}^{+\infty}\lambda_k(\boldsymbol{q})^{1/2}\left\langle u_k(\cdot,\boldsymbol{q}),\overline{u}_j\right\rangle_\Omega\eta_k\right)\right),\\ &= \left(\overline{\lambda}_i\overline{\lambda}_j\right)^{-1/2}\sum_{k,k'=1}^{+\infty}\left\langle u_k(\cdot,\boldsymbol{q}),\overline{u}_i\right\rangle_\Omega\left\langle u_{k'}(\cdot,\boldsymbol{q}),\overline{u}_j\right\rangle_\Omega\lambda_k(\boldsymbol{q})^{1/2}\lambda_{k'}(\boldsymbol{q})^{1/2}\mathbb{E}_\Theta\left(\eta_k\eta_{k'}\right),\\ &= \left(\overline{\lambda}_i\overline{\lambda}_j\right)^{-1/2}\left\langle\left\langle k(\cdot,\cdot,\boldsymbol{q}),\overline{u}_i\right\rangle_\Omega,\overline{u}_j\right\rangle_\Omega, \end{aligned} \tag{19}$$

where the last line is obtained thanks to Mercer's theorem [36]. The reference coordinates $\boldsymbol{\xi}$ follow a prior normal distribution $\mathcal{N}(0,\Sigma(\boldsymbol{q}))$, where the covariance matrix $\Sigma(\boldsymbol{q})\in\mathbb{R}^{r\times r}$ is defined by

$$\forall 1\leqslant i,j\leqslant r,\ \forall\boldsymbol{q}\in\mathbb{H},\qquad \Sigma(\boldsymbol{q})_{ij}=\left(\overline{\lambda}_i\overline{\lambda}_j\right)^{-1/2}\left\langle\left\langle k(\cdot,\cdot,\boldsymbol{q}),\overline{u}_j\right\rangle_\Omega,\overline{u}_i\right\rangle_\Omega. \tag{20}$$

The covariance of the coordinates $\boldsymbol{\xi}$ for a given $\boldsymbol{q}$ is obtained by projecting the $\boldsymbol{q}$-dependent autocovariance on the span of the r-dimensional reference basis. This covariance matrix is invertible if $k(\boldsymbol{q})$ is positive definite, which is the case in the following applications. However, note that the conditionning of $\Sigma(\boldsymbol{q})$ can be quite bad if r is very large.

The principles of the CoC and CoM methods are schematically represented on Fig. 1. For both methods, the model predictions only depends on the coordinates $\boldsymbol{\xi}$ in the reference basis. The main advantage of the CoM method is to split the posterior construction in two distincts parts: i) the computation of the likelihood $\mathcal{L}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\xi})$ which is independent from the hyperparameters $\boldsymbol{q}$ and ii) the computation of the prior probability distribution $\pi(\boldsymbol{\xi}|\boldsymbol{q})\pi_{\mathbb{H}}(\boldsymbol{q})$ with the covariance matrix $\Sigma(\boldsymbol{q})$. As the covariance matrix $\Sigma(\boldsymbol{q})$ only depends on the reference eigenelements which are computed offline, its computation is less expensive than the computation of $B(\boldsymbol{q})$ which requires to solve an eigenvalue problem at each step. In addition, the approximation of $\Sigma(\boldsymbol{q})$ by a surrogate model benefits from the regularity of the $\boldsymbol{q}$-dependent autocovariance function. The CoM representation also provides a better physical interpretation in comparison with the CoC one. Indeed, the CoM approach better distinguishes the coordinates from the hyperparameters, in the sense that changing the hyperparameters does not modify the field realization but its probability.

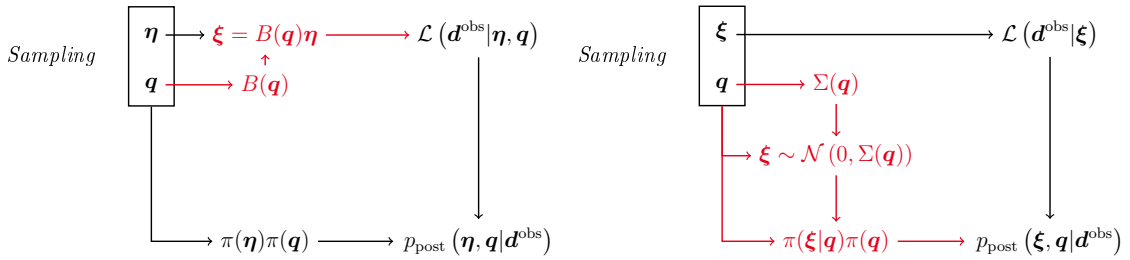


Figure 1: Workflow of the CoC (left) and the CoM (right) methods. The quantities that are specific to the method are indicated in red.

3.4. MCMC for the change of measure

The logarithm of the posterior probability distribution is written as

$$\log\pi_{\text{post}}(\boldsymbol{\xi},\boldsymbol{q}|\boldsymbol{d}_{\text{obs}})\propto\log\mathcal{L}(\boldsymbol{d}_{\text{obs}}|\boldsymbol{\xi})+\log\pi(\boldsymbol{\xi}|\boldsymbol{q})+\log\pi_{\mathbb{H}}(\boldsymbol{q}). \tag{21}$$

A Gaussian likelihood is considered, assuming a centered *independent and identically distributed* (i.i.d.) observation noise, with variance σ_ε^2: the noise level σ_ε is the standard deviation of the observation noise $\varepsilon=\boldsymbol{d}_{\text{obs}}-M(f_{\text{true}})$ and is assumed to be i.i.d. Gaussian,

$$\varepsilon\sim\mathcal{N}(0,\sigma_\varepsilon^2\mathrm{I}_N), \tag{22}$$

where I_N is the identity matrix of size N. The likelihood writes

$$\mathcal{L}(\boldsymbol{d}_{\text{obs}}|\boldsymbol{\xi}) \coloneqq \frac{1}{\sqrt{(2\pi\sigma_\varepsilon^2)^N}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2}\left\|\boldsymbol{d}_{\text{obs}} - M(\boldsymbol{\xi})\right\|_{l_2}^2\right). \tag{23}$$

Other likelihood functions and error models can be used [37, 38]. Since $\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma(\boldsymbol{q}))$, the prior log-density of the random vector $\boldsymbol{\xi}$ is

$$\log \pi(\boldsymbol{\xi}|\boldsymbol{q}) \propto -\frac{1}{2}\left(\text{logdet}_\Sigma(\boldsymbol{q}) + \boldsymbol{\xi}(\boldsymbol{q})^\top \Sigma(\boldsymbol{q})^{-1}\boldsymbol{\xi}(\boldsymbol{q})\right), \tag{24}$$

where $\text{logdet}_\Sigma(\boldsymbol{q}) \coloneqq \log\left(\det\left[\Sigma(\boldsymbol{q})\right]\right)$. The noise level σ_ε is estimated during the inference, assuming a Jeffreys prior [39, 40, 41], $\pi_{\sigma_\varepsilon}(\sigma_\varepsilon) \propto 1/\sigma_\varepsilon$.

The CoM posterior has a hierarchical structure [42]. In particular, since $\pi(\boldsymbol{\xi}|\boldsymbol{q})$ can be highly sensitive to $\boldsymbol{q}$, it can be difficult to define a proposal structure suitable over the whole sampling space. The *Metropolis–Hastings* (MH) algorithm, even in its adaptive version [43, 44], can be challenged by the CoM hierarchical structure without a proposal distribution π_{tr} depending on the current state of the chain $(\boldsymbol{\xi}^{(n)}, \boldsymbol{q}^{(n)})$. Our strategy to overcome this difficulty consists in sampling an auxiliary variable $\overline{\boldsymbol{\xi}}$ independent from $\boldsymbol{q}$, using a fixed proposal structure and a $\boldsymbol{q}$-dependent linear transformation to obtain $\boldsymbol{\xi}$. Specifically, we consider $\overline{\boldsymbol{\xi}} \sim \mathcal{N}(0, \Sigma_{\overline{\boldsymbol{\xi}}})$ and set $\boldsymbol{\xi}$ through

$$\boldsymbol{\xi} = \Sigma(\boldsymbol{q})^{1/2}\Sigma_{\overline{\boldsymbol{\xi}}}^{-1/2}\overline{\boldsymbol{\xi}}. \tag{25}$$

At each MCMC step, the MH acceptance criterion is used. Denoting $\boldsymbol{Y}^{(n)} = (\boldsymbol{\xi}^{(n)}, \boldsymbol{q}^{(n)}, \sigma_\varepsilon^{(n)})$ the current state and $\boldsymbol{Y}^* = (\boldsymbol{\xi}^*, \boldsymbol{q}^*, \sigma_\varepsilon^*)$ the proposed one, the proposition is accepted with probability

$$p_{\text{MH}} = \min\left(\frac{\pi_{\text{post}}(\boldsymbol{Y}^*)\pi_{\text{tr}}(\boldsymbol{Y}^{(n)}|\boldsymbol{Y}^*)}{\pi_{\text{post}}(\boldsymbol{Y}^{(n)})\pi_{\text{tr}}(\boldsymbol{Y}^*|\boldsymbol{Y}^{(n)})}, 1\right), \tag{26}$$

where $\pi_{\text{tr}}(\boldsymbol{Y}|\boldsymbol{X})$ is the transition probability from $\boldsymbol{X}$ to $\boldsymbol{Y}$. In our case, this transition probability is not symmetric,

$$\frac{\pi_{\text{tr}}(\boldsymbol{Y}^{(n)}|\boldsymbol{Y}^*)}{\pi_{\text{tr}}(\boldsymbol{Y}^*|\boldsymbol{Y}^{(n)})} = \left(\frac{\det\Sigma(\boldsymbol{q}^*)}{\det\Sigma(\boldsymbol{q}^{(n)})}\right)^{1/2}. \tag{27}$$

Note that $\det\Sigma(\boldsymbol{q})$ is computed for the prior, such that the evaluation of the transition probability ratio add no computational cost. Further, as $\Sigma(\boldsymbol{q})$, $\Sigma(\boldsymbol{q})^{1/2}$ is independent of the indexation and orientation of the eigenelements.

4. Accelerating sampling in the CoM framework

The Bayesian inference of $\overline{G}^r(\boldsymbol{x}, \boldsymbol{\xi})$ with $\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma(\boldsymbol{q}))$ requires the sampling of the $r + |q|$-dimensional space of $(\boldsymbol{\xi}, \boldsymbol{q}) \in \Xi \times \mathbb{H}$ with a MCMC algorithm. As previously mentioned, one advantage of the CoM is that the predictions, and therefore the likelihood, only depends on the coordinates $\boldsymbol{\xi}$. In this section, we briefly present PC expansions used to replace the set of forward model predictions. The coefficients and determinant of the CoM matrix used to define the likelihood function are also approximated with PC expansions.

