



**HAL**  
open science

# Polynomial surrogates for Bayesian traveltime tomography

Pierre Sochala, Alexandrine Gesret, Olivier Le Maitre

► **To cite this version:**

Pierre Sochala, Alexandrine Gesret, Olivier Le Maitre. Polynomial surrogates for Bayesian traveltime tomography. *GEM - International Journal on Geomathematics*, 2021, 12 (1), 10.1007/s13137-021-00184-0. hal-03384145

**HAL Id: hal-03384145**

**<https://hal.science/hal-03384145v1>**

Submitted on 18 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Polynomial surrogates for Bayesian traveltime tomography

P. Sochala<sup>a,\*</sup>, Alexandrine Gesret<sup>c</sup>, O. Le Maître<sup>c</sup>

<sup>a</sup>*BRGM, 3 avenue Claude Guillemin, 45060 Orléans, France*

<sup>b</sup>*Mines ParisTech, PSL - Research University, Centre de Géosciences, 77300 Fontainebleau, France*

<sup>c</sup>*CMAP, CNRS, Inria, École Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France*

---

## Abstract

This paper tackles the issue of the computational load encountered in seismic imaging by Bayesian traveltime inversion. In Bayesian inference, the exploration of the posterior distribution of the velocity model parameters requires extensive sampling. The computational cost of this sampling step can be prohibitive when the first arrival traveltime prediction involves the resolution of an expensive number of forward models based on the eikonal equation. We propose to rely on polynomial chaos surrogates of the traveltimes between sources and receivers to alleviate the computational burden of solving the eikonal equation during the sampling stage. In an offline stage, the approach builds a functional approximation of the traveltimes from a set of solutions of the eikonal equation corresponding to a few values of the velocity model parameters selected in their prior range. These surrogates then substitute the eikonal-based predictions in the posterior evaluation, enabling very efficient extensive sampling of Bayesian posterior, for instance, by a Markov Chain Monte Carlo (MCMC) algorithm. We demonstrate the potential of the approach using numerical experiments on the inference of two-dimensional domains with layered velocity models and different acquisition geometries (microseismic and seismic refraction contexts). The results show that, in our experiments, the number of eikonal model evaluations required to construct accurate surrogates of the traveltimes is low. Further, an accurate and complete characterization of the posterior distribution of the velocity model is possible, thanks to the generation of large sample sets at a low cost. Finally, we discuss the extension of the current approach to more realistic velocity models and operational situations.

*Keywords:* Inverse problem, Surrogate models, Seismic tomography, Statistical methods

---

## 1. Introduction

The present article focuses on seismic tomography, a method broadly used for imaging the seismic velocity structure of the Earth's subsurface. In traveltime tomography (see [55] for a review) with active sources, a set of sources generates seismic waves propagating in the subsurface and the observations used to infer the velocity model parameters are a set of traveltimes recorded at a network of receivers. The Bayesian inference (see, for instance, [64]) is a robust technique for this type of parameter estimation problem. Previous works on inverse problems in seismic tomography have considered Bayesian inference, see among others [51, 44, 6, 7]. The Bayesian approach estimates the (posterior) distribution of the velocity model parameters by comparing a set of informative observations (measured traveltimes) with the predictions of a forward model, consisting of an eikonal solver in our case.

Compared to an inversion solved with an optimization technique generally based on the computation of the gradient of a misfit function, the advantage of sampling the posterior distribution is threefold. First, it provides detailed information in the form of the full posterior distribution of the velocity model parameters,

---

\*Corresponding author

*Email addresses:* p.sochala@brgm.fr (P. Sochala), alexandrine.gesret@mines-paristech.fr (Alexandrine Gesret), olivier.le-maitre@polytechnique.edu (O. Le Maître)

therefore quantifying the posterior uncertainty in the estimated parameters and subsequent predictions based on this model. In contrast, maximum likelihood, maximum a posteriori, and generic best-fit estimators entirely disregard this information and focus on a single optimal velocity model without characterizing its robustness. Second, Bayesian methods often have intuitive priors that produce physically admissible prior ranges, whereas alternative inverse methods often require ad hoc regularization techniques that have no rigorous justifications. Finally, deterministic inverse methods often lead to the resolution of complex non-convex minimization problems, which can be challenging with computed estimators that are sub-optimal as a result. By aiming at producing the whole posterior distribution, the Bayesian inference is, in practice, computationally much more robust in these situations.

Despite these advantages, the Bayesian approach has to deal with a substantial computational cost to correctly estimate the posterior distributions of the model parameters. The posterior distribution estimation relies on sampling, typically using a Markov-Chain Monte-Carlo (MCMC) method such as the Metropolis-Hastings algorithm [49, 28]. The MCMC methods form a flexible and robust family of algorithms designed to sample distribution, but they suffer from slow convergence, in particular when the parameters of the chains are not well-tuned. The slow convergence requires to run long chains with, as a consequence, multiple evaluations of the forward model and a high computational cost. Several samplers have then been proposed to improve the convergence of MCMC methods (see section 2.3 for more details). Rather than improving the sampling strategy itself, we develop an alternative approach in this paper that directly considers the reduction of the complexity source, the computational cost of the forward model, by relying on surrogate models following [45]. In our context, the surrogate approach consists of building functional approximations of the first arrival traveltimes dependences on the model parameters, that are easy to evaluate but faithful, in order to replace the forward model during the MCMC sampling of the posterior. More precisely, we first construct Polynomial Chaos (PC) surrogates of the traveltimes, which have negligible evaluation cost, and use them in the MCMC algorithm in place of the solution of the eikonal equation. The substitution is innocuous, provided that the surrogate error is small [46].

Uncertainty quantification methods routinely use surrogate models to emulate complex forward models and enable their extensive sampling in order to characterize the impact of parametric uncertainties and perform global sensitivity analysis. Many works over the last decades focused on the development of accurate surrogate models from minimal size ensemble of model simulations, and these advances have triggered new application domains, such as the surrogate-based resolution of optimization, calibration, and inverse problems. In this work, we consider Polynomial Chaos (PC) expansions to approximate the functional dependencies between a quantity of interest and the random parameters [24, 39]. The PC expansions have been successfully applied to multiple problems in many application areas and found particularly effective for situations where the quantity of interest has smooth dependencies. Previous examples of applications in geoscience of PC expansions are, for instance, in [22, 32, 41, 61] (uncertainty quantification) and [20, 25, 52, 14] (inverse or calibration problems). To our knowledge, the use of PC surrogates for seismic tomography is new to the present work, but we stress that alternative surrogate construction could be employed [56, 17].

The structure of the paper is as follows. Section 2 presents the inversion framework with the Bayesian formulation of the problem, the discrepancy model, and the eikonal solver. Section 3 describes the polynomial chaos expansions used to construct the surrogates of the traveltimes. We assess the methodology for different problems with layered velocity models and different acquisition geometries: Section 4 concerns microseismic acquisition while Section 5 considers seismic refraction acquisition. Discussion and conclusion of the present work are reported in Section 6.

## 2. Framework

One crucial question faced in seismic tomography concerns the representation of the wave-velocity fields, which typically are infinite-dimensional. In this respect, one can distinguish two distinct approaches. The first set of approaches uses an explicit discretization of the velocity model with a fixed and finite number of parameters. In contrast, the second set avoids the explicit specification of the velocity model and imbeds the identification of its form within the inference process.

Explicit parametrizations are well suited to simple situations where the model is naturally specified using a limited number of unknown parameters, as considered below. More general situations, such as smooth velocity models, use truncated optimal decomposition of the a priori velocity model. We mention, in particular, the Karhunen-Loeve (KL) decomposition [35, 42] (also known as Principal Component Analysis, or Principal Orthogonal Decomposition). The inference problem using explicit finite parametrizations is recast in terms of the posterior distribution of these parameters, leading to the sampling of a finite-dimensional joint-distribution.

Alternative approaches do not specify a priori the structure of the velocity model, which evolves along the inference process. An example of these approaches is the so-called transdimensional method that can be applied with different parametrizations such as Voronoi cells [6], Delaunay tessellations [29], Johnson–Mehl tessellations [3] and wavelets [30]. The number of centroids, their location velocity value, are jointly evolving during the posterior sampling. These approaches are appropriate in situations corresponding to velocity models with complex structures hardly amenable to a simple parametrization.

In this work, we restrict ourselves to explicit parametrizations because the fixed form is essential to the practical construction of surrogate models. This paper considers only situations with trivial parametrizations in order to assess the potentiality of the approach. It should be mentioned that such layered velocity models are often used in the context of microseismic monitoring of hydraulic fracturing projects where they are constructed from a sonic log, which is then usually calibrated with perforation shots from known locations [19, 47]. We delay to future work the extension of the approach to more complex situations requiring larger (higher-dimensional) parametrizations, and possibly the introduction of hyper-parameters [62].

## 2.1. Bayesian formulation

### 2.1.1. Bayes' rule

The velocity model is assumed to be described by a random vector of  $M$  independent parameters denoted  $\mathbf{m} \in L^2(\mathbb{R}^M)$ . In the case of a layered planar geological structure,  $\mathbf{m}$  contains the velocities and the depths of the layer boundaries. We denote  $\mathbf{t}_{\text{obs}} \in L^2(\mathbb{R}_+^N)$  the vector of  $N$  measured traveltimes used to infer  $\mathbf{m}$  using the Bayes' rule [2]. In its simplest form, the Bayes' rule reads

$$\pi(\mathbf{m}|\mathbf{t}_{\text{obs}}) = \frac{\mathcal{L}(\mathbf{t}_{\text{obs}}|\mathbf{m})\pi_m(\mathbf{m})}{\mathcal{L}(\mathbf{t}_{\text{obs}})}, \quad (1)$$

where  $\pi_m(\mathbf{m})$  is the prior distribution of the parameters,  $\mathcal{L}(\mathbf{t}_{\text{obs}}|\mathbf{m})$  the likelihood of the measured traveltimes,  $\pi(\mathbf{m}|\mathbf{t}_{\text{obs}})$  the posterior distribution of  $\mathbf{m}$ , and  $\mathcal{L}(\mathbf{t}_{\text{obs}})$  the marginal likelihood (or evidence) of the observations. In practice, the evidence is just a normalization constant, ensuring that the posterior is a density, and it can remain undetermined, writing the proportionality between the posterior and the product of the likelihood of the observations and the prior of  $\mathbf{m}$ . The likelihood  $\mathcal{L}(\mathbf{t}_{\text{obs}}|\mathbf{m})$  gives the probability of the observed traveltimes given a particular value  $\mathbf{m}$  of the model parameters. It is classically a function of the discrepancies between measurements and the model predictions. Its derivation is now detailed.