4.1. Polynomial chaos surrogates

Surrogate models are widely used in uncertainty propagation and Bayesian inference [45, 46]. We rely on PC expansions [24, 25, 26] which have been used in various fields [11, 19, 4, 47, 48, 18, 49]. Let f be a second order functional of a r-dimensional vector $\boldsymbol{\zeta}$ with independent components such that $p_{\boldsymbol{\zeta}}(\boldsymbol{y}) = \prod_{i=1}^{r} p_i(y_i)$. The PC expansion $\widetilde{f}$ of f is a linear combination of polynomials in $\boldsymbol{\zeta}$,

$$f(\boldsymbol{\zeta}) \simeq \widetilde{f}(\boldsymbol{\zeta}) = \sum_{a\in\mathcal{A}} f_a \Psi_a(\boldsymbol{\zeta}), \tag{28}$$

where the PC basis functions $\{\Psi_a(\boldsymbol{\zeta})\}_{a\in\mathcal{A}}$ are multivariate orthogonal polynomials with respect to the density of $\boldsymbol{\zeta}$, $\{f_a\}_{a\in\mathcal{A}}$ are deterministic PC coefficients and $\mathcal{A}$ is a set of multi-indexes. The PC basis functions are the product of orthonormal univariate polynomials, $\Psi_{a=(a_1,\dots a_d)}(\boldsymbol{\zeta}) = \Pi_i \Psi^i_{a_i}(\boldsymbol{\zeta}_i)$, where $\Psi^i_{a_i}$ is a polynomial of degree a_i in $\boldsymbol{\zeta}_i$. Several methods can be implemented to estimate the PC coefficients. In this work, we use the *pseudo spectral projection* (PSP) method to approximate the projection coefficients $\{f_a\}_{a\in\mathcal{A}}$ of $f(\boldsymbol{\zeta})$ [50, 51]. The PSP method relies on sequences of nested 1-dimensional discrete projection operators into spaces of increasing polynomial degrees. The 1-d projections are sparsely tensorized, using the Smolyak formula. The PSP yields sparse projection operators, free of internal aliasing for $f \in \mathrm{Span}\{\Psi_a,\ a \in |\mathcal{A}|\}$, where $\mathcal{A}$ depends on the composing sequences and their retained tensorizations. In practice, the PSP method uses evaluations of f at the nodes $\boldsymbol{\zeta}^{(l)}$ of a sparse grid $S = \{\boldsymbol{\zeta}^{(k)}\}_{1\leqslant k\leqslant N_{\mathrm{PSP}}}$ and computes the PC coefficients f_a through

$$\forall a \in \mathcal{A}, \quad f_a = \sum_{k=1}^{N_{\mathrm{PSP}}} \Pi_{ak} f(\boldsymbol{\zeta}^{(k)}). \tag{29}$$

The surrogate model accuracy is assessed by estimating the *relative root mean squared error* (RRMSE) with a validation set of N_v samples $\mathcal{Z} = \left\{\boldsymbol{\zeta}^{(i)}\right\}_{1\leqslant i\leqslant N_v}$ drawn from $p_{\boldsymbol{\zeta}}$,

$$\mathrm{RRMSE}(f,\widetilde{f}) = \sqrt{\sum_{\boldsymbol{\zeta}\in\mathcal{Z}} \left\| f(\boldsymbol{\zeta}) - \widetilde{f}(\boldsymbol{\zeta})\right\|^2 / \sum_{\boldsymbol{\zeta}\in\mathcal{Z}} \|f(\boldsymbol{\zeta})\|^2}. \tag{30}$$

4.2. Surrogate models for likelihood computation

As seen in Sections 3.3 and 3.4, the model predictions in the likelihood only depend on the coordinates of the field $\boldsymbol{\xi}$. Therefore, we construct a surrogate model of the prediction vector $\boldsymbol{d}(\boldsymbol{\xi})$ with the PSP method:

$$\boldsymbol{d}(\boldsymbol{\xi}) \simeq \sum_{a\in\mathcal{A}} d_a \Psi_a(\boldsymbol{\xi}). \tag{31}$$

Recall that the coordinates $\boldsymbol{\xi}$ follow $\mathcal{N}(0,\Sigma(\boldsymbol{q}))$ in the CoM method. To avoid the $\boldsymbol{q}$-dependency of $\boldsymbol{\xi}$, we consider that $\boldsymbol{\xi} \sim \mathcal{N}(0,\mathrm{I}_r)$ for the PC surrogate construction. This choice corresponds to a simple approximation of the marginal coordinates distribution over the hyperparametric domain since $\mathbb{E}_{\mathbb{H}}\left(\mathbb{E}_{\Theta}(\boldsymbol{\xi}\,|\boldsymbol{q})\right) = 0$ and $\mathbb{E}_{\mathbb{H}}(\Sigma(\cdot)) = \mathrm{I}_r$ from equations (10),(20). Note that this is not the true marginal coordinates distribution, since the sum of Gaussian distributions is not Gaussian (see AppendixA for more details). The model predictions at the PSP points are computed by i) building the field associated with the PSP point $g^{(l)}(\boldsymbol{x}) = \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} u_i(\boldsymbol{x}) \xi_i^{(l)}$, ii) solving the forward model with $g^{(l)}$ and evaluate $\boldsymbol{d}^{(l)} = d(\boldsymbol{\xi}^{(l)})$.

4.3. Surrogate models for prior computation

The CoM formulation and its sampling require the computation of $\Sigma(\boldsymbol{q})^{-1}$, $\mathrm{logdet}_{\Sigma}(\boldsymbol{q})$, and $\Sigma(\boldsymbol{q})^{1/2}$ at each step. We construct surrogate models of these three quantities to accelerate the MCMC sampling. To ensure that the approximated $\Sigma(\boldsymbol{q})^{-1}$ is non negative, we approximate its square root. To ensure the positivity of the $\Sigma(\boldsymbol{q})^{-1}$ surrogate, we adopt the approach of [52] for the PC approximations of semi positive definite operators in the context of domain decomposition methods for stochastic partial differential equations. The approach is based on the spectral decomposition of $\Sigma(\boldsymbol{q})$ in

$$\Sigma(\boldsymbol{q}) = U(\boldsymbol{q})\Lambda(\boldsymbol{q})U(\boldsymbol{q})^{\top},\ \text{with } U(\boldsymbol{q})U(\boldsymbol{q})^{\top} = U(\boldsymbol{q})^{\top}U(\boldsymbol{q}) = \mathrm{I}_r, \tag{32}$$

and defines

$$\Sigma(\boldsymbol{q})^{\pm 1/2} \coloneqq U(\boldsymbol{q})\Lambda(\boldsymbol{q})^{\pm 1/2}U(\boldsymbol{q})^{\top}. \tag{33}$$

The advantage of this approach is that the PSP approximation

$$\widetilde{\Sigma}^{\pm 1/2} = \sum_{a\in\mathcal{A}} \Sigma_a^{\pm 1/2} \Psi_a(\boldsymbol{q}),\ \text{with } \Sigma_a^{\pm 1/2} = \sum_{k=1}^{N_{\mathrm{PSP}}} \Pi_{ak} \Sigma(\boldsymbol{q}^{(k)})^{\pm 1/2} \tag{34}$$

is not affected by the indexation and orientation of the $\Sigma(\boldsymbol{q}^{(k)})$ eigenelements at the different *sparse grid* (SG) points. Finally, we use

$$\widetilde{\Sigma}^{-1}(\boldsymbol{q}) \coloneqq \widetilde{\Sigma}^{-1/2}\widetilde{\Sigma}^{-1/2}. \tag{35}$$

In addition, the PSP approximation of $\mathrm{logdet}_\Sigma(\boldsymbol{q})$ writes as

$$\widetilde{\mathrm{logdet}}_\Sigma(\boldsymbol{q}) = \sum_{a\in\mathcal{A}} l_a \Psi_a(\boldsymbol{q}), \text{ with } l_a = \sum_{k=1}^{N_{\mathrm{PSP}}} \Pi_{ak} \mathrm{logdet}_\Sigma(\boldsymbol{q}^{(k)}). \tag{36}$$

When the prior range is large and the number of modes r is high, the prior $\pi(\boldsymbol{\xi}|\boldsymbol{q})$ can become highly stretched with respect to $\boldsymbol{q}$. In that case, the PC approximation of the conditional law can be difficult and requires a high PSP level. In practice, we adapt the PSP method level to ensure a RRMSE of less than 0.1% on the surrogates of $\Sigma(\boldsymbol{q})^{-1}$ and $\mathrm{logdet}_\Sigma(\boldsymbol{q})$. As for the surrogate predictions, the PC error should be dominated by the observation error. This is verified *a posteriori*, and the analysis is led with a higher PSP level if needed.

5. Application to a transient diffusion problem

In this section, the CoM method is applied to infer a diffusivity field in the one-dimensional *transient diffusion* (TD) problem presented in [21]. This problem is summarized in Section 5.1, the behavior and convergence of the PC surrogates of $\Sigma^{\pm 1/2}$ and logdet_Σ are analysed in Section 5.2 and the inference results are presented in Section 5.3, where a comparison with the CoC method is proposed.

5.1. Case presentation

The 1D TD equation writes

$$\forall t \in (0, T = 0.05),\ \boldsymbol{x} \in \Omega = (0,1), \quad \frac{\partial U(\boldsymbol{x},t)}{\partial t} = \frac{\partial}{\partial \boldsymbol{x}}\left(\nu(\boldsymbol{x})\frac{\partial U(\boldsymbol{x},t)}{\partial \boldsymbol{x}}\right), \tag{37}$$

with the boundary and initial conditions,

$$U(\boldsymbol{x}=0,t) = -1, \quad U(\boldsymbol{x}=1,t) = 1, \quad \text{and} \quad U(\boldsymbol{x},t=0) = 0. \tag{38}$$

The field ν is such that $0 < \nu(\boldsymbol{x}) < +\infty$ almost everywhere in Ω, is unknown and to be inferred. For a given ν, the transient diffusion equation is solved with a $\mathbb{P}_1$-finite element method in space and a second order implicit time-integration scheme. The objective of the inference is to learn the diffusitivity field ν from observations of U. The observations consist in $N = N_x \times N_t = 18 \times 13 = 234$ noisy evaluations of U, uniformly distributed in space and time. They are obtained by solving Pb. (37) and then synthetically corrupted by random independent Gaussian centered measurement noises $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 \coloneqq 0.01)$. For the inference, we consider *a priori* a log-normal stationnary field ν, and then infer $g \coloneqq \log \nu$ using a Gaussian prior with zero mean and with a squared exponential autocovariance function $k(\boldsymbol{q})$,

$$\forall (\boldsymbol{x},\boldsymbol{y}) \in \Omega^2, \quad \boldsymbol{q} = \{A, l\} \in \mathbb{R}_+^2, \quad k(\boldsymbol{x},\boldsymbol{y},\boldsymbol{q}) \coloneqq A \exp\left(-\|\boldsymbol{x}-\boldsymbol{y}\|^2/(2l^2)\right), \tag{39}$$

where A denotes the amplitude and l the correlation length. The prior distributions of these two hyperparameters are

$$A \sim \mathrm{InvGamma}(3,1) \quad \text{and} \quad l \sim \log\text{-}\mathcal{U}(0.1, 0.7). \tag{40}$$

The decomposition in the reference basis shown in Eq. (11)leads to the following approximation,

$$g_{\mathrm{true}} \simeq g_{\mathrm{kl}}(\boldsymbol{\xi}) = \overline{G}^r(\cdot,\boldsymbol{\xi}), \text{ with } \xi_i = \left\langle \overline{\lambda}_i^{-1/2}\overline{u}_i, g_{\mathrm{true}} \right\rangle_\Omega. \tag{41}$$

The number of coordinates r is set equal to 8 and captures 99.8% of the field prior variance.

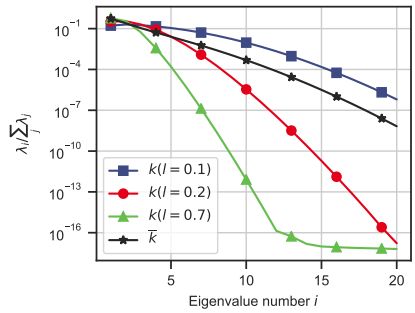


Figure 2: TD case - Eigenvalues decay for different correlation lengths l as indicated.

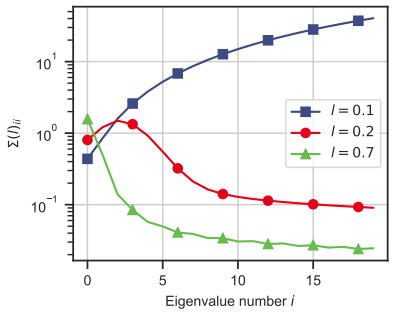


Figure 3: TD case - Diagonal terms of $\Sigma(l)$ corresponding to the prior variance of $\boldsymbol{\xi}$.

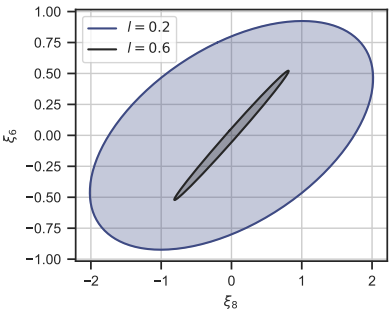


Figure 4: TD case - Prior Covariance $\Sigma(l)$ for two correlation lengths. Projection on the $6th$ and $8th$ coordinates.

r	l_-	l_+	$\sum_{i=1}^{r}\overline{\lambda}_i / \sum_{i=1}^{+\infty}\overline{\lambda}_i$	$\left\|\Sigma(\boldsymbol{q})^{-1}\right\|_{l_+^2}$	RRMSE$(\Sigma(l)^{-1})$
8	0.1	0.7	99.8%	6.7×10^{3}	1.0×10^{-3}
8	0.05	0.7	97.1%	3.3×10^{4}	6.4×10^{-3}
8	0.1	0.9	99.8%	1.1×10^{4}	1.5×10^{-3}
6	0.1	0.7	99.0%	6.4×10^{3}	4.5×10^{-4}

Table 1: TD case - Quantities of interest for different numbers of coordinates and prior ranges. RRMSE is computed for PC order equal to 15.

5.2. Behavior of the change of measure

The aim of this section is to analyze the behavior of the CoM's surrogates for the TD case. It focuses on the analysis of $\Sigma(\boldsymbol{q})$ and its derived quantities $\Sigma^{\pm 1/2}(\boldsymbol{q})$ and $\text{logdet}_{\Sigma}(\boldsymbol{q})$ when l varies, since the A-dependency is reduced to a multiplicative factor.

The eigenvalues of the covariance for different correlation lengths are plotted on Fig. 2: the higher the correlation length, the faster the decay. In particular, for large correlation lengths, the information of the l-dependent basis is concentrated in the few first modes. The spectrum of the averaged covariance (*i.e.* leading to the reference basis) is intermediate. Figures 3 and 4 show how changes in l affect the prior coordinates $\boldsymbol{\xi}$. Figure 3 reports the diagonal coefficients $\Sigma_{ii}(\boldsymbol{q})$ for three values of l: the higher the correlation length, the smaller the variance of the last coordinates. A coordinate variance is close to zero when the associated mode is irrelevant for the approximation of the prior field. Figure 4 shows the resulting 90% confidence interval for ξ_6 and ξ_8 and two values of l, highlighting the strong dependency of $\Sigma(\boldsymbol{q})$ on l. The CoM induces a high sensitivity of the $\boldsymbol{\xi}$ prior distribution to l. Tab. 1 highlights the influence of the prior range and the number of coordinates on the eigenvalues and covariance matrix of the CoM method. Two conclusions can be drawn: i) diminishing r increases the error on the field prior, and diminishing l_- builds a reference basis that is closer to small correlation length bases, each effect increasing the truncation error and ii) increasing the distance between l_- and l_+ or the number of terms r yields a higher l^2-norm. The numerical invertibility of the covariance matrix is compromised when the hyperparameter space or the number of coordinates is too large. Finally, the RRMSE of $\Sigma^{1/2}, \Sigma^{-1}$ and logdet_{Σ} are plotted on Fig. 5 for different PC order and number of terms. For $r = 8$, a PC order of 15 ensures that the error is below 0.1% for the three quantities. A Tikhonov regularization is implemented in order to prevent numerical instability when increasing the number of terms.

5.3. Inference results

We test the CoM for two different true log-diffusivity fields, namely i) a sinusoidal profile: $g^{\text{sin}}(\boldsymbol{x}) = \sin(2\pi\boldsymbol{x})$ and ii) a step function: $g^{\text{step}}(\boldsymbol{x}) = \begin{cases} -1/2, \text{ if } \boldsymbol{x} < 0.5, \\ 1/2 \text{ else} \end{cases}$. The projections of the true fields in

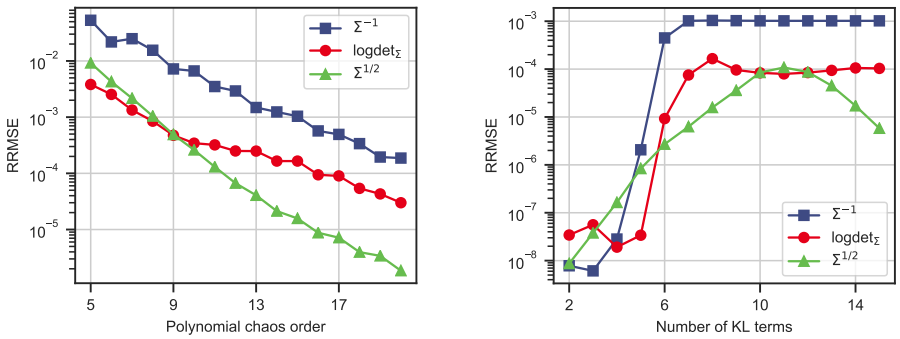


Figure 5: TD case - RRMSE values (left) according PC order ($r = 8$), (right) according number of terms (PC order = 15). $1,000$ random validation points.

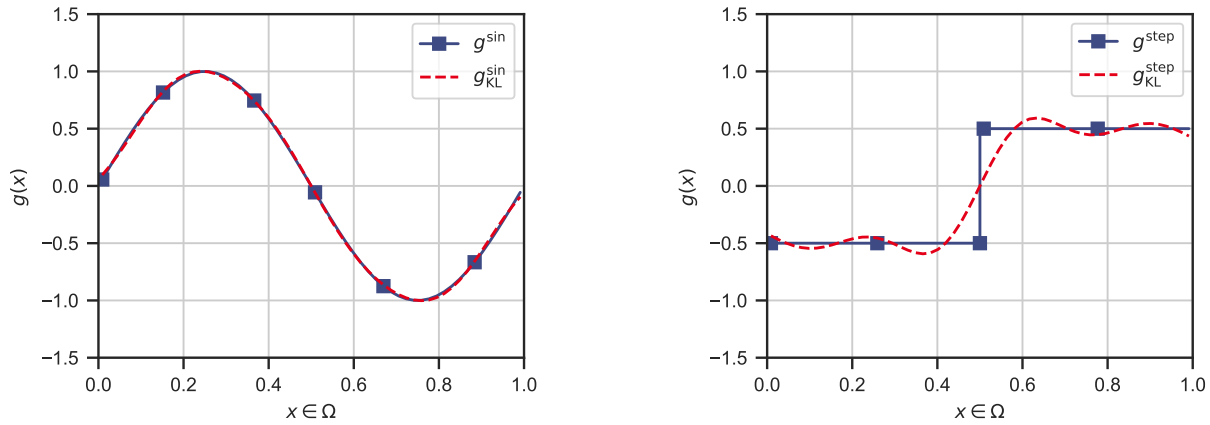


Figure 6: TD case - True diffusivity fields g_{true} and their projections g_{kl}. (left) g^{sin}, (right) g^{step}

the reference basis (Eq. (41)) are shown on Fig. 6. As expected, the field g^{sin} is well approximated in the reference basis whereas the field g^{step} is considerably smoothed and presents oscillations, because this non-smooth field is not a likely realization from the prior.

We use the MCMC presented in Section 3.4 to sample the posterior distribution. The CoC method introduced in Section 3.2 is also implemented for validation purposes. The proposal matrix is adapted each $2.5{\times}10^4$ steps during a burning phase of $2.5{\times}10^5$ steps, then $1{\times}10^6$ steps are used to sample the posterior distribution. This choice leads to a multi effective sample size [53] around $10,000$ for both CoM and CoC methods. We start the analysis by examining the field coordinates. The marginal posterior distributions of five coordinates are plotted on Fig. 7 for the g^{sin} field. For the comparison, the CoC samples $(\boldsymbol{\eta}, \boldsymbol{q})$ are translated into samples of $\boldsymbol{\xi}$ using the CoC $\boldsymbol{\xi} = B(\boldsymbol{q})\boldsymbol{\eta}$. The $\boldsymbol{\xi}$ marginal prior distribution $\pi(\boldsymbol{\xi}) = \int_{\mathbb{H}} \pi(\boldsymbol{\xi}|\boldsymbol{q})\pi_{\mathbb{H}}(\boldsymbol{q})d\boldsymbol{q}$ is also plotted as reference. We observe that the two methods yield similar results, with the posterior distributions for the first 5 coordinates centered on the best projections of the true field. The other coordinates are less informed by the observations. The noise level posterior distribution is peaked around the true value.

Figure 8 presents the posterior distributions confidence intervals of the inferred fields. The mean, median, *maximum a posteriori* (MAP) as well as the 1%-99% and 5%-95% quantiles are plotted. The MAP corresponds here to the chain sample maximizing the posterior distribution. As for the coordinates estimations, the results of the two methods are very close for g^{sin} and g^{step}. Note that the g^{sin} medians are shifted

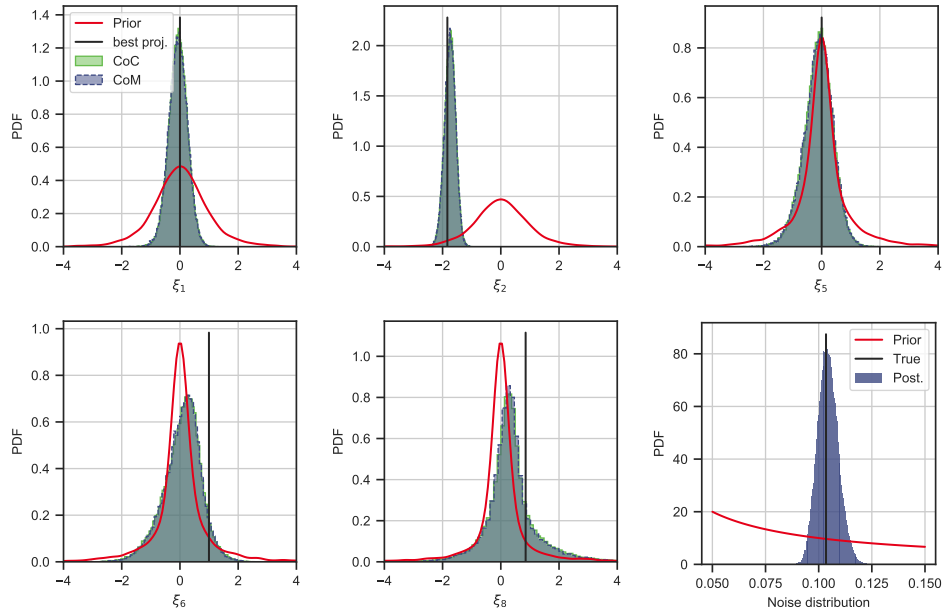


Figure 7: TD case for $g^{\sin}$ - Marginal posterior distribution of coordinates 1, 2, 5, 6 and 8 and noise level (coordinates 3, 4 and 7 are not shown for brevity). Best projection corresponds to the coordinates used to plot Fig. 6 and true noise is the standard deviation of the synthetical noises.

relatively to the true field, this gap is reduced when decreasing the noise level (see Fig 9 where σ_ε is set to 0.02). Figure 9 assesses the convergence of the CoM inference result with respect to the PC order for the forward model, the number of coordinates and the PC order for the CoM prior quantities. Capturing the stochastic nonlinearities of the forward model is crucial (left figure) while the median field appears to be less sensitive to the approximation of the CoM prior quantities (middle figure). It is clear that a too low number of coordinates is detrimental to the field inference (right figure).