### 2.1.2. Likelihood function

In the following we denote  $\mathbf{t}(\mathbf{m})$  the vector of traveltimes predicted for the velocity model with parameters  $\mathbf{m}$ . The differences between the measured and predicted traveltimes has two contributions,

$$\mathbf{t}_{\text{obs}} - \mathbf{t}(\mathbf{m}) = \mathbf{e}_{\text{mod}} + \mathbf{e}_{\text{obs}}, \quad (2)$$

the model error  $\mathbf{e}_{\text{mod}}$  and the measurement (or observation) error  $\mathbf{e}_{\text{obs}}$ . The error  $\mathbf{e}_{\text{mod}}$  is rooted in the simplifications of physics and the various misspecifications of the numerical model while, in the tomography context, the measurement errors typically arise from the manual picking of the traveltimes ([5]). As discussed in the introduction, we shall substitute the forward model predictions of the traveltimes  $\mathbf{t}(\mathbf{m})$  with surrogate approximations  $\tilde{\mathbf{t}}(\mathbf{m})$ , introducing another source of discrepancy  $\mathbf{e}_{\text{sur}}$ , leading to

$$\mathbf{t}_{\text{obs}} = \tilde{\mathbf{t}}(\mathbf{m}) + \boldsymbol{\epsilon}, \quad (3)$$

where  $\boldsymbol{\epsilon} = \mathbf{e}_{\text{obs}} + \mathbf{e}_{\text{mod}} + \mathbf{e}_{\text{sur}}$  is the *discrepancy* vector between the observations and the surrogate predictions.

In practice, the discrepancy vector is unknown, and one needs to rely on a statistical model. Hereafter we assume that the observation error  $\mathbf{e}_{\text{obs}}$  dominates the other sources of uncertainty. In that case,  $\boldsymbol{\epsilon}$  amounts to the manual picking errors which can be considered unbiased and independent from an observation to another. In the numerical experiments below, we rely on synthetic observations consisting of eikonal solutions corrupted by centred independent Gaussian noise with diagonal covariance matrix  $\mathbf{C}_{\text{obs}} \in \mathbb{R}^{N,N}$  defined as

$$[\mathbf{C}_{\text{obs}}(\alpha)]_{i,j} = \begin{cases} (\alpha t_i)^2 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \quad (4)$$

where  $t_i$  denotes the  $i$ th travel time and  $\alpha \in \mathbb{R}_+$  controls the noise level. In other words, the manual picking error is Gaussian with a variance proportional to the true traveltime. We shall assume that the noise level is not known and we extend the the Bayes' rule to

$$\pi(\mathbf{m}, \alpha | \mathbf{t}_{\text{obs}}) \propto \mathcal{L}(\mathbf{t}_{\text{obs}} | \mathbf{m}, \alpha) \pi_m(\mathbf{m}) \pi_\alpha(\alpha), \quad (5)$$

where  $\pi_\alpha$  is the prior of the hyper-parameter  $\alpha$ . Neglecting the surrogates and model error contributions, the misfit function  $\chi(\mathbf{m}, \alpha | \mathbf{t}_{\text{obs}})$  corresponds to the (squared) Mahalanobis distance,

$$\chi(\mathbf{m}, \alpha) = (\mathbf{t}_{\text{obs}} - \tilde{\mathbf{t}}(\mathbf{m}))^\top \mathbf{C}_{\text{obs}}(\alpha)^{-1} (\mathbf{t}_{\text{obs}} - \tilde{\mathbf{t}}(\mathbf{m})). \quad (6)$$

Finally, the Gaussian likelihood function of the Bayes' rule is expressed as

$$\mathcal{L}(\mathbf{t}_{\text{obs}} | \mathbf{m}, \alpha) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}_{\text{obs}}(\alpha)|}} \exp\left(-\frac{\chi(\mathbf{m}, \alpha)}{2}\right). \quad (7)$$

where  $|\mathbf{C}_{\text{obs}}|$  denotes the determinant of  $\mathbf{C}_{\text{obs}}$ . Using the likelihood function (7), the maximum likelihood estimator of  $\mathbf{m}$  corresponds to the solution of the (weighted) least-squares fit problem, minimizing  $\|\mathbf{t}_{\text{obs}} - \tilde{\mathbf{t}}(\mathbf{m})\|_{\mathbf{C}_{\text{obs}}^{-1}}$  (the norm induced by the inverse of  $\mathbf{C}_{\text{obs}}$ ).

The observation noise distribution will be unknown in real-life applications for which the model error is not necessarily negligible. In such a case, one has to introduce an appropriate treatment of model error (see for instance [36]) and identify the likelihood structure involving possibly several hyper-parameters. For instance, symmetric or asymmetric Laplace distributions may be more representative of the operator errors in the picking process [67], while more general covariance matrix  $\mathbf{C}_{\text{obs}}$  with a limited number  $n_\alpha \ll N$  of hyper-parameters to be inferred can better fit for the discrepancy between the predictions and the observations.

The Bayesian inference result depends on the choice of the likelihood [8] and the approach proposed here is no exception. However, a key advantage of our surrogates approach is that one can use the surrogates to test different forms of likelihood. Indeed, the construction of the surrogates will demand most of the computational effort, while subsequent MCMC sampling has a low computational cost. As a result, our approach is suitable to compare several likelihoods and determine the best ones from objective criteria such as Bayes factors. Similarly, the capacity to sample at a low computational cost the posterior enables advanced analyses such as for the posterior sensitivity on the priors [40].

### 2.1.3. Prior distributions

We assume that the velocity model parameters are a priori independent, such that the prior  $\pi_m(\mathbf{m})$  factorizes to

$$\pi_m(\mathbf{m}) = \prod_{i=1}^M \pi_i(m_i), \quad (8)$$

where  $\pi_i(m_i)$  is the prior of the parameter  $m_i$ . The particular form of these priors will be detailed in the test cases sections. Regarding the hyper-parameter  $\alpha$  of the covariance matrix, we rely on a Jeffreys' prior,

$$\pi_\alpha(\alpha) \propto \begin{cases} \alpha^{-1} & \text{if } \alpha > 0, \\ 0 & \text{if } \alpha \leq 0. \end{cases} \quad (9)$$

The Jeffreys prior [33, 34] is adapted to situations with no a priori information on the scale (order of magnitude) of the parameter, since it is invariant under any monotone transformation (or re-parametrization through change of coordinates). By using Eq. (4) to Eq. (9), the posterior distribution finally reads

$$\pi(\mathbf{m}, \alpha | \mathbf{t}_{\text{obs}}) \propto \frac{1}{(2\pi)^{N/2} \alpha^{N+1}} \prod_{i=1}^N \frac{1}{|t_{\text{obs},i}|} \exp \left[ -\frac{1}{2} \left( \frac{t_{\text{obs},i} - \tilde{t}_i(\mathbf{m})}{\alpha t_{\text{obs},i}} \right)^2 \right]. \quad (10)$$

## 2.2. Eikonal solver

As explained below, we shall rely on a non-intrusive method to construct the surrogates  $\tilde{\mathbf{t}}(\mathbf{m})$  of the  $N$  traveltimes. This step requires the computation of the forward model  $\mathbf{t}(\mathbf{m})$  at selected points  $\mathbf{m}$  within the a priori range. The  $i$ th component  $t_i(\mathbf{m})$  of  $\mathbf{t}(\mathbf{m})$  corresponds to the first arrival traveltime from the source point  $\mathbf{x}_{\text{src},i}$  to reach the receiver point  $\mathbf{x}_{\text{rec},i}$ . The two points are in a domain  $\Omega \subseteq \mathbb{R}^d$  where  $d$  is the spatial dimension of the domain. The velocity model  $v(\mathbf{x}|\mathbf{m}) \in \mathbb{R}_+$  describes the wave propagation velocity at all  $\mathbf{x} \in \Omega$ . With this notation, we define the eikonal problem associated to the  $i$ th source point by

$$\begin{cases} \|\nabla \tau_i(\mathbf{x}|\mathbf{m})\|_2^2 = \frac{1}{v(\mathbf{x}|\mathbf{m})^2}, & \forall \mathbf{x} \in \Omega, \\ \tau_i(\mathbf{x}|\mathbf{m}) = 0, & \text{for } \mathbf{x} = \mathbf{x}_{\text{src},i}. \end{cases} \quad (11)$$

The solution  $\tau_i$  gives the time it takes for the wave with source  $\mathbf{x}_{\text{src},i}$  to reach any point  $\mathbf{x} \in \Omega$  in a medium with velocity  $v(\cdot|\mathbf{m})$ . The traveltime to the receiver is then simply given by  $t_i(\mathbf{m}) = \tau_i(\mathbf{x}_{\text{rec},i}|\mathbf{m})$ .

The first-order nonlinear partial differential equation (11) can be numerically solved by different approaches, including finite difference schemes [65], fast marching method [58], fast sweeping method [68], and discontinuous Galerkin scheme [38]. In this paper, we employ a numerical method combining a spherical wave approximation in the source neighborhood and a plane wave approximation farther away [53]. More precisely, the approach uses several local finite difference operators and a global fast sweeping method. This approach allows treating velocity models with sharp vertical and horizontal contrasts accurately. Naturally, when the domain of interest is instrumented with a set of  $N_s$  sources and  $N_p$  receivers, the eikonal equation is solved  $\min(N_s, N_p)$  times (using the principle of traveltime reciprocity if  $N_s > N_p$  to reduce the computational cost).

## 2.3. MCMC sampling

The posterior distribution in (10) does not have a closed-form expression and must be sampled. The MCMC algorithms are efficient to draw samples from complicated distributions. From the original Metropolis-Hasting (MH) algorithm [28], several improvements have been introduced over the years. These improvements include, in particular, the Hamiltonian (or hybrid) MC [16, 21], the Langevin MC [4], the adaptive MCMC [26, 57], and the parallel MC methods [66]. All of these methods can be applied to our surrogate-based approach, retaining all of its computational benefits. Our surrogates being differentiable, it is well suited to the Hamiltonian and Langevin MC methods which use the gradient of the posterior when the estimation of the gradients of the traveltimes may be delicate and costly to compute using the forward model (eikonal equation) only.

For the sake of simplicity, we restrict here the exposition to the MH algorithm and denote  $(\mathbf{m}, \alpha)_k$  the current state of the chain. At each step, the same procedure is repeated. First, a new state  $(\mathbf{m}, \alpha)_*$  is proposed at random from a transition probability  $\pi_{\text{tr}}((\mathbf{m}, \alpha)_* | (\mathbf{m}, \alpha)_k)$ . Second, the new state is accepted with a probability  $r$  given by

$$r = \min \left( 1, \frac{\pi(\mathbf{m}_*, \alpha_* | \mathbf{t}_{\text{obs}}) \pi_{\text{tr}}((\mathbf{m}, \alpha)_* | (\mathbf{m}, \alpha)_k)}{\pi(\mathbf{m}_k, \alpha_k | \mathbf{t}_{\text{obs}}) \pi_{\text{tr}}((\mathbf{m}, \alpha)_k | (\mathbf{m}, \alpha)_*)} \right). \quad (13)$$

If the new state is accepted, we set  $(\mathbf{m}, \alpha)_{k+1} = (\mathbf{m}, \alpha)_*$ , otherwise  $(\mathbf{m}, \alpha)_{k+1} = (\mathbf{m}, \alpha)_k$ . When the transition probability is symmetric, the proposed state is always accepted if it yields a higher posterior, while accepting

a state with a much lower posterior value is unlikely. Using this simple random rule,  $(\mathbf{m}, \alpha)_k$  follows asymptotically for  $k \rightarrow \infty$  the posterior distribution. Therefore, the successive states of the chain, after large enough  $k$ , form samples set drawn from the posterior. The improvement listed below aims at improving the mixing properties and minimizing the chain's correlation. Indeed, determining an (explicit or implicit) transition probability function  $\pi_{\text{tr}}$  is critical to obtain an effective MCMC method. Without loss of generality, we use an adaptive Metropolis algorithm with symmetric Gaussian proposal distributions,

$$\pi_{\text{tr}}((\mathbf{m}, \alpha)_* | (\mathbf{m}, \alpha)_k) = \dots \frac{1}{|2\pi\mathbf{C}_{\text{tr}}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{m}_k - \mathbf{m}_*, \alpha_k - \alpha_*)^\top \mathbf{C}_{\text{tr}}^{-1}(\mathbf{m}_k - \mathbf{m}_*, \alpha_k - \alpha_*)\right), \quad (14)$$

where the transition covariance  $\mathbf{C}_{\text{tr}}$  is sequentially updated during the initial **burn-in** stage of the chain, using information from previous steps of the chain [26].