6. Application to seismic tomography

The CoM method is applied to a *seismic tomography* (ST) problem based on the propagation of the P-waves in the first kilometer of the Earth's crust. After introducing the test case in Section 6.1, the results are presented in Section 6.2. Results for different field shapes are presented in Section 6.3. Section 6.4 provides a comparison with inference results with fixed hyperparameters.

6.1. Case presentation

We consider the inference of a 2D continuous seismic velocity field v from observations of traveltimes between known sources and receivers. The observation vector $\boldsymbol{d}_{\text{obs}}$ corresponds to the 5 sources and 23 receivers depicted in Fig. 10, such that the dimension of $\boldsymbol{d}_{\text{obs}}$ is $N = 115$. The forward model is the eikonal equation [54] that relates the traveltimes map $t_s(\boldsymbol{x})$ from a source at $\boldsymbol{x} = s$ to the velocity field v,

$$|\nabla t_s(\boldsymbol{x})|^2 = \frac{1}{v^2(\boldsymbol{x})}, \text{ with } t_s(s) = 0 \tag{42}$$

where $\boldsymbol{x} \in \Omega$ is the spatial position. The quantities are expressed with SI units: $t \sim$ [s], $x \sim$ [m] and $v \sim [\text{m} \cdot \text{s}^{-1}]$. The observations $\boldsymbol{d}_{\text{obs}}$ are synthetically generated by solving Eq. (42) with $v = v_{\text{true}}$ and corrupting the traveltimes with an i.i.d. additive noise having a centered Gaussian distribution with standard deviation equal to $\sigma_\varepsilon = 0.002 \sim$ [s],

$$\boldsymbol{d}_{\text{obs}} = \boldsymbol{d}_{\text{true}} + \varepsilon, \quad \varepsilon = \mathcal{N}(0, \sigma_\varepsilon^2 \mathrm{I}_N). \tag{43}$$

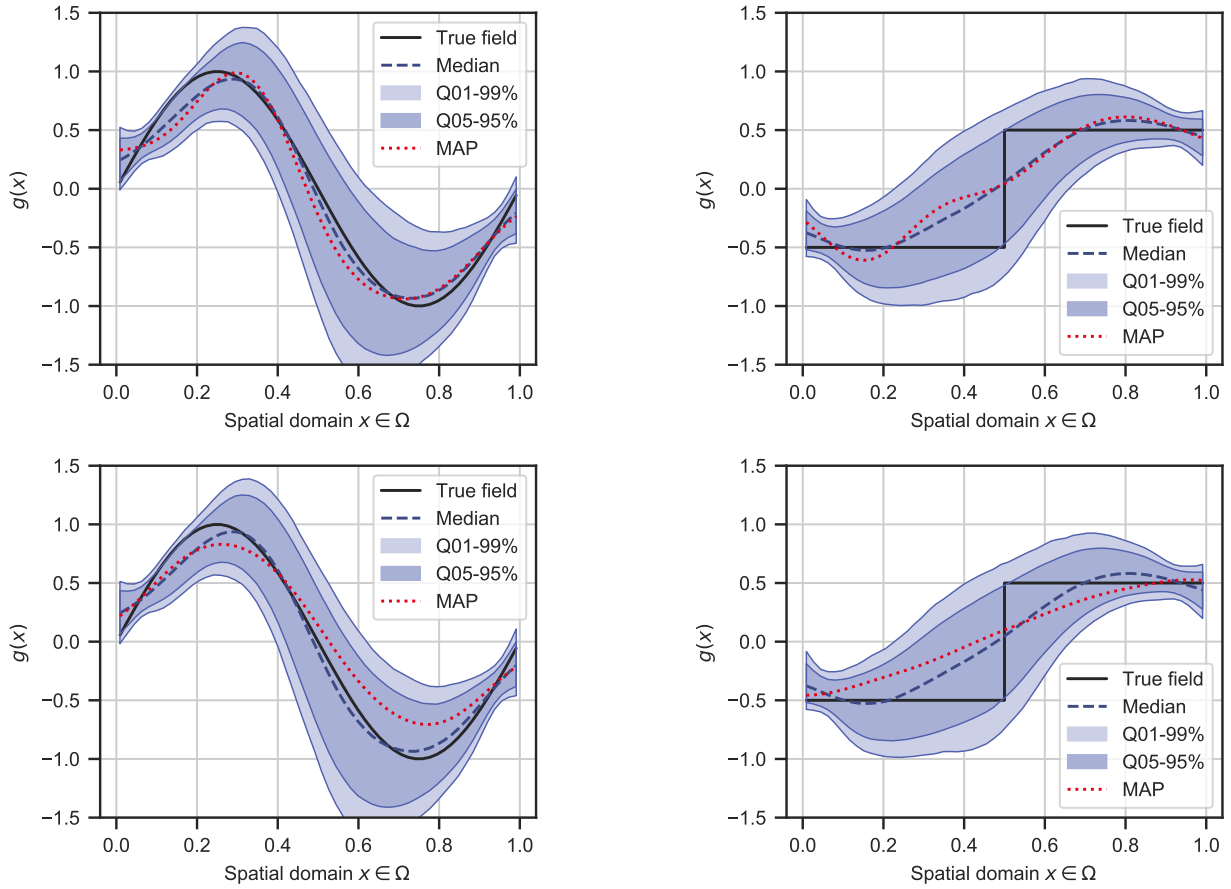


Figure 8: TD case - Posterior distribution of the inferred fields. Black line: g_{true}; star blue line: median field; dashed red line: MAP; blue intervals: $1\% - 99\%$ and $5\% - 95\%$ quantiles. (left) g^{sin} case / (right) g^{step} case ; (top) Change of coordinates / (bottom) Change of measure

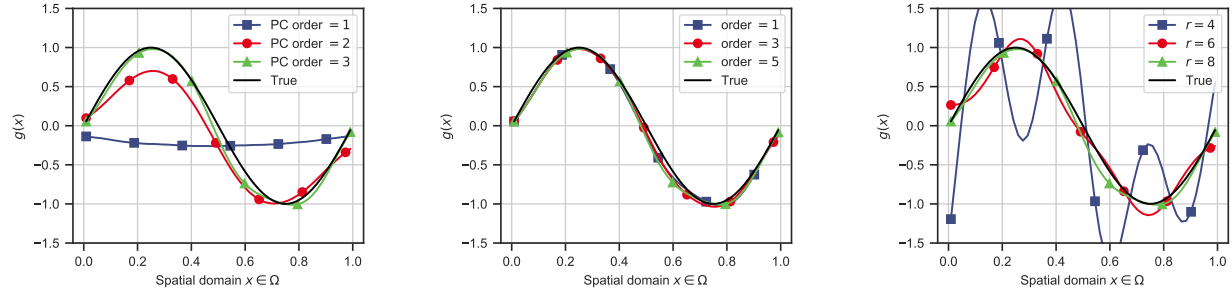


Figure 9: TD case for g^{sin} - Convergence of MCMC posterior median according to different parameters (CoM method): (left) PC order for the forward model, (middle) PC order for the CoM prior quantities, (right) number of coordinates. For fixed parameters, inference realized with $r = 8$, $\sigma_\varepsilon = 0.02$, PC order $\widetilde{\mathcal{M}} = 5$, PC order prior $= 15$.

For the sources-receivers geometry and the true field we consider, the noise is around 1% of the true times. In this work, $v_{\text{true}} : \Omega \to \mathbb{R}_+$ is adapted from [55] and depends only on the vertical coordinate. We again consider a log-normal prior distribution:

$$v(\boldsymbol{x}) = \exp(g(\boldsymbol{x})), \quad g \sim \mathcal{N}(c, k), \quad k(\boldsymbol{x}, \boldsymbol{y}) = A \exp\left(\frac{-\|\boldsymbol{x} - \boldsymbol{y}\|^2}{2l^2}\right), \tag{44}$$

where $c \in \mathbb{R}$ is a constant trend to infer. The field $g(\boldsymbol{x})$ writes

$$g(\boldsymbol{x}) = c + \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x}) \xi_i, \tag{45}$$

with $(\overline{u}_i, \overline{\lambda}_i)_{1 \leqslant i \leqslant r}$ the eigenelements of $\mathbb{E}_{\mathbb{H}}(k)$. In that case, the forward model surrogate depends on the coordinates $\boldsymbol{\xi}$ and on the constant trend c. *A priori*, the constant trend is independent from the coordinates and from the other hyperparameters, leading to the following expression for the posterior distribution

$$\begin{aligned}
\pi_{\text{post}}(\boldsymbol{\xi}, \boldsymbol{q}, c, \sigma_\varepsilon) &\propto \mathcal{L}(\boldsymbol{d}^{\text{obs}} | \boldsymbol{\xi}, \boldsymbol{q}, c, \sigma_\varepsilon) \pi(\boldsymbol{\xi} | \boldsymbol{q}, c, \sigma_\varepsilon) \pi(\boldsymbol{q}, c, \sigma_\varepsilon) \\
&\propto \frac{1}{\sqrt{(2\pi\sigma_\varepsilon^2)^N}} \exp\left(\frac{-\left\|\boldsymbol{d}^{\text{obs}} - \boldsymbol{d}(\boldsymbol{\xi}, c)\right\|_{l_2}^2}{2\sigma_\varepsilon^2}\right) \times \frac{1}{\sqrt{(2\pi)^r \left|\Sigma(\boldsymbol{q})\right|}} \exp\left(\frac{-\boldsymbol{\xi}^\top \Sigma(\boldsymbol{q})^{-1} \boldsymbol{\xi}}{2}\right) \\
&\quad \times \pi(c) \pi_{\mathbb{H}}(\boldsymbol{q}) \pi_{\sigma_\varepsilon}(\sigma_\varepsilon)
\end{aligned}$$

The prior distributions of the constant trend and the two hyperparameters are

$$c \sim \mathcal{U}(6.9, 8.1), \quad A \sim \text{InvGamma}(21, 1) \quad \text{and} \quad l \sim \mathcal{U}(10, 100). \tag{46}$$

These priors have been chosen to generate a large set of plausible velocity fields, see Fig. 10 (left). The number of reduced coordinates is $r = 20$; it accounts for 96.8% of the prior variance and allows building accurate and tractable surrogate models with sparse grids. For $r = 20$, the best approximation of v_{true},

$$v_{\text{best}}^r(\boldsymbol{x}) = \exp\left(c_b + \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x}) v_i\right),$$

where c_b is the spatial mean of $\log v_{\text{true}}$ and the coordinates v_i result from the projection of $\log v_{\text{true}} - c_b$ on the reference basis as done in Eq. (11). This approximation yields to low relative mean squared errors

$$\|v_{\text{true}} - v_{\text{best}}^r\| \,/\, \|v_{\text{true}}\| \simeq 2\% \text{ and } \|\boldsymbol{d}(v_{\text{true}}) - \boldsymbol{d}(v_{\text{best}}^r)\| \,/\, \|\boldsymbol{d}(v_{\text{true}})\| \simeq 0.1\%. \tag{47}$$

In other words, the component of $\log v_{\text{true}}$ that is orthogonal to the span of the reduced basis has only a weak impact on the traveltimes.

6.2. Inference results

A PC order of 15 have been selected for the surrogate model of $A^{\pm 1/2} \Sigma^{\pm 1/2}(l)$ and $r \log A + \text{logdet}_\Sigma(l)$ ensuring a RRMSE around 10^{-3}. Regarding the traveltimes, a level 3 sparse grid with $15,135$ points (dimension $r + 1 = 21$), requiring $75,675$ solves of the eikonal equation (42) because of the 5 sources, has been used with a RRMSE of 1%.

With the surrogate, a MCMC chain burn-in of 10^6 steps is performed, where the $(r + 4)^2$ MCMC covariance proposal is adapted every $5{\times}10^4$ steps. The acceptance rate is converging around 24% at the end of the burn-in. After the burn-in, $5{\times}10^6$ steps are performed. The analysis shows an effective sample size around $6,000$ using the method proposed in [53]. Figure 11 depicts the posterior marginal histograms of the coordinates together with their respective prior. The first coordinates, except for ξ_1, are rather well inferred

with narrow distributions containing the projection coordinates while the last coordinates stay near their prior. The posterior of the constant trend concentrates. The first coordinate ξ_1 and the constant trend c are difficult to distinguish since $\overline{u}_1$ is nearly constant. This explains why the first coordinate is not peaked in contrast to $\boldsymbol{\xi}_{2-10}$. The posterior of the hyperparameters are also plotted on Fig. 11. We remark that the observations are weakly informative for the prior variance whose posterior distribution remains close to its prior. The posterior distribution of the correlation length is more concentrated in the middle of the prior range and is asymetrical with a higher probability for low values. The fact that the posterior distribution of l is not peaked highlights the interest of exploring the hyperparameters space to propose a large variety of plausible fields. Furthermore, we computed the probability of the true log field to be a Gaussian process realization with autocovariance function $k(\boldsymbol{q})$. The best hyperparameters couple is $(0.04, 28)$ which is broadly consistent with the posterior distributions in spite of the noise level, that can affect these best values [31]. The noise level σ_ε is correctly inferred with a distribution peaking at 1.8ms with a standard deviation of 0.1ms, consistent with the true value of 2ms.

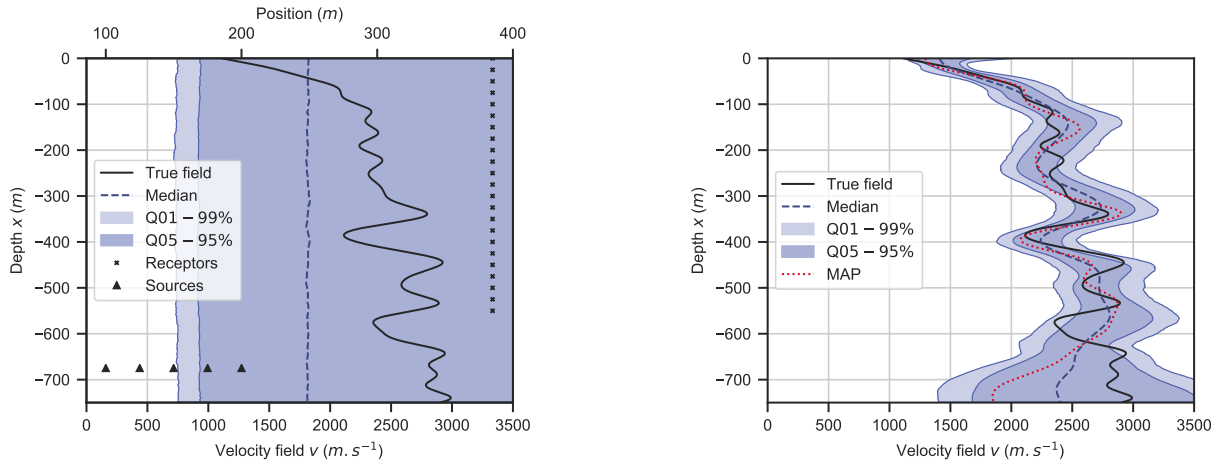


Figure 10: ST case - Distribution of the velocity field f prior (left) and posterior (right).

We now present our procedure to obtain the posterior distribution of the velocity field. To take into account the uncertainty due to the truncation, we add the K next terms in the approximation as done in [17]. *A priori*,

$$g(\boldsymbol{x}) = c + \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x}) \xi_i + \sum_{i=r+1}^{r+K} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x}) X_i, \tag{48}$$

with the $R \coloneqq r + K$ coordinates $(\boldsymbol{\xi}, \boldsymbol{X}) \sim \mathcal{N}(0, \Sigma(\boldsymbol{q}))$, where $\Sigma(\boldsymbol{q}) \in \mathbb{R}^{R \times R}$ is defined as in Eq. (20) and is partitioned as follows

$$\Sigma(\boldsymbol{q}) = \begin{pmatrix} \Sigma_{rr}(\boldsymbol{q}) & \Sigma_{rK}(\boldsymbol{q}) \\ \Sigma_{Kr}(\boldsymbol{q}) & \Sigma_{KK}(\boldsymbol{q}) \end{pmatrix}. \tag{49}$$

Thus, $\boldsymbol{X}|\boldsymbol{\xi} \sim \mathcal{N}(\overline{\mu}, \overline{\Sigma}(\boldsymbol{q}))$, where

$$\overline{\mu} = \Sigma_{Kr}(\boldsymbol{q})\Sigma_{rr}(\boldsymbol{q})^{-1}\boldsymbol{\xi} \text{ and } \overline{\Sigma}(\boldsymbol{q}) = \Sigma_{KK}(\boldsymbol{q}) - \Sigma_{Kr}(\boldsymbol{q})\Sigma_{rr}(\boldsymbol{q})^{-1}\Sigma_{rK}(\boldsymbol{q}). \tag{50}$$

In practice, we use $R = r + K = 20 + 60 = 80$. Since generating $\boldsymbol{X}$ knowing $\boldsymbol{\xi}$ and $\boldsymbol{q}$ requires the construction of $\Sigma_{Kr}(\boldsymbol{q})$ and $\Sigma_{KK}(\boldsymbol{q})$ and the decomposition of $\overline{\Sigma}(\boldsymbol{q})$, the cost can be significant. Subsamples of the MCMC chains are extracted and their velocity field are generated randomly. In practice, one sample over 100 is selected. The median, MAP and two quantiles range of the field posterior and prior distributions are represented on Fig. 10. The MAP is the sample from the chain that maximizes the likelihood. The median and the MAP globally capture the main variations of the true field. The confidence intervals are quite tight

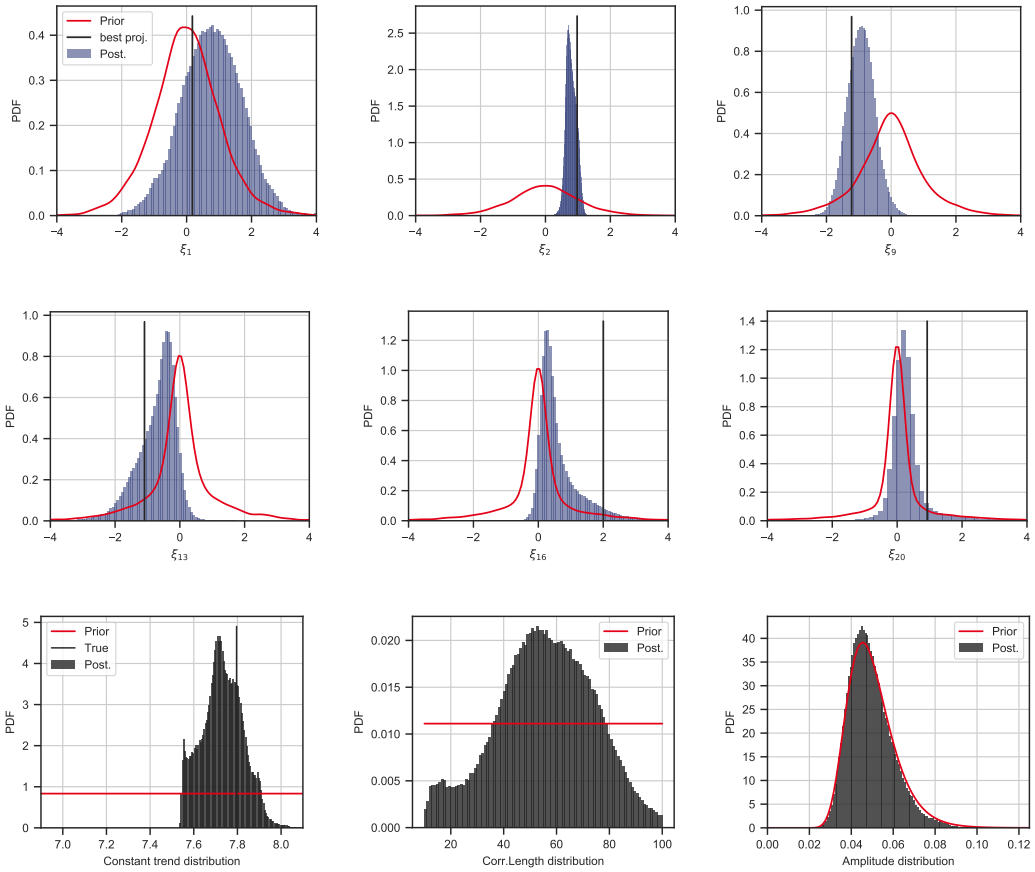


Figure 11: ST case - Marginal posterior distributions of some coordinates $\boldsymbol{\xi}$ and (hyper)parameters. Best projection coordinates are g_{kl} coordinates (sparse noised observations not taken into account).

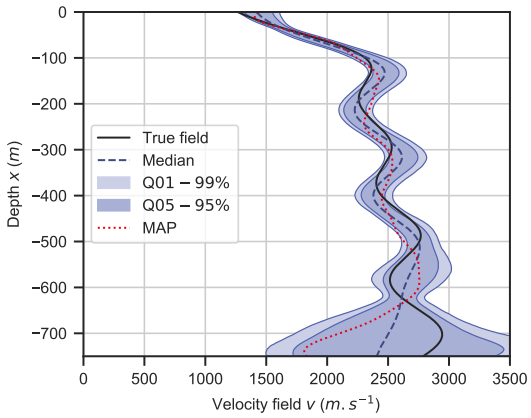
(especially at the top of the domain) and remain consistent with the true field. Larger uncertainties remains at the bottom of the domain because of the lack of observations there.

6.3. Comparison for different fields

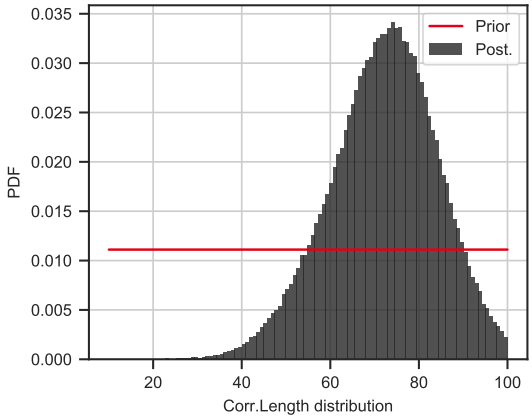
To evaluate the robustness of the CoM method, we repeat the inference for two versions of v_{true}: i) v_{LW} obtained by smoothing v_{true} to remove short wavelengths, and ii) v_{SW} obtained by perturbing v_{true} with short wavelengths. Posterior fields and correlation lengths distributions are shown on Fig. 12. The posterior uncertainties depend on the wavelengths of the field: they are rather smooth and tight for v_{LW} and less regular and larger for v_{SW}. This behaviour is consistent with the modified true field shapes. Further, the correlation length distributions are impacted in a coherent way: the small correlation lengths are much less likely for v_{LW} than for v_{SW} whereas for v_{SW} the whole prior range remains plausible. Clearly, v_{SW} is more difficult to infer than v_{LW} because of its broader range of samples profile. The inference of short wavelengths structures would require a denser arrangement of sources and receivers and possibly more information observations with incidence angle at arrival and/or full waveform analysis.