The computational time associated with the MCMC procedure is proportional to the (averaged) computational time of the forward model with the length  $\mathcal{M}$  of the chain necessary for the exploration of the posterior. The actual value of  $\mathcal{M}$  depends on the problem considered and the particular MCMC method employed. However, we expect  $\mathcal{M}$  to vary from  $10^4$  to  $10^6$ , or even more than that for very challenging problems. Consequently, MCMC is feasible only if the forward model is fast enough. If the eikonal solver is used to make all predictions, with average computational cost  $T_{\text{eik}}$ , the resulting computational times  $\mathcal{M} \times T_{\text{eik}}$  is quickly prohibitive. In contrast, using the surrogates' prediction, the sampling time becomes  $\mathcal{M} \times T_{\text{sur}}$  where  $T_{\text{sur}} \ll T_{\text{eik}}$ . The complete complexity analysis has to account for the cost of constructing the surrogates. As discussed below, the construction time is proportional to the number of  $N_q$  of the eikonal solutions involved in the procedure, so the computational complexity of the surrogate approach is  $N_q \times T_{\text{eik}} + \mathcal{M} \times T_{\text{sur}}$ , which remains much less than the cost of the model-based approach provided that  $N_q \ll \mathcal{M}$ . The latter condition will be satisfied if the dependencies with  $\mathbf{m}$  of the traveltimes are not too complex, and the dimensionality of  $\mathbf{m}$  is not too high.

### 3. Surrogate Models

In this section, we introduce the PC expansions of the traveltimes  $\mathbf{t}(\mathbf{m})$ . We rely on non-intrusive methods, meaning that the construction of the PC expansions uses a set of deterministic solutions of the eikonal equation for selected velocity models with parameters  $\mathbf{m}$ . In the following, we denote  $\Xi \subseteq \mathbb{R}^M$  the a priori range of  $\mathbf{m}$ . As mentioned before, we restrict ourselves to the case where the parameters in  $\mathbf{m}$  are mutually independent, such that their prior can be factorized to (8). Similarly, the range  $\Xi$  is the product of the ranges  $\Xi_i$  of each velocity parameter  $m_i$ .

#### 3.1. Polynomial Chaos expansion

For notational convenience, let  $f(\mathbf{m})$  be one of the traveltime to be approximated. We assume that  $f(\mathbf{m}) \in L^2(\Xi, \pi_m)$ , where  $L^2(\Xi, \pi_m)$  is the space of second order functions in  $\mathbf{m}$ :

$$f(\mathbf{m}) \in L^2(\Xi, \pi_m) \quad \Leftrightarrow \quad \mathbb{E}(f^2) = \int_{\Xi} f^2(\mathbf{m})\pi_m(\mathbf{m})d\mathbf{m} < \infty. \quad (15)$$

We shall equip  $L^2(\Xi, \pi_m)$  with the inner product coinciding with the correlation and denoted with brackets:

$$\langle f, f' \rangle \doteq \mathbb{E}(f(\mathbf{m})f'(\mathbf{m})) = \int_{\Xi} f(\mathbf{m})f'(\mathbf{m})\pi_m(\mathbf{m})d\mathbf{m}. \quad (16)$$

Following [9], any second-order random function in  $\mathbf{m}$  has a convergent spectral expansion of the form

$$f(\mathbf{m}) = \sum_k f_k \Psi_k(\mathbf{m}), \quad (17)$$

where the  $f_k \in \mathbb{R}$  are the expansion coefficients or stochastic modes of  $f$ , and  $\{\Psi_k\}$  is a complete orthogonal set constituting a basis of  $L^2(\Xi, \pi_m)$ . Generalized PC expansions result when selecting random polynomials in  $\mathbf{m}$  for the basis of  $L^2(\Xi, \pi_m)$ . To construct this basis, we denote for  $i = 1, \dots, M$  the orthonormal set of uni-variate polynomials  $\{\phi_k^i(m_i), k = 0, 1, 2, \dots\}$ , where  $\phi_k^i$  is a polynomial of degree  $k$  in  $m_i$ . The orthonormal character implies that

$$\mathbb{E}(\phi_k^i \phi_l^i) = \int_{\Xi_i} \phi_k^i(m_i) \phi_l^i(m_i) \pi_i(m_i) dm_i = \delta_{k,l}, \quad (18)$$

when  $m_i$  has a uniform (resp. Gaussian) distribution, the  $\phi_k^i$  are scaled Legendre (resp. Hermite) polynomials [1].

Then, introducing the integer-valued multi-index  $\mathbf{k} = (k_1, \dots, k_M) \in \mathbb{N}^M$ , we define the  $M$ -variate polynomial  $\Psi_{\mathbf{k}}(\mathbf{m})$  by

$$\Psi_{\mathbf{k}}(\mathbf{m}) = \prod_{i=1}^M \phi_{k_i}^i(m_i). \quad (19)$$

One can easily check that the polynomials  $\Psi_{\mathbf{k}}(\mathbf{m})$  are orthonormal:

$$\langle \Psi_{\mathbf{k}}, \Psi_{\mathbf{l}} \rangle = \int_{\Xi} \Psi_{\mathbf{k}}(\mathbf{m}) \Psi_{\mathbf{l}}(\mathbf{m}) \pi_m(\mathbf{m}) d\mathbf{m} = \prod_{i=1}^M \int_{\Xi_i} \phi_{k_i}^i(m_i) \phi_{l_i}^i(m_i) \pi_i(m_i) dm_i = \delta_{\mathbf{k}, \mathbf{l}}. \quad (20)$$

The total degree of the multi-variate polynomial  $\Psi_{\mathbf{k}}$  is  $\|\mathbf{k}\|_1$  while the maximum degree is  $\|\mathbf{k}\|_{\infty}$ .

The *truncated* PC expansion of  $f$  is then defined using a finite set  $\mathcal{K}$  of multi-indices in the expansion:

$$f(\mathbf{m}) \approx f^{\mathcal{K}}(\mathbf{m}) := \sum_{\mathbf{k} \in \mathcal{K}} f_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{m}), \quad (21)$$

In the following, we denote  $|\mathcal{K}|$  the size of the PC basis. The truncation error  $f(\mathbf{m}) - f^{\mathcal{K}}(\mathbf{m})$  depends on the truncation strategy and decreases when  $\mathcal{K} \subset \mathbb{N}^M$  increases. In fact, we have

$$\mathbb{E}(|f - f^{\mathcal{K}}|^2) = \sum_{\mathbf{k} \in \mathbb{N}^M \setminus \mathcal{K}} f_{\mathbf{k}}^2. \quad (22)$$

In practice, a convergence analysis is mandatory to verify that the truncation error is sufficiently small.

### 3.2. Computation of the spectral coefficients

Several methods are available to compute the spectral coefficients of the PC expansion. In this work, we use two non-intrusive approaches in order to exploit their respective advantages depending on the situation. The first method, the sparse pseudo-spectral projection, is very efficient for smooth functions but suffers in the presence of localized features (such as loss of differentiability, discontinuity or plateau) that tend to plague the whole approximation. In contrast, the least-squares regression is usually less accurate for the same computational complexity, but is more robust and offers more flexibility in the selection of the evaluation points. Indeed, the sample set of the projection approach consists of nested sparse grids whereas the least-squares regression can rely on more general set of points.

#### 3.2.1. Pseudo-Spectral Projection

A direct consequence of the orthogonal expansion in (21) is that the *deterministic* expansion coefficients  $f_{\mathbf{k}}$  are defined by

$$f_{\mathbf{k}} = \langle f, \Psi_{\mathbf{k}} \rangle, \quad \forall \mathbf{k} \in \mathbb{N}^M. \quad (23)$$

Using (23), the non-intrusive spectral projection method approximates the  $M$ -dimensional integral involved in the inner-product definition (see (16)) by a deterministic quadrature formula with  $N_q$  points,

$$f_{\mathbf{k}} \approx \mathcal{Q}_{N_q}(f \phi_{\mathbf{k}}) = \sum_{q=1}^{N_q} w^{(q)} f(\mathbf{m}^{(q)}) \Psi_{\mathbf{k}}(\mathbf{m}^{(q)}),$$



where  $\mathbf{m}^{(q)} \in \Xi$  and  $w^{(q)}$  are the quadrature nodes and weights of the formula. We denote  $\mathcal{X} = \{\mathbf{m}^{(q)}, q = 1, \dots, N_q\}$  the set of nodes. The quadrature rule must be accurate enough to provide a correct estimate of the expansion coefficients. An important condition that is generally required is that  $\mathcal{Q}_{N_q}$  yields the discrete orthogonality of the polynomials, that is

$$Q_{N_q}(\Psi_{\mathbf{k}}\Psi_{\mathbf{l}}) = \delta_{\mathbf{k},\mathbf{l}}, \quad \forall \mathbf{k}, \mathbf{l} \in \mathcal{K}. \quad (24)$$

This condition ensures a projection which is free of any internal aliasing.