6.4. Comparison with fixed hyperparameters inference

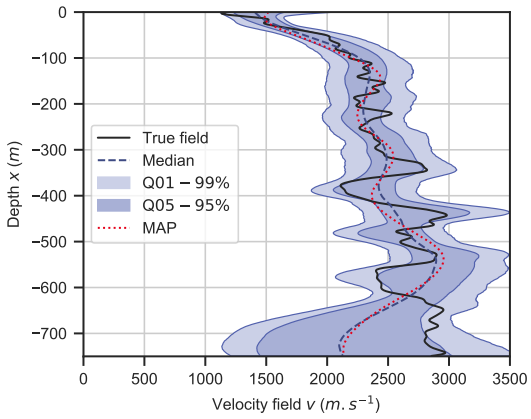
To assess the interest of the CoM method, we point out the limitations of using bases with fixed correlation lengths to infer v_{LW} and v_{SW}. Two bases with different correlation lengths are built, i) $l = 10$ to capture short wavelengths, and ii) $l = 80$ to focus on long wavelengths. For the two bases, an inference is led considering



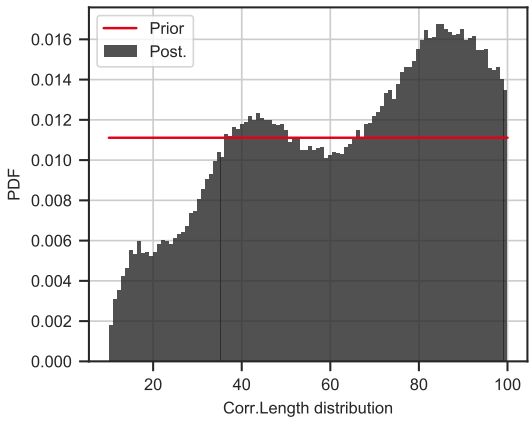
(a) Posterior field distribution for v_{LW}



(b) Posterior corr.length distribution for v_{LW}



(c) Posterior field distribution for v_{SW}



(d) Posterior corr.length distribution for v_{SW}

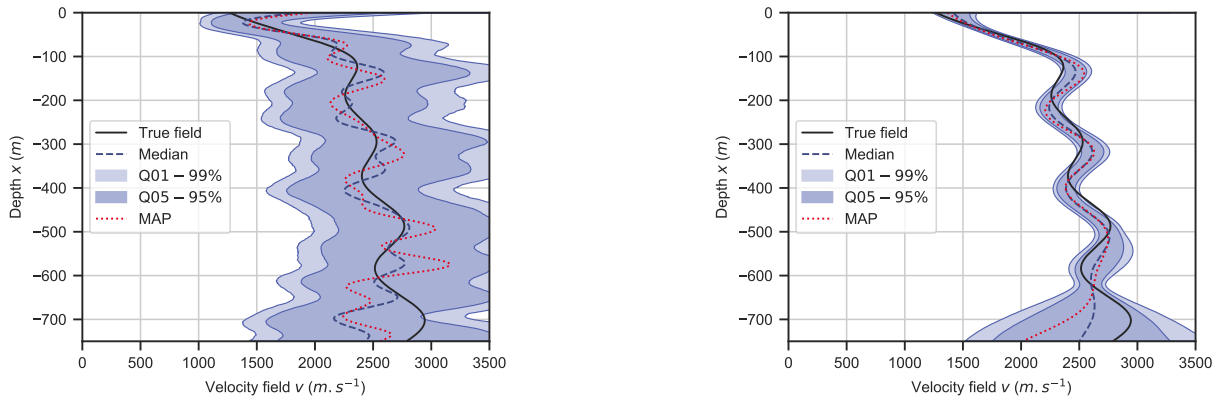
Figure 12: ST case - Inference results for two fields with various wavelengths. Posterior distributions are obtained adding $K = 60$ terms after the inference.

only the standard deviation noise σ_ε, the constant trend c and the amplitude A as hyperparameters, and inferring the coordinates in the basis parametrized with $l = 10$ or $l = 80$, with a standard Gaussian prior.

Figure 13 displays the posterior field confidence interval obtained by summing $r = 20$ (and $K = 60$) coordinates. We see that a larger correlation length reduces the posterior uncertainties where the true field is well correlated. However, large l values are unable to capture the small structures of v_{SW}. On the contrary, smaller correlation lengths can account for small structures of v_{SW} but without reducing much the prior uncertainties. This result can be explained by the unsufficient information to learn the short wavelength features and the number of terms $r = 20$ used. Comparing Fig. 12 and Fig. 13 shows the importance of allowing the inference of the covariance hyperparameters, since the hyperparameters choice is impacting the posterior distribution a lot. In turn, the exploration of the hyperparameters space encountered in the CoM method improves the estimation of the field.

7. Application to a 2D groundwater flow problem

The CoM method is finally applied to a two-dimensional case with more hyperparameters to illustrate the approach's feasibility on a more realistic problem. The proposed test case introduced in Section 7.1 is a



(a) v_{LW}, $l = 10$

(b) v_{LW}, $l = 80$

(c) v_{SW}, $l = 10$

(d) v_{SW}, $l = 80$

Figure 13: ST case - Posterior field distribution for different field and inference bases. $r = 20$, $K = 60$ coordinates are added after the inference.

source-sink groundwater flow problem inspired from [34]. The inference results discussed in Section 7.2 show that the CoM method is able to deal with a parametrization adapted for anisotropic permeability fields.

7.1. Case presentation

A steady-state groundwater flow problem with impermeable boundary condition is considered,

$$\begin{cases} -\nabla \cdot (\kappa(\boldsymbol{x}) \nabla u(\boldsymbol{x})) = f(\boldsymbol{x}), & \boldsymbol{x} \in \Omega = [0,1]^2, \\ \nabla u(\boldsymbol{x}) \cdot \boldsymbol{n} = 0, & \boldsymbol{x} \in \partial\Omega \backslash \mathbf{0}, \\ u(\mathbf{0}) = 0, \end{cases} \tag{51}$$

where u is the pressure field and κ is the permeability field. We set $u(\mathbf{0}) = 0$ to ensure the well-posedness of the problem. The right-hand side term f, plotted in Fig. 14 (left), consists of a source in the top-left corner and a sink in the bottom-right corner, with rates following a squared-exponential shape centered on the corners and having a characteristic width of 0.1. Given a realization of κ, the forward problem is solved by the $\mathbb{P}_2$ triangular finite element method using a mesh generated using [56].

The objective is to infer the permeability field κ. The prior of $\log \kappa$ is a Gaussian distribution with an

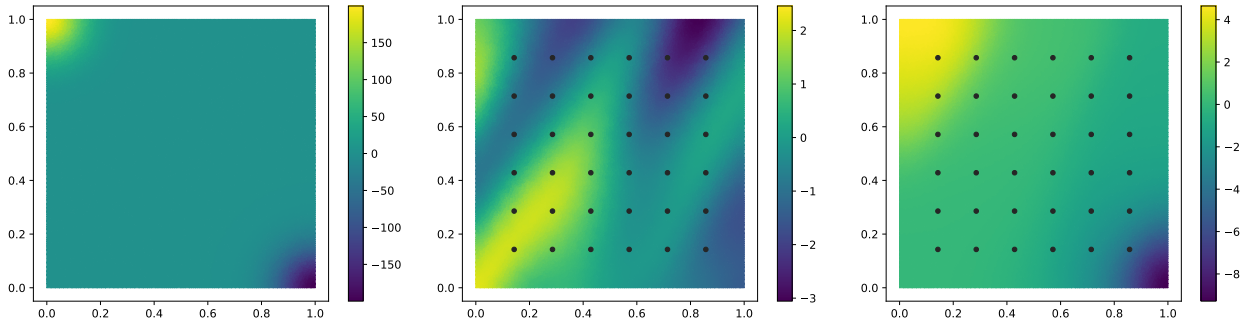


Figure 14: Case presentation: (left) Source term f, (middle) True log-field $\log \kappa_{\text{true}}$, (right) True solution u_{true}. Black circles are observation locations.

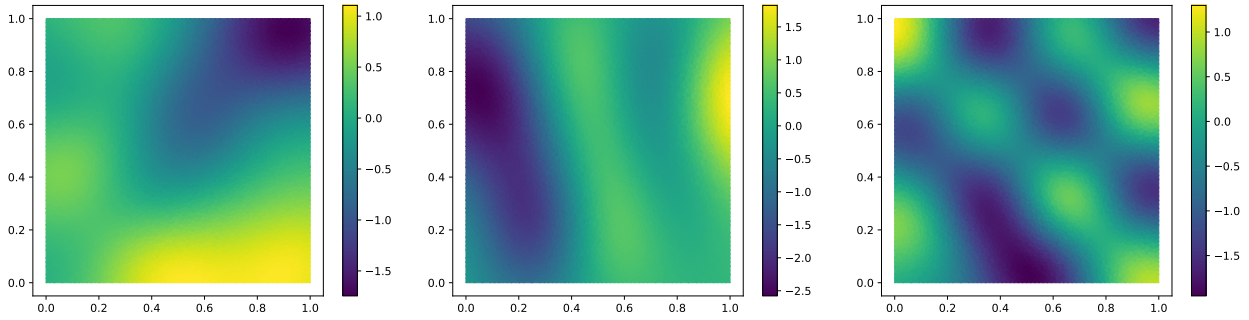


Figure 15: Few samples $\log \kappa$.

anisotropic squared exponential autocovariance function,

$$\log \kappa \sim \mathcal{N}(0, k) \quad \text{with} \quad \begin{cases} k(\boldsymbol{x}, \boldsymbol{y}) = A \exp\left(\dfrac{-(\boldsymbol{x}-\boldsymbol{y})^\top K (\boldsymbol{x}-\boldsymbol{y})}{2}\right), \quad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \Omega, \\ K = R(\theta) \begin{pmatrix} 1/l_1^2 & 0 \\ 0 & 1/l_2^2 \end{pmatrix} R(\theta)^\top, \end{cases} \tag{52}$$

where $R(\theta)$ denotes the 2D rotation matrix with orientation θ and l_1 and l_2 are the two correlation lengths characterizing the anisotropy. The four autocovariance hyperparameters priors are

$$A \sim \text{IG}(3, 14), \quad l_1 \sim \mathcal{U}(0.1, 0.6), \quad l_2 \sim \mathcal{U}(0.1, 0.6), \quad \theta \sim \mathcal{U}(0, \pi/2). \tag{53}$$

The true field κ_{true} used to generate the observations is a particular realization of the random process described in Eq. (52) with $\{A^\star,\ l_1^\star,\ l_2^\star,\ \theta^\star\} = \{1.3,\ 0.4,\ 0.15,\ \pi/3\}$. This realization is shown in the center plot of Fig. 14, along with its associated pressure field u_{true} (right plot) computed with a fine uniform triangulation of the domain with $27,763$ elements. The observations correspond to the pressure field u_{true} measured over a uniform grid of 6×6 sensors (see Fig. 14). An i.i.d. Gaussian noise of standard deviation $\sigma_\varepsilon = 0.02$ (around 1% of the average magnitude of the true values) is added to play the role of measurement errors.

The truncated KL decomposition used for the inference has $r = 20$ modes which capture 97% of the prior variance. The inference uses a coarser mesh composed of $7,875$ elements to introduce a model error term. Using this parametrization, some samples of $\log \kappa$ drawn from its prior are shown in Fig. 15 to highlight the rich structure of the *a priori*.