A straightforward approach to constructing such quadrature rules consists of the tensorization of high-order one-dimensional quadrature formulas adapted to the measure  $\pi_i$ . Typical examples are the Gauss and Clenshaw–Curtis formulas for Gaussian and uniform distributions. For the first test case (see section 4), we use Féjèr’s quadrature formula with nodes corresponding to the zeroes of the Chebyshev polynomials. Full tensorization of the one-dimensional formulas is usually not an option as the number  $N_q$  of quadrature nodes would increase exponentially with the dimension  $M$  of the integration. Sparse tensorizations following the Smolyak’s formula [59] are necessary if  $M$  exceeds few units, leading to so-called Sparse Grid Quadrature (SGQ) methods [23]. One drawback of the SGQ is that some of the weights are negative. Moreover, using the same SGQ for the projection on all polynomials  $\Psi_{\mathbf{k}}, \mathbf{k} \in \mathcal{K}$  is not optimal, and adapting the quadrature rule to each mode yields better results. The Pseudo-Spectral Projection (PSP) method [12, 11] achieves this adaptation by applying the Smolyak’s formula on sequences on nested projection operators, rather than on sequences of quadrature rules. The PSP procedure results in mode dependent weights in the quadrature formulas for the coefficients, which can be recast in

$$f_{\mathbf{k}} \approx \sum_{q=1}^{N_q(\ell)} w_{\mathbf{k}}^{(q)} f(\mathbf{m}^{(q)}) \Psi_{\mathbf{k}}(\mathbf{m}^{(q)}). \quad (25)$$

For the same sparse grid, the PSP method leads to a larger multi-index set  $\mathcal{K}$  satisfying (24), compared to the SGQ method. In practice, the PSP method uses a level index  $\ell \in \mathbb{N}_+$  to control the size  $N_q(\ell)$  of the sparse grid  $\mathcal{X}(\ell)$ . The corresponding set  $\mathcal{K}(\ell)$  of polynomial multi-indices satisfying (24) is explicitly fixed by the construction. In the remainder of the paper, we rely on nested quadrature rules which ensure nested sparse grids,  $\mathcal{X}(\ell) \subset \mathcal{X}(\ell+1)$ , and multi-index sets,  $\mathcal{K}(\ell) \subset \mathcal{K}(\ell+1)$ . Besides the computational complexity aspect, the nested character of PSP is interesting to monitor the convergence of the PC approximation with increasing  $\ell$ .

### 3.2.2. Ordinary Least Squares Regression

As an alternative to the PSP method, we will also use the Ordinary Least Squares (OLS) regression. In this approach, the expansion coefficients on a prescribed basis are obtained by the resolution of a regression problem. Let  $\mathcal{X} = \{\mathbf{m}^{(1)}, \dots, \mathbf{m}^{(N_q)}\}$  be a samples set of model parameters, each in  $\Xi$ . For example, the sample set can be generated by Monte Carlo method or Latin Hypercube Sampling [48]. For the second test case (see section 5), we use a quasi-Monte Carlo method with a standard Halton sequence [50]. In the OLS, one defines the PC coefficients as the minimizers of the following sum of squares:

$$LS(\{f_{\mathbf{k}}, \mathbf{k} \in \mathcal{K}\}) = \sum_{q=1}^{N_q} \left| f(\mathbf{m}^{(q)}) - \sum_{\mathbf{k} \in \mathcal{K}} f_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{m}^{(q)}) \right|^2.$$

Denoting  $\mathbf{f} \in \mathbb{R}^{|\mathcal{K}|}$  the vector of PC coefficients  $f_{\mathbf{k}}$  and  $\mathbf{y} \in \mathbb{R}^{N_q}$  the vector of traveltimes  $f(\mathbf{m}^{(q)})$ , the LS solution satisfies the system of normal equations

$$\mathbf{Z}^{\top} \mathbf{Z} \mathbf{f} = \mathbf{Z}^{\top} \mathbf{y}, \quad (26)$$

where  $\mathbf{Z} \in \mathbb{R}^{N_q \times |\mathcal{K}|}$  is the matrix of regressors  $\Psi_{\mathbf{k}}(\mathbf{m}^{(q)})$ . Contrary to the PSP case, where  $\mathcal{K}$  is fixed by the choice of level, the OLS allows to choose  $\mathcal{K}$  independently of  $\mathcal{X}$ . Most often, one fixes  $\mathcal{K}$  by deciding a maximal degree  $d^{\circ}$  in the polynomial expansion. For the total degree truncation, that is using

$$\mathcal{K}(d^{\circ}) = \{\mathbf{k} \in \mathbb{N}^M, \sum_i k_i \leq d^{\circ}\}, \quad (27)$$

the dimension of the basis is

$$|\mathcal{K}(d^\circ)| = \frac{(M + d^\circ)!}{M!d^\circ!}.$$

However, the choice of  $d^\circ$  affects the conditioning of system (26), which can degrade if the samples set  $\mathcal{X}$  is not rich enough or adapted to  $\mathcal{K}$ . This is not the case for the present application, but if such a situation is suspected, one can rely on validation strategies (leave-one-out or k-fold cross-validation) to assess the stability of the regression problem and to prevent overfitting. Another option to avoid ill-conditioned matrices is to implement a weighted least squares method with a coherence-optimal sampling [27, 10].

### 3.3. Statistical information

The spectral expansion on orthonormal bases allows the direct estimation of the first statistical moments of  $f$  and more involved statistics such as the sensitivity indices. For instance, exploiting the orthogonality of the PCs, the mean  $\mathbb{E}(f)$  and variance  $\mathbb{V}(f)$  are estimated by

$$\mathbb{E}(f) \approx \mathbb{E}(f^\mathcal{K}) = \langle f^\mathcal{K}(\mathbf{m}), 1 \rangle = f_{\mathbf{0}}, \quad (28)$$

$$\mathbb{V}(f) \approx \mathbb{V}(f^\mathcal{K}) = \langle f^\mathcal{K}(\mathbf{m}) - \mathbb{E}(f^\mathcal{K}), f^\mathcal{K}(\mathbf{m}) - \mathbb{E}(f^\mathcal{K}) \rangle = \sum_{\mathbf{k} \in \mathcal{K} \setminus \mathbf{0}} f_{\mathbf{k}}^2, \quad (29)$$

where  $\mathbf{0}$  is the multi-index with all its indexes equal to zero.

The decomposition of the variance is another useful information deriving directly from the spectral expansion that allows useful global sensitivity analyses [63, 13]. Following the Sobol–Hoeffding decomposition [31], variance-based global sensitivity analysis identifies the contribution of each input parameter in  $\mathbf{m}$ , or group of input parameters, to the variance of the outputs through the definition of the so-called sensitivity indices [60]. The most common indices are the first-order indices, that measure the isolated contribution of a subset of parameters, and the total order indices that also include possible joint effects with the remaining parameters. As an example, the first-order index  $S_i$  associated with the  $i$ th parameter is given by the following formula,

$$S_i(f^\mathcal{K}) := \frac{V_{S_i}}{\mathbb{V}(f^\mathcal{K})} = \frac{\sum_{\mathbf{k} \in \mathcal{K}_i} f_{\mathbf{k}}^2}{\sum_{\mathbf{k} \in \mathcal{K} \setminus \mathbf{0}} f_{\mathbf{k}}^2}, \quad (30)$$

where  $V_{S_i}$  is the partial variance due to the  $i$ th parameter (only) and is computed from the following set of multi-indices  $\mathcal{K}_i = \{\mathbf{k} \in \mathcal{K}, k_{j=i} > 0, k_{j \neq i} = 0\}$ . In the results section, we shall exploit the first-order sensitivity indices to interpret and explain the differences in the information gain on the different parameters of the velocity model.

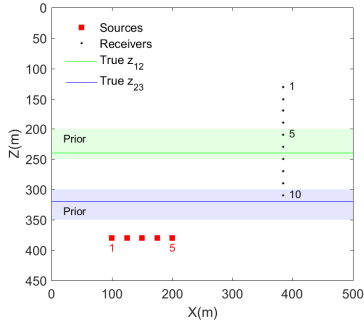
## 4. Microseismic test case

### 4.1. Test case description

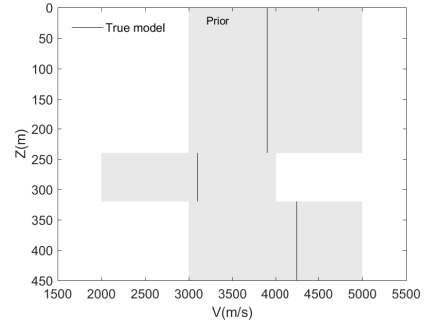
We first consider an idealized two dimensional medium,  $\Omega = [0, 500] \times [0, 450]$  m<sup>2</sup>, with three horizontal geological layers having constant wave propagation velocities. The velocity model has  $M = 5$  parameters: the velocities in the top ( $v_1$ ), central ( $v_2$ ) and bottom ( $v_3$ ) layers as well as the locations  $z_{12}$  and  $z_{23}$  of the interfaces between the layers,

$$\mathbf{m} = (v_1, v_2, v_3, z_{12}, z_{23}). \quad (31)$$

The observations stem from a microseismic acquisition geometry with five sources and ten receivers as represented in the left plot of Fig. 1. In total, the inference will use  $N = 5 \times 10 = 50$  traveltimes. We generate synthetic observations solving the eikonal equation for particular values of the parameters referred to as  $\mathbf{m}_{\text{true}}$ . These numerical traveltimes are subsequently corrupted, adding independent Gaussian noise to simulate the measurement errors. Therefore, the inference problem involves no model error but only the surrogate errors and measurement noise. Regarding the prior distribution of  $\mathbf{m}$ , we consider independent uniform distributions for all the parameters, with ranges encompassing  $\mathbf{m}_{\text{true}}$ . Table 1 reports the prior range for  $\mathbf{m}$  and the values of  $\mathbf{m}_{\text{true}}$ , while Figure 1 depicts the prior and true velocity models. For the eikonal



(a) Sources and receivers layout and a priori ranges of the interface locations.



(b) A priori velocity ranges and true interface location  $z_{12}$  and  $z_{23}$ .

Figure 1: Schematical description of the microseismic test case.

Parameter	$v_1$ (m/s)	$v_2$ (m/s)	$v_3$ (m/s)	$z_{12}$ (m)	$z_{23}$ (m)
Prior range	$[3, 5] \times 10^3$	$[2, 4] \times 10^3$	$[3, 5] \times 10^3$	$[200, 250]$	$[300, 350]$
True value	3,900	3,100	4,250	240	320

Table 1: Microseismic test case: a priori ranges and true values of the velocity model parameters.

equation, the solver uses a resolution of 1 meter, leading to a mesh with 225,000 square cells. This mesh is not conforming with the (uncertain) interfaces between the layers. The velocity of cells intersecting with an interface is defined as the average velocity equal to the average slowness using a logarithmic transformation of the variable.

## 4.2. Surrogates analysis

### 4.2.1. PC surrogates validation

For this test case, the PSP method (see section 3.2.1) is used to construct the PC expansions of the traveltimes. Table 2 reports the number  $N_q$  of PSP points (corresponding to the number of eikonal solver evaluations) and the dimension of the PC basis  $P$ , for the PSP levels  $\ell = 2, 4$  and 6. Before substituting the eikonal solver with the surrogates of the traveltimes in the Bayesian inference, we assess the PC approximations' accuracy and predictive capability. To estimate the surrogates' error, we randomly generate a validation set of  $N_v = 500$  velocity model parameters  $\mathbf{m}^{(k)} \in \Xi$ , relying on a Latin Hypercube Sampling (LHS) method. For each pair of source and receiver, we measure the error in the traveltime approximation  $t^{\mathcal{K}}$  by the Root Mean Squared Relative Error (RMSRE),

$$\text{RMSRE} := \sqrt{\frac{1}{N_v} \sum_{k=1}^{N_v} \left( \frac{t(\mathbf{m}^{(k)}) - t^{\mathcal{K}}(\mathbf{m}^{(k)})}{t(\mathbf{m}^{(k)})} \right)^2}, \quad (32)$$

where  $t(\mathbf{m}^{(k)})$  denotes the traveltime computed from the eikonal solver with velocity model given by  $\mathbf{m}^{(k)}$ . Figure 2 reports the RMSRE of the 50 traveltimes for different PSP levels. The traveltimes are grouped by receivers (10 groups), each containing the five sources. The plot shows that RMSRE decreases with  $\ell$  (note the logarithmic scale on the error). At a given PSP level, the errors depend on the considered receiver. For instance, for  $\ell = 2$  and source 1, the PSP error is  $3.22 \times 10^{-3}$  to receiver 1 and  $1.54 \times 10^{-2}$  to receiver 6. These differences are not due to the normalization of the errors as all the traveltimes have a comparable magnitude. The magnitude and uncertainty level in the traveltimes can be appreciated from Fig. 3, which reports the mean and standard deviation of the traveltimes PSP estimates at level  $\ell = 6$ . The significant differences between the RMSRE associated with different couples of source and receiver come from the

PSP level ( $\ell$ )	2	4	6
SG points ( $N_q$ )	71	1471	18943
PC basis dimension ( $P$ )	26	321	2972
Maximum total degree ( $d^\circ$ )	3	15	63

Table 2: Characteristics of the PSP method for different levels. Isotropic SG in 5 dimensions.

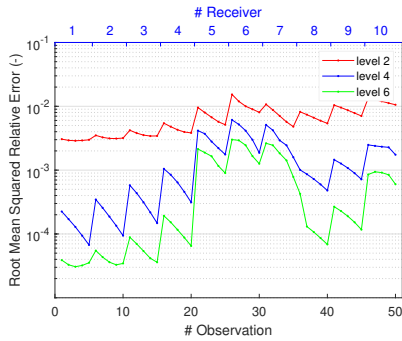


Figure 2: Microseismic case: RMSRE of the PSP approximation of the traveltimes at level  $\ell = 2, 4$  and 6.

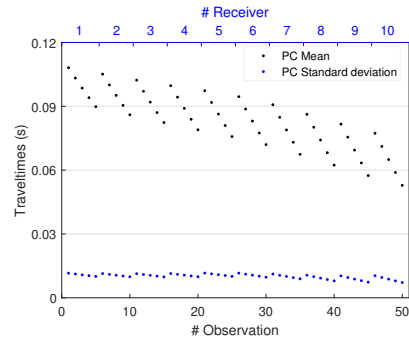


Figure 3: Microseismic case: mean and standard deviation of PC estimates of the traveltimes. PSP with level  $\ell = 6$ .

different complexities in the dependencies of the traveltimes on the parameters of the velocity model, as further analyzed in the next section.

#### 4.2.2. One-at-a-time dependencies

Figure 4 depicts the One-At-a-Time (OAT) variations with the components of  $\mathbf{m}$ , of the traveltimes from source 1 to receivers 1 and 6. The objective is to illustrate the complexity of the dependencies and the regularity of the traveltimes. The OAT variations are anchored at the mid-interval value of  $\mathbf{m}$  (a priori mean value). Figure 4a shows that the traveltime to receiver 1 depends smoothly on the five parameters of the velocity model. This smoothness ensures a high convergence rate of the PC approximation with the PSP level. In contrast, when the location of the interface  $z_{12}$  varies, the receiver 6 is either in the first or in the second layer. Also for high  $v_1$ , the faster raypath corresponds to the refracted wave travelling along  $v_1$  before reaching the receiver located in the second layer. Consequently, the traveltime from source 1 is independent of  $z_{12}$  and  $v_1$  over much of their prior range and presents localized singularities. The non-differentiable points are visible in Fig. 4b for  $v_1$  at 90% and  $z_{12}$  at 60% of their respective prior range. Non-smooth dependencies deteriorate the convergence rate of the PC expansions. However, their impact is not too severe in the present example, with RMSRE much lower than  $10^{-2}$  for PSP levels 4 and 6. This error level should be compared to the precision of the observed traveltimes, and the influence of the PC error on the Bayesian inference results is further analyzed below.

#### 4.2.3. A priori global sensitivity analysis

Before proceeding with the Bayesian inference, we present the a priori global sensitivity analysis of the traveltimes using the PC surrogates. Figure 5 reports the first-order indices (30) of the traveltimes. The first-order indices differ significantly depending on the considered receiver. For instance,  $v_1$  has no impact on the traveltimes to receivers 6-10 but explains roughly 20% of the traveltimes' variability to receivers 1 and 2. Similarly,  $v_2$  is the most influencing parameter for receivers 1-8, while it is  $v_3$  for others. Finally, the uncertainties in the interface locations have limited impact on the traveltimes uncertainty,  $z_{12}$  alone inducing no variance on the traveltimes to receivers 7-10. Also, the sum of the first-order sensitivity indices, which is close to 1 for all source and receiver couples, indicates that interaction effects are weak (less than 5% of the global variance over the prior range). The impacts of the uncertain parameters on the traveltimes are consistent with the sources-receivers configuration shown in Fig. 1a. We can already anticipate that

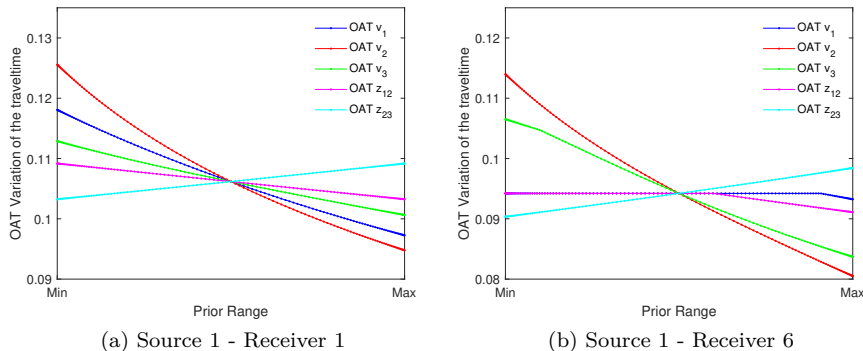


Figure 4: Microseismic case: One-At-a-Time (OAT) variations of traveltimes.

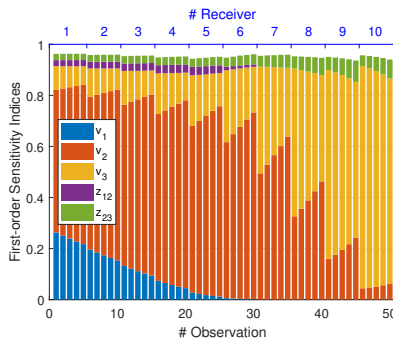


Figure 5: Microseismic case: first-order sensitivity indices estimated with PSP at level 6.

the interface locations will be more difficult to infer than the velocities, as they have a lower impact on the traveltimes.

### 4.3. Inference

We start the inference with synthetic observations corrupted with an additive Gaussian noise following the structure of the discrepancy model, with 2% noise level:  $\mathbf{e}_{\text{obs}} \sim \mathcal{N}(0, \mathbf{C}_{\text{obs}}(0.02))$  where  $\mathbf{C}_{\text{obs}}$  is defined in (4) with  $t_{\text{obs},i}$  substituted with  $t_i(\mathbf{m}_{\text{true}})$ .

#### 4.3.1. MCMC validation

In all subsequent analyses, the reported results use  $N_c = 10^6$  steps in the MCMC sampler, after the proposal covariance adaptation and burn-in phase. We recall that the chains use the PC surrogates of the traveltimes, such that dealing with long chains is not a problem. We have systematically performed complete assessments of the mixing and convergence of the MCMC chains for all inference results presented in the paper; we only detail the diagnostics in this section in the case of the PSP surrogate at level  $\ell = 6$  (results are similar for lower levels). Figure 6 details a typical chain of the MH algorithm after the covariance adaptation and burn-in. The left plots show the whole chain for the five parameters of the velocity model and noise level ( $\alpha$ ). Consistently with the results of the a priori sensitivity analysis, the chain explores most of the prior range for the interface locations ( $z_{12}$  and  $z_{23}$ ) and only a limited part of the velocity prior ranges. The right plots focus on a subset of 500 steps to appreciate the chain's mixing rate, which has an acceptance rate of about 26%. Figure 7 shows the autocorrelation functions of the chain for a quantitative assessment of its mixing properties. The figure reports the autocorrelation  $\rho(k)$ , where  $k$  is the lag expressed in the number of steps. The autocorrelations of the sampled parameters decay quickly and become negligible after  $k \approx 75$  steps. Following [37], the effective sample size incurring to the correlation between the steps can be

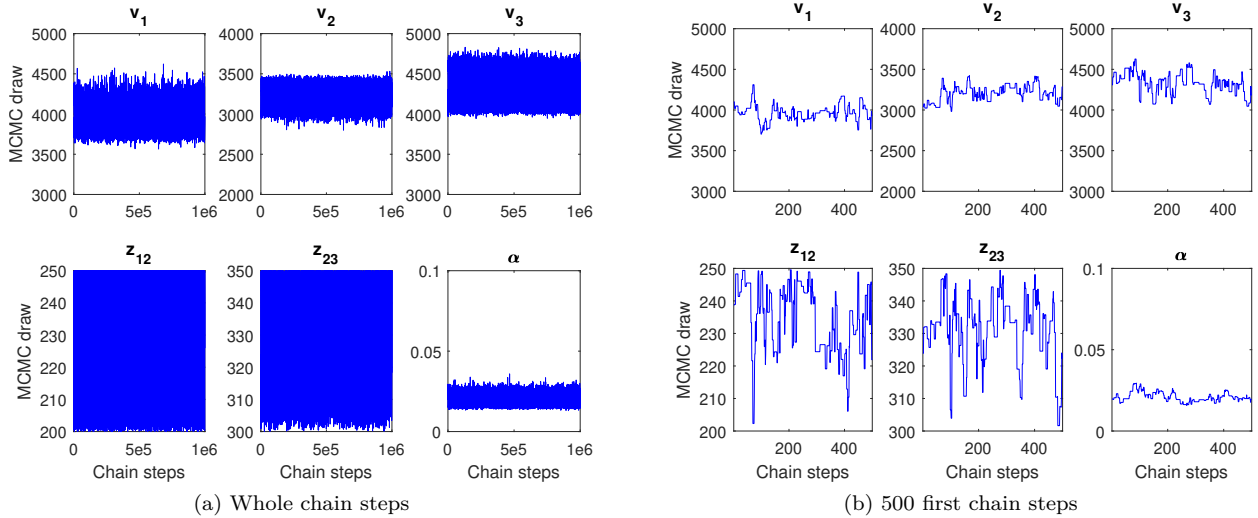


Figure 6: Microseismic case: MCMC chains for PSP surrogates at level 6. The ranges of the y-axis correspond to the prior ranges of the velocity model parameters.