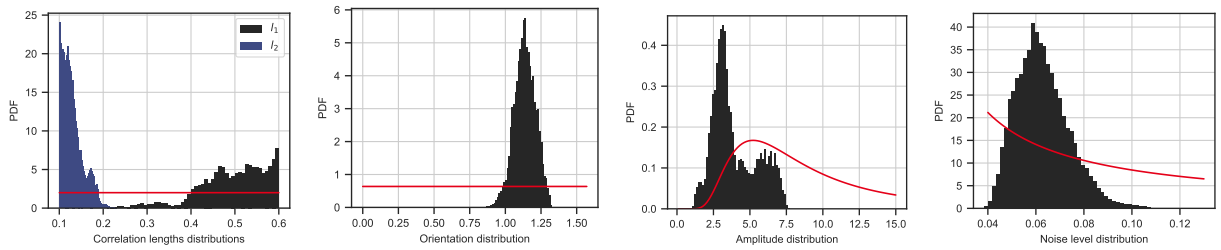


Figure 16: Marginal distributions of kernel hyperparameters (left to right) first and second correlation lengths l_1, l_2, orientation θ, amplitude A, noise level σ_ε

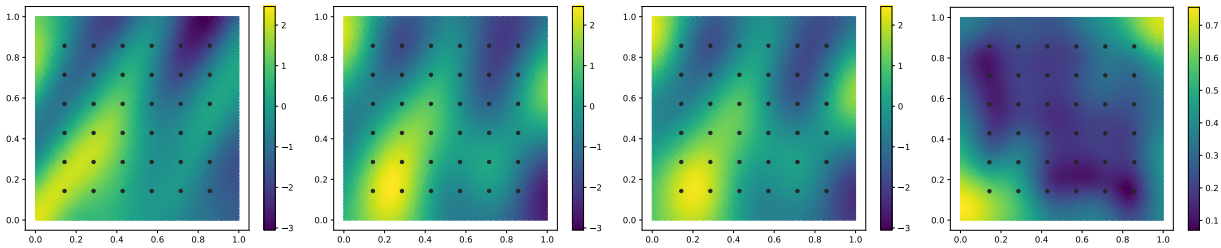


Figure 17: (second plot) MAP (third plot) mean and (fourth plot) standard deviation of the posterior $\log \kappa$ distribution. The true log-field is recalled (first plot) in order to facilitate the comparison.

7.2. Results

The parameters to infer are $r = 20$ reduced coordinates, the amplitude A, the two correlation lengths l_1 and l_2, the orientation θ as well as the noise standard deviation σ_ε, leading to a sampling space of dimension 25. As for the previous test cases, PC surrogates are built for $\Sigma(\boldsymbol{q})^{\pm 1/2}/A^{\pm 1/2}$ and $\mathrm{logdet}_\Sigma(\boldsymbol{q}) - r \log A$ where $\boldsymbol{q} = \{l_1,\ l_2,\ \theta\}$ (see Section 4.3). A RRMSE lower than 1% is achieved with a 12th order PC expansion (total degree truncation) fitted using a Latin hypercube sampling of size $1,000$. The $\mathbb{P}_2$ finite element method used to solve Pb. (51) is fast enough to not require the use of a forward model surrogate.

The posterior marginals of the five hyperparameters are reported in Fig. 16. The marginal of the orientation peaks around $\pi/3$ while the marginal of the two correlation lengths are distinct. The extent of the marginal of A is sensibly reduced compared to its prior distribution. These marginals are somehow consistent with the particular values of hyperparameters used for generating the true field. Finally, the marginal of the noise level shows likely values slightly higher than the value used for generating the observations, as expected due to a non-vanishing model error induced by using a coarser mesh and a truncation of the field representation for the inference.

Figure 17 shows the true log field (first plot), its MAP estimate (second plot), posterior mean (third plot), and standard deviation (fourth plot). It is seen that the MAP and the posterior mean field are close to the true field. Further, the posterior uncertainty, as measured by the standard deviation, is higher in the corners far from the observations and the injection and pumping areas.

To highlight the benefit of using a parametrized prior, we compared the inferred fields with two other approaches based on an isotropic covariance ($l_1 = l_2$, no orientation). The first relies on the CoM method with parameters $\{A, l\}$ and the second uses a KL parametrization with a fixed correlation length $l = 0.3$ and amplitude $A = 1.5$. In all cases, we use $r = 20$ reduced coordinates in the inference. We assess the inference quality by reporting the posterior distribution of the norm of the difference between the inferred log-fields and the true log-field. These distribbutions, reported in Fig. 18, show that a richer parametrization of the prior significantly reduces the posterior distance to the true field. The error on the MAP estimates, shown with the vertical lines, exhibits consistent behavior. These results demonstrate the importance of effectively treating hyperparameters to enable more general priors.

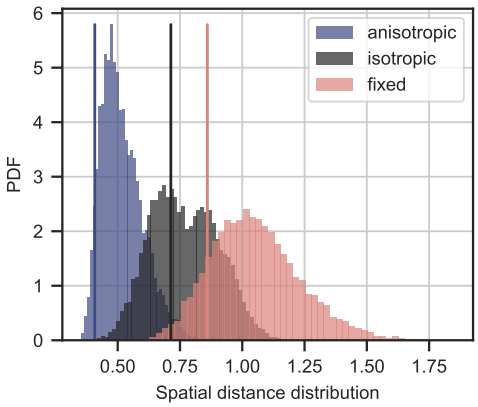


Figure 18: Posterior distributions of the spatial distance between the true log-field and posterior realizations $||\log \kappa_{true} - \log \kappa^{(i)}||_{L^2(\Omega)}$ for the (blue) anisotropic (gray) isotropic (pink) fixed hyperparameters case. Distances between the true log-field and the MAP log-fields are indicated with lines.

8. Conclusion

In this paper, we developed and implemented a novel approach to efficiently deal with hyperparameters in Bayesian inference of Gaussian fields. Specifically, the approach relies on a representation of the field in a fixed reduced basis, carefully selected, and the prior's covariance of the reduced coordinates is seen as a function of the hyperparameters. The cornerstone of our approach is a change of measure designed to avoid the dependency of the likelihood on the hyperparameters. From a methodological viewpoint, the change of measure yields a posterior that depends smoothly on the hyperparameters and reduced coordinates, compared to alternative approaches such as the change of coordinates which can exhibit discontinuities.

This smoothness enables to use surrogate models to decrease the computational cost of the MCMC sampling. In practice, the CoM results in a hierarchical Bayes formulation with a closed-form expression for the hyperparameters-dependent prior: it allows the efficient sampling of the joint distribution with an auxiliary variable. The method was demonstrated on three test cases with various complexities and geometries. The numerical experiments demonstrate that the hyperparameters space exploration is i) highly valuable because it provides a better estimation of the field uncertainties and ii) computationally tractable thanks to the CoM formulation combined with surrogate models.

Further works should focus on the extension of CoM methods to non-Gaussian random processes with possibly non stationnary prior autocovariance functions. In that case, the change of measure formulation needs to be generalized to account for possible non Gaussian dependencies in the priors. For seismic tomography applications, it would be interesting to add other types of observations, such as the direction of propagation, the maximum amplitude or even the full waveform, to improve the estimation of the reduced coordinates with higher indices. More generally, the extension to three-dimensional cases is not conceptually limited by the CoM method but by the number of KL modes required to accurately represent the field of interest. Current work focuses on the development of adaptive dimension reduction techniques to tackle this issue.

[1] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, Vol. xii, SIAM, 2005.

[2] D. Sivia, J. Skilling, Data Analysis: A Bayesian Tutorial, Oxford science publications, Oxford, 2006.

[3] A. Malinverno, Parsimonious Bayesian Markov Chain Monte Carlo Inversion in a Nonlinear Geophysical Problem, Geophys. J. Int. 151 (2002) 675–688.

[4] P. Sochala, A. Gesret, O. Le Maitre, Polynomial Surrogates for Bayesian Traveltime Tomography, GEM - Int. J. Geomath. 12 (1) (Dec. 2021).

[5] D. Grana, L. Azevedo, L. de Figueiredo, P. Connolly, T. Mukerji, Probabilistic Inversion of Seismic Data for Reservoir Petrophysical Characterization: Review and Examples, Geophysics 87 (5) (2022) M199–M216.

[6] D. R. Rounce, T. Khurana, M. B. Short, R. Hock, D. E. Shean, D. J. Brinkerhoff, Quantifying Parameter Uncertainty in a Large-Scale Glacier Evolution Model Using Bayesian Inference: Application to High Mountain Asia, J. Glaciol. 66 (256) (2020) 175–187.

[7] Y. Ying, J. Maddison, J. Vanneste, Bayesian Inference of Ocean Diffusivity from Lagrangian Trajectory Data, Ocean Model. 140 (2019) 101401.

[8] A. Doucet, A. Smith, N. de Freitas, N. Gordon, Sequential Monte Carlo Methods in Practice, Information Science and Statistics, Springer NY, 2013.

[9] Y. Marzouk, H. Najm, L. Rahn, Stochastic Spectral Method for Efficient Bayesian Solution of Inverse Problems, J. Comput. Phys. 224 (2007) 560–586.

[10] M. Navarro, O. P. Le Maître, I. Hoteit, D. L. George, K. T. Mandli, O. M. Knio, Surrogate-Based Parameter Inference in Debris Flow Model, Comput. Geosci. 22 (6) (2018) 1447–1463.

[11] Y. Marzouk, H. Najm, Dimensionality Reduction and Polynomial Chaos Acceleration of Bayesian Inference in Inverse Problems, J. Comput. Phys. 228 (2009) 1862–1902.

[12] K. Karhunen, Zur Spektraltheorie Stochastischer Prozesse, Ann. Acad. Sci. Fenn. (1946).

[13] M. Loève, Probability Theory I, Vol. 45 of Graduate Texts in Mathematics, Springer NY, 1977.

[14] M. Meyer, H. G. Matthies, Efficient Model Reduction in Non-Linear Dynamics Using the Karhunen-Loève Expansion and Dual-Weighted-Residual Methods, Comput. Mech. 31 (2003) 179–191.

[15] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, The MIT Press, 2005.

[16] F. Uribe, I. Papaioannou, W. Betz, D. Straub, Bayesian Inference of Random Fields Represented with the Karhunen–Loève Expansion, Comput. Methods Appl. Mech. Eng. 358 (2020) 112632.

[17] G. A. Meles, N. Linde, S. Marelli, Bayesian Tomography with Prior-Knowledge-Based Parametrization and Surrogate Modelling, Geophys. J. Int. 231 (1) (2022) 673–691.

[18] S. Khatoon, J. Phirani, S. S. Bahga, Fast Bayesian Inference for Inverse Heat Conduction Problem Using Polynomial Chaos and Karhunen–Loeve Expansions, Applied Thermal Engineering 219 (2023) 119616.

[19] E. Laloy, B. Rogiers, J. A. Vrugt, D. Mallants, D. Jacques, Efficient Posterior Exploration of a High-Dimensional Groundwater Model from Two-Stage Markov Chain Monte Carlo Simulation and Polynomial Chaos Expansion, Water Resour. Res. 49 (5) (2013) 2664–2682.

[20] P. M. Tagade, H.-L. Choi, A Generalized Polynomial Chaos-Based Method for Efficient Bayesian Calibration of Uncertain Computational Models, Inverse Problems in Science and Engineering 22 (4) (2014) 602–624.

[21] I. Sraj, O. P. Le Maître, O. M. Knio, I. Hoteit, Coordinate Transformation and Polynomial Chaos for the Bayesian Inference of a Gaussian Process with Parametrized Prior Covariance Function, Comput. Methods Appl. Mech. Eng. 298 (2016) 205–228.

[22] J. Latz, M. Eisenberger, E. Ullmann, Fast Sampling of Parameterised Gaussian Random Fields, Comput. Methods Appl. Mech. Eng. 348 (2019) 978–1012.

[23] A. Siripatana, O. Le Maître, O. Knio, C. Dawson, I. Hoteit, Bayesian Inference of Spatially Varying Manning's n Coefficients in an Idealized Coastal Ocean Model Using a Generalized Karhunen-Loève Expansion and Polynomial Chaos, Ocean Dyn. 70 (2020).

[24] N. Wiener, The Homogeneous Chaos, Am. J. Math. 60 (4) (1938) 897–936.