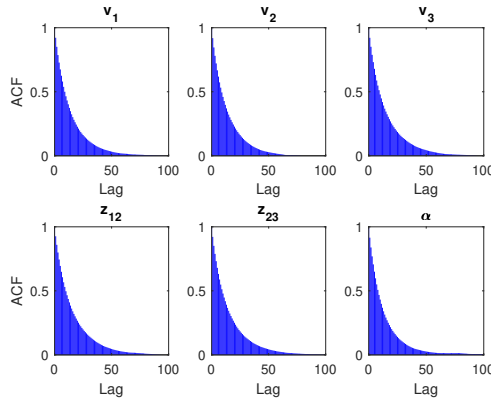


Figure 7: Microseismic case: autocorrelation functions for the PSP surrogates at level 6.

estimated as  $ESS := N_c / (1 + 2 \sum_{k=1}^{\infty} \rho(k))$ . Table 3 reports the effective fraction of samples  $ESS/N_c$  of the parameters for the present case. The table shows that the effective fraction is  $\gtrsim 3\%$ , a value that is quite satisfying.

Additionally, we have compared our results with those directly obtained from the eikonal solver. Figure 8 plots the misfit function obtained with the PC surrogates of the traveltimes versus the exact misfit function for a thousand samples of  $\mathbf{m}$  drawn from the MCMC chain. These so-called Q-Q plots confirm the convergence of the PC approximation with the PSP level and display a good agreement for the level 6 surrogates. This agreement, in turn, ensures that the posterior distribution based on the PC surrogates is close to the posterior obtained with the original eikonal solver. The eikonal solver’s marginal posteriors were also successfully contrasted with their PC-based counterparts (not shown).

#### 4.3.2. Marginal distributions

Figure 9 plots the marginals of the posterior distributions obtained for three PSP levels. The marginals in Fig. 9 are estimated with a standard Kernel Density Estimation (KDE) method [54]. The plots also show the maximum (over the MCMC chains) a posteriori (MAP) estimates of  $\mathbf{m}$  and the true model parameters  $\mathbf{m}_{\text{true}}$ . We observe a significant concentration of the marginals around the true values for the velocities,

Parameter	$v_1$	$v_2$	$v_3$	$z_{12}$	$z_{23}$	$\alpha$
ESS/ $N_c$ (%)	3.58	3.52	3.10	2.99	3.27	3.55

Table 3: Microseismic case: fraction of effective samples for the PSP surrogates at level 6.

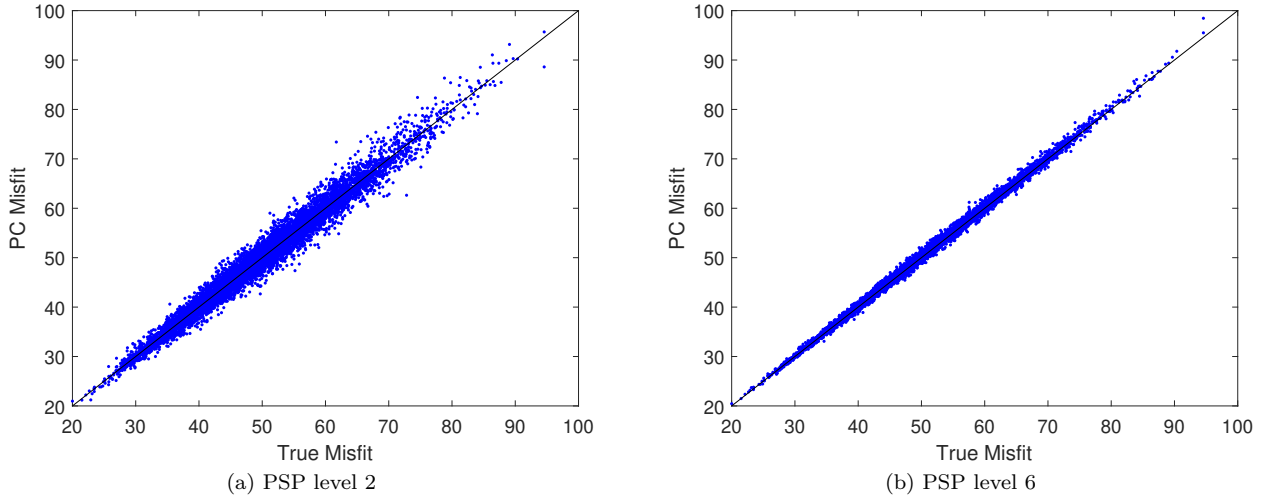


Figure 8: Microseismic case: true misfit function versus approximated PC misfit function plotted for 1,000 independent values for  $\mathbf{m}$  extracted from the MCMC sampling.

especially for  $v_1$  and  $v_2$ . In contrast, the shapes of the posterior marginals of the two interface locations cover the whole prior range, reflecting the weak influence of  $z_{12}$  and  $z_{23}$  on the traveltimes (see Fig. 5). We also see that the MAP over the chains may not always converge to the true model parameters with respect to the maximum PC total degree whereas the MAP of the marginals are closer to the true values. The parameter  $\alpha$  is correctly estimated, and the inference results are robust to the selected SG level, except for the interface  $z_{12}$ , which seems to be more sensitive to  $\ell$ .

#### 4.3.3. Two-dimensional marginals

Figure 10 shows, for  $\ell = 6$ , KDE approximations of two-dimensional posterior marginals to analyze the correlations between the inferred velocity model parameters. Concerning the wave velocities, we found very weak correlations with estimated coefficients of determination  $R^2(v_1, v_2) = 0$ ,  $R^2(v_1, v_3) = 5\%$ , and  $R^2(v_2, v_3) = 3\%$ . We also notice some correlations between the interface locations and the velocities in the layers. The first interface location is correlated with the velocities of layers 1 and 2 ( $R^2(z_{12}, v_1) = 25\%$ ,  $R^2(z_{12}, v_2) = 38\%$ , while  $R^2(z_{12}, v_3) = 2\%$ ). The second interface location, on the contrary, is correlated with the velocities of layers 2 and 3 ( $R^2(z_{23}, v_2) = 26\%$ , and  $R^2(z_{23}, v_3) = 83\%$ ), but not with the first layer velocity ( $R^2(z_{23}, v_1) = 3\%$ ). Finally and unsurprisingly, the noise parameter  $\alpha$  is not correlated with the velocity model parameters (the coefficients of determination are lower than 3% for the five model parameters).

#### 4.4. Decreasing of measurement errors

To better understand the effect of the surrogates error and its impact depending on the observation noise, we repeat the previous inference problem using synthetic observations generated with a noise level set to 0.01%. This situation corresponds to observations with unrealistically low in situ measurement error. As a result, the surrogates' error dominates the discrepancy model in the likelihood. Figure 11 shows the posterior marginals of  $v_1$ ,  $z_{12}$ ,  $z_{23}$ , and  $\alpha$  for PSP levels  $\ell = 2, 4$  and 6 (the two other velocities' marginals behave as for  $v_1$ ). Compared to the previous test, we first observe the substantial reductions in the supports

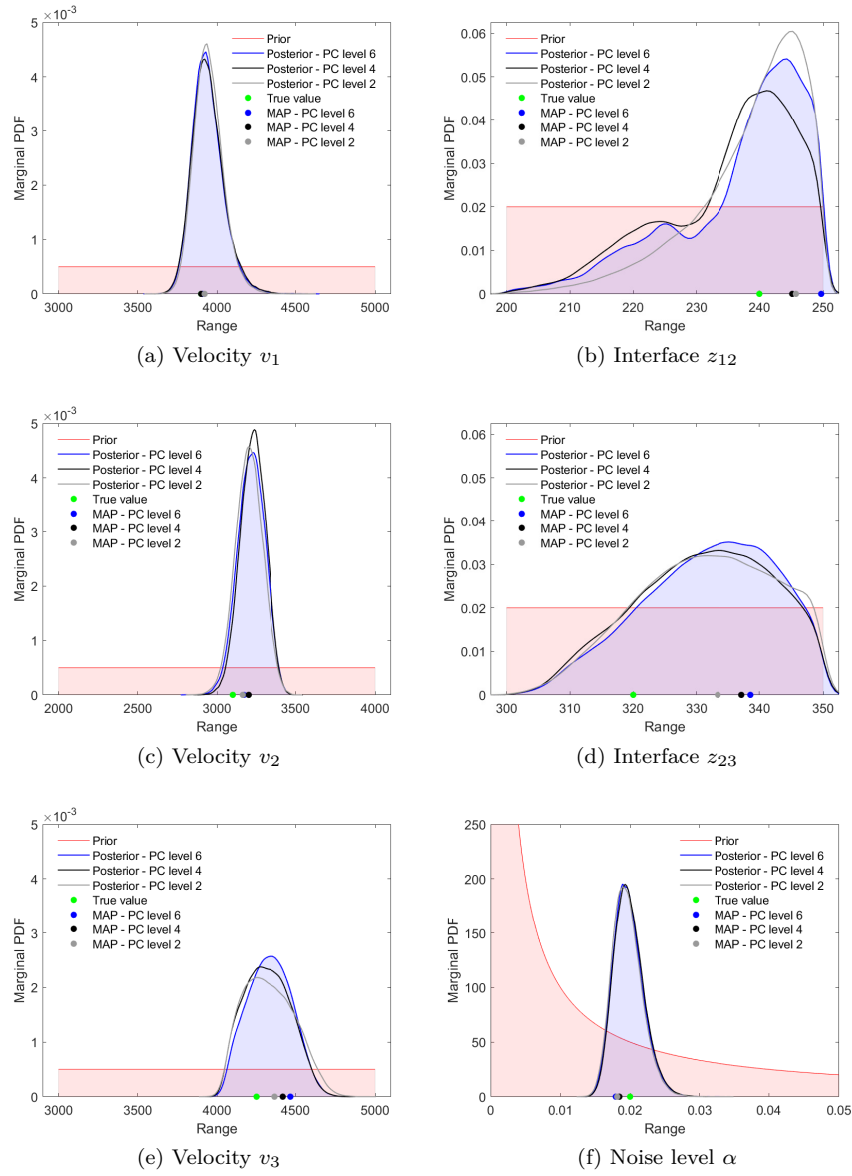


Figure 9: Microseismic case: priors and posterior marginals inferred with SG levels  $\ell = 2$  (gray curves),  $\ell = 4$  (black curves) and  $\ell = 6$  (blue curves). The noise level of the synthetic observations is 2%.



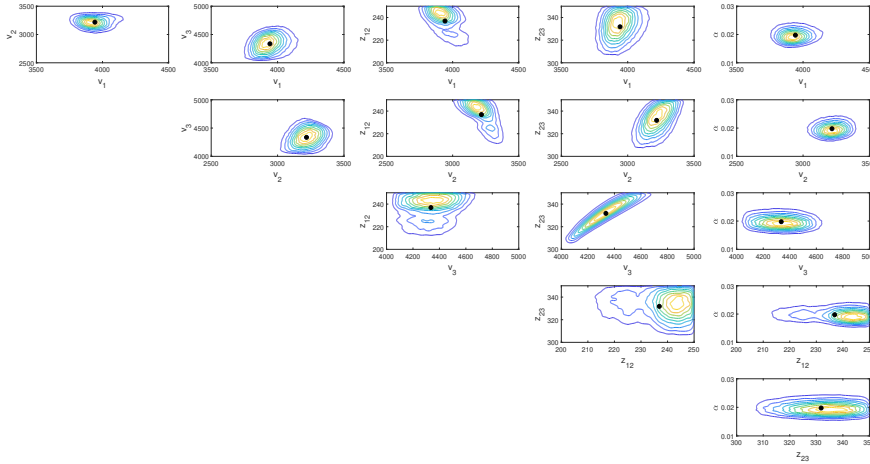


Figure 10: Microseismic case: two dimensional posterior marginals of the model and noise parameters. KDE estimation with PSP level 4. The noise level of the synthetic observations is 2%.

of the marginals (note the change of ranges compared to Fig. 9), which denote a much narrower posterior. Further, the marginals shrink as  $\ell$  increases and the estimate of  $\alpha$  reduces. This behavior is consistent with the reduction of the surrogates' error as  $\ell$  increases. As shown in Fig. 11d, the Bayesian inference assimilates the surrogates' error as observations' error, leading to an overestimation of  $\alpha$  and an artificially increased posterior uncertainty for the other parameters.

It is interesting to note a better estimation of the interface locations, which points to the interest of acquiring data as accurately as possible to infer the least influential parameters. However, Fig. 11b shows that the PSP level  $\ell = 6$  is not enough to correctly estimate  $z_{12}$  since the true value is not squarely within the support of the posterior marginal. This remark highlights an essential drawback of the inference in the presence of significant surrogates' error: it can be possible to find values  $\mathbf{m} \neq \mathbf{m}_{\text{true}}$  that fit the observations significantly better than  $\mathbf{m}_{\text{true}}$ . This situation can be detected by checking the convergence with the PSP level of the inference results. If the convergence is not satisfactory, the surrogates can be refined by increasing  $\ell$  and exploiting the sparse grid's nested character. An alternative option consists of reducing the surrogates' error over the support of the posterior distribution, as opposed to uniformly over the prior, following an adaptive strategy, as illustrated in the next section.

## 5. Seismic refraction test case

### 5.1. Description

The second test case uses the same velocity model structure, with three layers having unknown interface locations, but it differs from the previous setting by a different acquisition geometry and a computational domain extended to  $[0, 10] \times [0, 3]$  km<sup>2</sup>. The spatial mesh now involves 1,200,000 square cells with  $5 \times 5$  m<sup>2</sup> size. The sources and receivers are uniformly spaced at the surface, with 100 receivers deployed every 100 m and 5 sources located every 2.4 km; Figure 12a illustrates the acquisition geometry. This acquisition configuration leads to  $N = 500$  observations of sources to receivers traveltimes. Table 4 presents the uniform prior ranges for the velocity model parameters, illustrated in Fig. 12, and the true values used to generate the synthetic observations.

### 5.2. Surrogate analysis

#### 5.2.1. PC validation

The surrogate construction of the traveltimes uses the least-squares method (see section 3.2.2) on a quasi-Monte Carlo sample set of  $N_q = 1,000$  realizations. Recall that the least-squares regression is more flexible in the choice of the sampling points and less sensitive to local losses of regularity than the spectral projection.

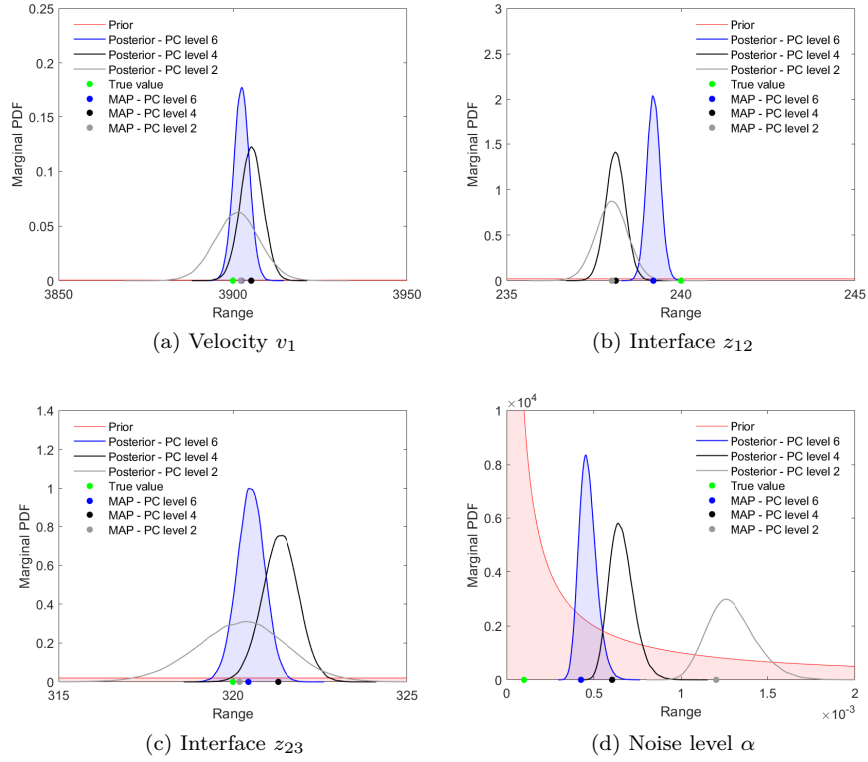


Figure 11: Microseismic case: posterior marginals of velocity  $v_1$ , interface locations  $z_{12}$  and  $z_{23}$ , and noise parameter  $\alpha$ , for PSP levels  $\ell = 2$  (gray curves),  $\ell = 4$  (black curves) and  $\ell = 6$  (blue curves). The noise level in the synthetic observations is 0.01%.

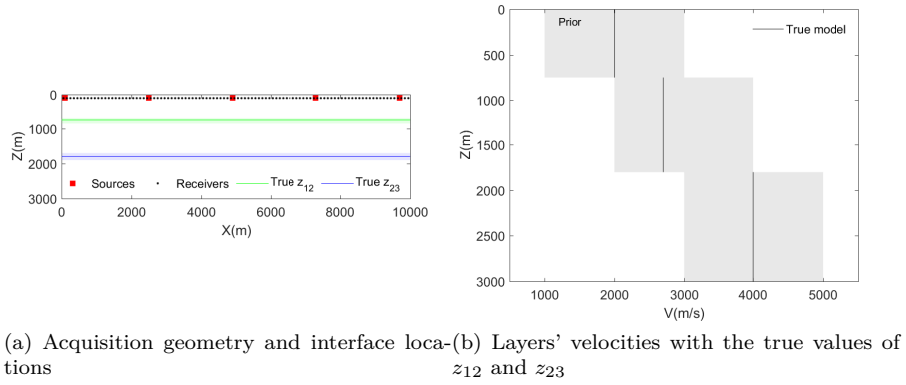


Figure 12: Schematic description of the refraction test case

Parameter	$v_1$ (m/s)	$v_2$ (m/s)	$v_3$ (m/s)	$z_{12}$ (m)	$z_{23}$ (m)
Prior ranges	$[1, 3] \times 10^3$	$[2, 4] \times 10^3$	$[3, 5] \times 10^3$	$[700, 850]$	$[1700, 1900]$
True values	2000	2700	4000	750	1800

Table 4: Refraction case: ranges of the uniform prior distributions of the velocity model parameters and true values used for the synthetic observations.

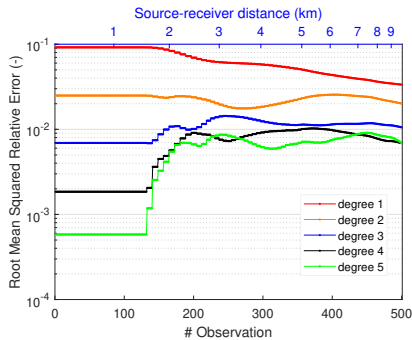


Figure 13: Refraction case: RMSRE of the PC surrogates of the traveltimes for maximum total degree  $d^\circ$  from 1 to 5.

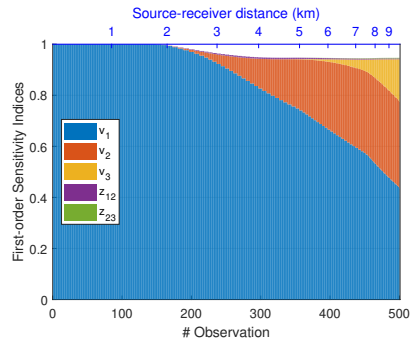


Figure 14: Refraction case: first-order sensitivity indices estimated with PC surrogates of maximum total degree  $d^\circ = 5$ .

The surrogate errors are estimated using an independent LHS of 500 realizations. The RMSRE (32) are computed for maximum total degree  $d^\circ$  from 1 to 5 leading to 6, 21, 56, 126, and 252 terms in the PC expansions, respectively. These numbers have to be compared to the  $N_q = 1,000$  realizations for the regression. Figure 13 shows the RMSRE of the 500 predicted traveltimes; the traveltimes are sorted by increasing source-receiver distance (see labels of the top axis). We observe that the error depends clearly on the source-receiver distance and two groups are identified. When the distance is less than 1.6 km, the traveltimes correspond to direct waves and depend only on the inverse of the first layer velocity (slowness) inducing an exponential convergence rate of the PC surrogates. For larger distances, first arrivals correspond to refracted waves with traveltimes having more complex dependencies on the uncertain parameters of the velocity model producing higher surrogate errors. In the subsequent results, we use a maximum total degree  $d^\circ = 5$  for which the error is around  $5 \cdot 10^{-4}$  for the first group and plateaus to  $10^{-2}$  for the second one.

### 5.2.2. *A priori global sensitivity analysis*

Figure 14 reports the first-order indices of the 500 PC traveltimes. As observed previously, the variance is only due to the uncertainty in the first layer's velocity for short source-receiver distance. We observe that the second (resp. third) layer's velocity influences the traveltimes for source-receiver distances higher than 2 km (resp. 5 km) corresponding to refracted waves travelling along the interfaces. The first-order sensitivity indices also indicate that the contribution to the variance of the layers' velocity decreases with their depth, suggesting that the inference of  $v_1$  will be more precise than for  $v_2$  and  $v_3$ . Also, we remark that the first-order sensitivity of the interface locations is negligible. Indeed, in the present case, their impact is limited to interaction effects with the velocities. Thus, one can expect some correlation between the inferred velocity and the interface locations.

### 5.3. Inference

We now proceed with the inference problem. We first generate synthetic observations of the 500 traveltimes by corrupting the traveltimes predicted by the model (solving the eikonal equation) using the true parameter values reported in Table 4. Like the previous section, an additive, centered, Gaussian noise is considered, with standard deviation amounting to 2% of the computed value ( $\alpha = 0.02$ ).

#### 5.3.1. *Prior-based first step*

Figure 15 shows the marginals of the posterior distribution of the velocity model parameters and noise parameter  $\alpha$ , based on the PC surrogates with maximum total degree  $d^\circ = 5$ . These marginals are estimated from  $N_c = 10^6$  steps of chain generated by MCMC, using a Gaussian proposal adapted to yield an acceptance rate of roughly 30%. The marginals of the first layer velocity  $v_1$  in Fig. 15a and noise parameter  $\alpha$  in Fig. 15f are highly peaked around the true values (2,000 m/s and 0.02 respectively). As expected from the sensitivity analysis, the second and third layer velocities marginals are less resolved than  $v_1$ , with increased

dispersion in their posterior, see Fig. 15c and 15e. Similarly, the observations are clearly less informative on the second interface location ( $z_{23}$ , Fig. 15d) compared to the first one ( $z_{12}$ , Fig. 15b).