[25] R. G. Ghanem, P. D. Spanos, Stochastic Finite Element Method: Response Statistics, in: Stochastic Finite Elements: A Spectral Approach, Springer NY, 1991, pp. 101–119.

[26] D. Xiu, G. E. Karniadakis, The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.

[27] P. C. Hansen, Discrete Inverse Problems, SIAM, 2010.

[28] T. Bodin, M. Sambridge, N. Rawlinson, P. Arroucau, Transdimensional Tomography with Unknown Data Noise, Geophys. J. Int. 189 (3) (2012) 1536–1556.

[29] N. Piana Agostinetti, G. Giacomuzzi, A. Malinverno, Local Three-Dimensional Earthquake Tomography by Trans-Dimensional Monte Carlo Sampling, Geophys. J. Int. 201 (3) (2015) 1598–1617.

[30] J. Belhadj, T. Romary, A. Gesret, M. Noble, B. Figliuzzi, New Parameterizations for Bayesian Seismic Tomography, Inverse Problems 34 (6) (2018) 065007.

[31] A. Musolas, S. T. Smith, Y. Marzouk, Geodesically Parameterized Covariance Estimation, SIAM J. Matrix Anal. Appl. 42 (2) (2021) 528–556.

[32] J.-L. Akian, L. Bonnet, H. Owhadi, É. Savin, Learning "Best" Kernels from Data in Gaussian Process Regression. With Application to Aerodynamics, J. Comput. Phys. 470 (2022) 111595.

[33] J. Zhang, B. Ellingwood, Orthogonal Series Expansions of Random Fields in Reliability Analysis, J. Eng. Mech. 120 (12) (1994) 2660–2677.

[34] T. Cui, J. Martin, Y. M. Marzouk, A. Solonen, A. Spantini, Likelihood-informed dimension reduction for nonlinear inverse problems, Inverse Problems 30 (11) (2014) 114015. doi:10.1088/0266-5611/30/11/114015.

[35] J. Li, Y. M. Marzouk, Adaptive Construction of Surrogates for the Bayesian Solution of Inverse Problems, SIAM Journal on Scientific Computing 36 (3) (2014) A1163–A1186. doi:10.1137/130938189.

[36] J. Mercer, A. R. Forsyth, XVI. Functions of Positive and Negative Type, and their Connection to the Theory of Integral Equations, Philos. Trans. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character 209 (441-458) (1909) 415–446.

[37] F. Zhang, R. Dai, H. Liu, Seismic inversion based on L1-norm misfit function and total variation regularization, J. Appl. Geophys. 109 (2014) 111 – 118.

[38] N. Leoni, Bayesian Inference of Model Error for the Calibration of Two-Phase CFD Codes, Theses, Institut Polytechnique de Paris (Apr. 2022).

[39] H. Jeffreys, An Invariant Form for the Prior Probability in Estimation Problems, Proc. R. Soc. Lond. Ser. Math. Phys. Sci. 186 (1007) (1946) 453–461.

[40] R. E. Kass, L. Wasserman, The Selection of Prior Distributions by Formal Rules, J. Am. Stat. Assoc. 91 (435) (1996) 1343–1370.

[41] N. Chopin, C. Robert, J. Rousseau, Harold Jeffreys' Theory of Probability Revisited, Stat. Sci. 107 (3) (2009) 141–172.

[42] M. J. Betancourt, M. Girolami, Hamiltonian Monte Carlo for Hierarchical Models (Dec. 2013).

[43] H. Haario, S. E, J. Tamminen, An Adaptive Metropolis Algorithm, Bernoulli 7 (2) (2001) 223–242.

[44] G. O. Roberts, J. S. Rosenthal, Examples of Adaptive MCMC, J. Comput. Graph. Stat. 18 (2) (2009) 349–367.

[45] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, SIAM Rev. 60 (3) (2018) 550–591.

[46] A. Al-Ghosoun, N. Moçayd, M. Seaid, A Surrogate Model for Efficient Quantification of Uncertainties in Multilayer Shallow Water Flows, Environ. Model. Softw. 144 (2021) 105176.

[47] Q. Han, P. Ni, X. Du, H. Zhou, X. Cheng, Computationally Efficient Bayesian Inference for Probabilistic Model Updating with Polynomial Chaos and Gibbs Sampling, Struct. Control Health Monit. 29 (6) (2022) e2936.

[48] P. Robbe, D. Andersson, L. Bonnet, T. A. Casey, M. W. D. Cooper, C. Matthews, K. Sargsyan, H. N. Najm, Bayesian Calibration with Summary Statistics for the Prediction of Xenon Diffusion in UO2 Nuclear Fuel, Comput. Materials Sci. 225 (2023) 112184.

[49] G. A. Meles, M. Amaya, S. Levy, S. Marelli, N. Linde, Bayesian tomography using polynomial chaos expansion and deep generative networks, Geophys. J. Int. 237 (1) (2024) 31–48.

[50] S. A. Orszag, Comparison of Pseudospectral and Spectral Approximation, Stud. Appl. Math. 51 (3) (1972) 253–259.

[51] P. G. Constantine, M. S. Eldred, E. T. Phipps, Sparse Pseudospectral Approximation Method, Comput. Methods Appl. Mech. Eng. 229–232 (2012) 1–12.

[52] J. Reis, O. Le Maître, P. Congedo, P. Mycek, Stochastic Preconditioning of Domain Decomposition Methods for Elliptic Equations with Random Coefficients, Comput. Methods Appl. Mech. Eng. 381 (2021) 113845.

[53] D. Vats, J. M. Flegal, G. L. Jones, Multivariate output analysis for Markov chain Monte Carlo, Biometrika 106 (2) (2019) 321–337.

[54] M. Noble, A. Gesret, N. Belayouni, Accurate 3-D finite difference computation of traveltimes in strongly heterogeneous media, Geophys. J. Int. 199 (3) (2014) 1572–1585.

[55] M. O'Brien, C. Regone, Amoco Tulsa Research Lab (1994).

[56] J. Shewchuk, Triangle: Engineering a 2d Quality Mesh Generator and Delaunay Triangulator, in: M. Lin, D. Manocha (Eds.), Applied Computational Geometry: Towards Geometric Engineering, Vol. 1148 of Lecture Notes in Computer Science, Springer-Verlag, 1996, pp. 203–222, from the First ACM Workshop on Applied Computational Geometry.

AppendixA. Marginal distribution of the coordinates

In Section 4, the approximation $\boldsymbol{\xi} \sim \mathcal{N}(0, \mathrm{I}_r)$ is used for the PC surrogate construction. This choice is justified in the following. The conditional distribution of the coordinates $\boldsymbol{\xi}$ according to the hyperparameters $\boldsymbol{q}$ writes

$$\boldsymbol{\xi}_{|\boldsymbol{q}} \sim \mathcal{N}(0, \Sigma(\boldsymbol{q})) \coloneqq \mathcal{N}_0(\boldsymbol{q}), \quad \boldsymbol{q} \sim \pi_{\mathbb{H}} \tag{A.1}$$

The marginal distribution of the coordinates $\boldsymbol{\xi}$ over the hyperparametric domain is a compound probability distribution. Note that this distribution has no reason to be Gaussian and its exact shape depends on the distribution of the hyperparameters. Here, we want to find the normal distribution $\mathcal{N}_1(\mu, C)$ such that it approximates the best $\boldsymbol{\xi}_{|\boldsymbol{q}}$. That is to say, we want $\mathcal{N}_1$ to be the closest normal distribution in average along $\boldsymbol{q}$ for a given distance. The Kullback–Leibler divergence between two multivariate Gaussian distribution is

$$2D_{\mathrm{KL}}(\mathcal{N}_0(\boldsymbol{q})||\mathcal{N}_1) = \mu^\top C^{-1}\mu + \mathrm{Tr}(C^{-1}\Sigma(\boldsymbol{q})) - \log\frac{\det(\Sigma(\boldsymbol{q}))}{\det(C)} - r, \tag{A.2}$$

where r is the dimension. Denoting $\overline{\Sigma}$ the expectation of $\Sigma(\boldsymbol{q})$ with respect to $\boldsymbol{q}$, we have

$$\overline{\Sigma}_{ij} = \int_{\mathbb{H}} \Sigma(\boldsymbol{q})_{ij}\pi_{\mathbb{H}}(\boldsymbol{q})d\boldsymbol{q} \overset{(20)}{=} (\overline{\lambda}_i\overline{\lambda}_j)^{-1/2}\left\langle\left\langle\int_{\mathbb{H}} k(\boldsymbol{q})\pi_{\mathbb{H}}(\boldsymbol{q})d\boldsymbol{q}, \overline{u}_i\right\rangle, \overline{u}_j\right\rangle \overset{(9)}{=} (\overline{\lambda}_i\overline{\lambda}_j)^{-1/2}\left\langle\left\langle\overline{k}, \overline{u}_i\right\rangle, \overline{u}_j\right\rangle \overset{(10)}{=} \delta_{ij}. \tag{A.3}$$

By linearity and using Eq. (A.3), we get

$$\mathbb{E}_{\mathbb{H}}(\mathrm{Tr}(C^{-1}\Sigma(\boldsymbol{q}))) = \mathrm{Tr}(C^{-1}\mathbb{E}_{\mathbb{H}}(\Sigma(\boldsymbol{q}))) = \mathrm{Tr}(C^{-1}\overline{\Sigma}) = \mathrm{Tr}(C^{-1}). \tag{A.4}$$

So, the $\boldsymbol{q}$-averaged Kullback–Leibler divergence simplifies as

$$\mathbb{E}_{\mathbb{H}}(2D_{\mathrm{KL}}(\mathcal{N}_0(\boldsymbol{q})||\mathcal{N}_1)) = \mu^\top C^{-1}\mu + \mathrm{Tr}(C^{-1}) + \log\det(C) - \mathbb{E}_{\mathbb{H}}(\log\det(\Sigma(\boldsymbol{q}))) - r. \tag{A.5}$$

Since C is symmetric positive definite, μ is set to zero to minimize the divergence. The covariance C^* minimizes therefore

$$C^* = \underset{C\in\mathcal{S}_+^r}{\mathrm{argmin}} \left[\mathrm{Tr}(C^{-1}) + \log\det(C)\right], \tag{A.6}$$

where $\mathcal{S}_+^r$ denotes the set of the symmetric positive definite matrices of size $r \times r$. Furthermore, $C \in \mathcal{S}_+^r$ can be decomposed

$$C = QLQ^{-1}, \tag{A.7}$$

with Q an unitary matrix and L a positive diagonal matrix. Then,

$$\log\det(C) = \sum_{i=1}^{r}\log L_{ii} \quad \text{and} \quad \mathrm{Tr}(C^{-1}) = \mathrm{Tr}(QL^{-1}Q^{-1}) = \sum_{i=1}^{r} L_{ii}^{-1} \tag{A.8}$$

The minimization consists in finding the positive diagonal elements of L, $L_{11}^*, \ldots, L_{rr}^*$ such that

$$L_{11}^*, \ldots, L_{rr}^* = \underset{L_{11},\ldots,L_{rr}}{\mathrm{argmin}} \left[\sum_{i=1}^{r} L_{ii}^{-1} + \log L_{ii}\right]. \tag{A.9}$$

It is a sum of positive terms whose minimum is achieved when all the terms are minimal. It corresponds to $L_{ii} = 1$ for $i = 1, \ldots, r$, leading to

$$C^* = QL^*Q^{-1} = Q\mathrm{I}_rQ^{-1} = \mathrm{I}_r \tag{A.10}$$

To conclude, the normal distribution that minimizes the $\boldsymbol{q}$-averaged Kullback–Leibler divergence to $\mathcal{N}_0(\boldsymbol{q})$ is $\mathcal{N}(0, \mathrm{I}_r)$.