Comparing the marginals for the two maximum PC total degrees in Fig. 15, we observe significant dependences on  $d^\circ$ , except for  $v_1$  and  $\alpha$ . Specifically, the marginals for  $d^\circ = 2$  and 5 of  $v_2$ ,  $v_1$  and  $z_{12}$  are barely overlapping, while the two marginals of  $z_{23}$  are not agreeing well. Also, the differences in the retrieved MAP values are significant, and the improvement in the MAP estimates for  $d^\circ = 5$  is unclear.

Besides the observations' limited information on the lowest layers parameters, discussed before, the PC surrogates explain the poor inference behavior. Indeed, the traveltime from a source to a receiver, sufficiently distant for the waves to travel in the second (resp. third) layer, depends on 3 (resp. 5) model parameters:  $v_1$ ,  $v_2$  and  $z_{12}$  (plus  $v_3$  and  $z_{23}$  resp.). Therefore, PC surrogates with  $d^\circ = 2$  that can only account for (linear) interaction between 2 variables are insufficient for source-receiver couples with high separation. Further, although PC surrogates with  $d^\circ = 5$  have a global error less than 1% (see Fig. 13), the errors are seen to affect the inference results significantly. A detailed analysis of the surrogates revealed that the isolated singularities in the traveltimes' dependencies, which are not differentiable at parameter values corresponding to bifurcations in the wave path, are the principal source of error. Although localized, these singularities pollute the whole PC surrogate because of its polynomial nature. Therefore, if increasing the maximum PC total degree  $d^\circ$  would improve the inference of the parameters, the improvement with  $d^\circ$  will be slow. Further, increasing  $d^\circ$  would increase the PC basis dimension and the simulations needed to construct the surrogates.

### 5.3.2. Posterior-based second step

Rather than attempting to construct highly accurate surrogates for the whole prior, it is much more efficient to improve the approximation accuracy in the regions with high posterior density. Focusing the approximation effort in regions with a posteriori high probability enables lower degree polynomial expansions and excludes most of the singularities. Of course, the posterior density being unknown, one must rely on an iterative approach, as discussed in [43]. The results presented below consider surrogates adapted just once, as subsequent iterations lead to similar inference results (not shown). Our approach thus consists of two steps. The first step corresponds to the PC surrogates construction, with a maximum total degree  $d^\circ$ , as performed above. In the second step, a subset of 1,000 independent values for  $\mathbf{m}$  is extracted from the MCMC sampling of the posterior using the first step PC surrogates. The traveltimes associated with the subset are computed using the eikonal model, providing data to fit with the second step surrogates. Because the sampled components of  $\mathbf{m}$  are not independent and have the complex marginals depicted in Fig. 15, an orthogonal expansion is not considered at this stage. Instead, the sample set is centered and standardized by diagonalizing the empirical covariance, and we formulate the surrogates in terms of these uncorrelated and normalized coordinates. We then rely on the OLS method to fit polynomial surrogates in the transformed coordinates. Here, without loss of generality, we used Hermite polynomial bases with a maximum total degree  $d_2^\circ$ , possibly different than  $d^\circ$ . Figure 16 reports the RMSRE of the 500 predicted traveltimes, for the first step ( $d^\circ = 2$  and 5) and the second step ( $d_2^\circ = 2$ , adapted) surrogates. The estimation of the RMSRE relies on an additional MCMC sampling of the posterior to draw 1,000 independent samples, distinct from the sample used for the construction of the adapted surrogates. The Q-Q plots of the corresponding misfit function in Fig. 17 clearly highlight the improvement brought by the adaptation, which yields satisfactory results.

Finally, we repeat the inference problem with the new polynomial surrogates substituting the original PC surrogates of the traveltimes. Figure 18 shows the corresponding posterior's marginals of  $v_3$ ,  $z_{12}$  and  $z_{23}$ , for  $d^\circ = 5$  and  $d_2^\circ = 2$ . The estimation of these marginals uses a chain of  $N_c = 10^6$  MCMC samples. The other marginals are identical to those obtained with the initial surrogates with  $d^\circ = 5$  (not shown). We observe that the posterior marginals remain quite broad, reflecting the limited information brought by the observations. The distance of the MAP values of the interfaces to the true MAP values is appreciable but consistent with the low sensitivity of the traveltimes on these quantities, making them more susceptible to the observation noise.

While not in the scope of the present paper, the adaptive approach outlined above can drastically reduce the computational cost of building the surrogate, while achieving a small error on the posterior of the

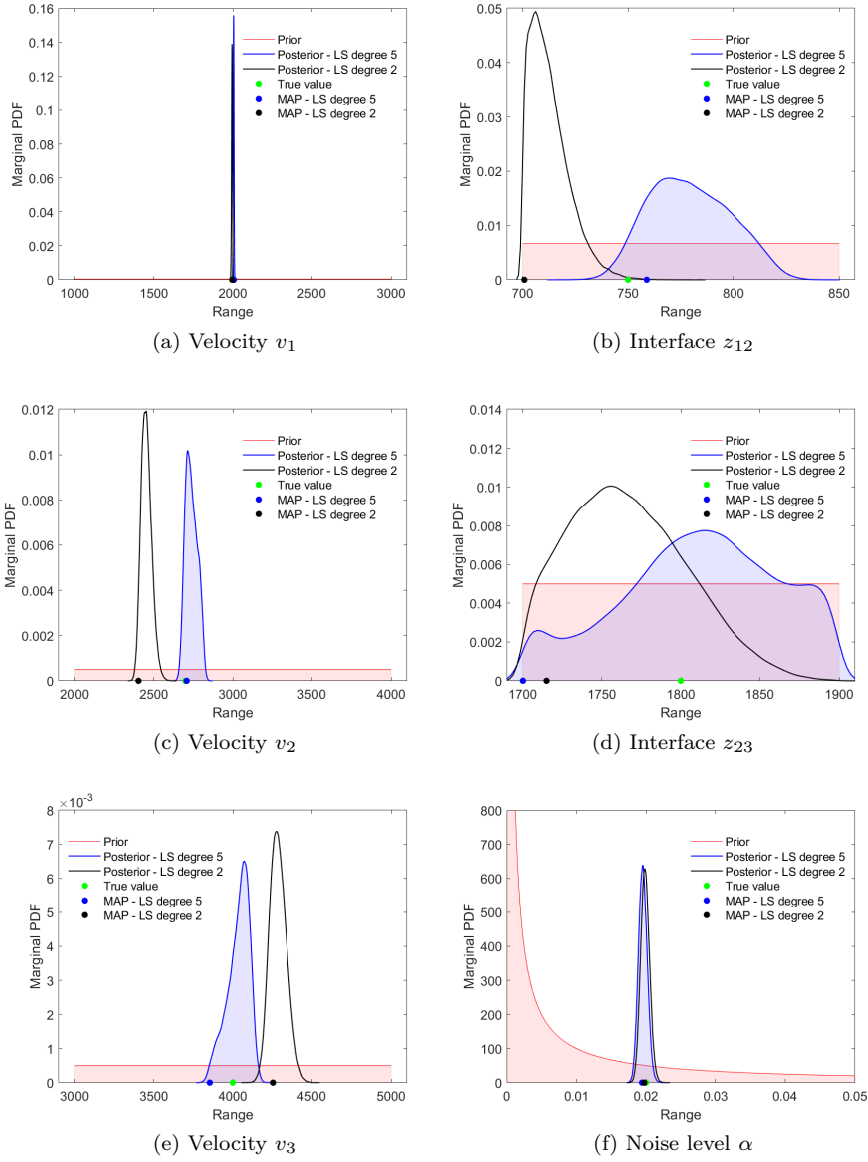


Figure 15: Refraction case: priors and posterior marginals of the velocity model parameters and noise level  $\alpha$ , based on the PC surrogates of maximum total degree  $d^o = 2$  and 5.

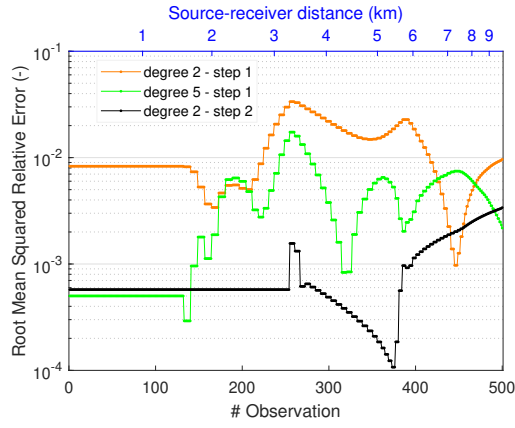


Figure 16: Refraction case: RMSRE of the PC surrogates of the traveltimes for the two steps of the adaptation strategy.

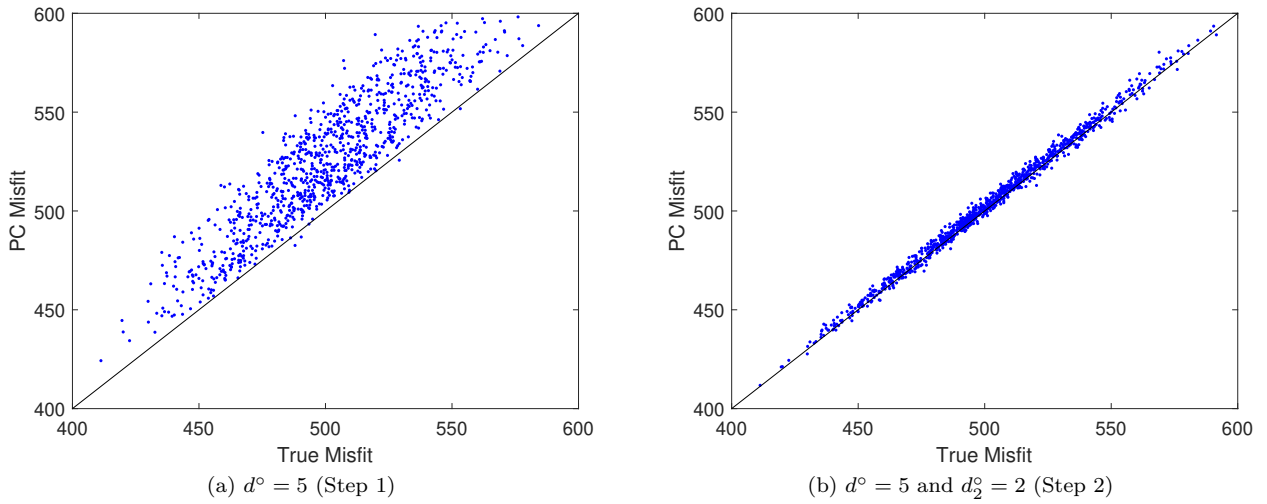


Figure 17: Refraction case: true misfit function versus approximated PC misfit function plotted for 1, 000 independent values for  $\mathbf{m}$  extracted from the MCMC sampling.

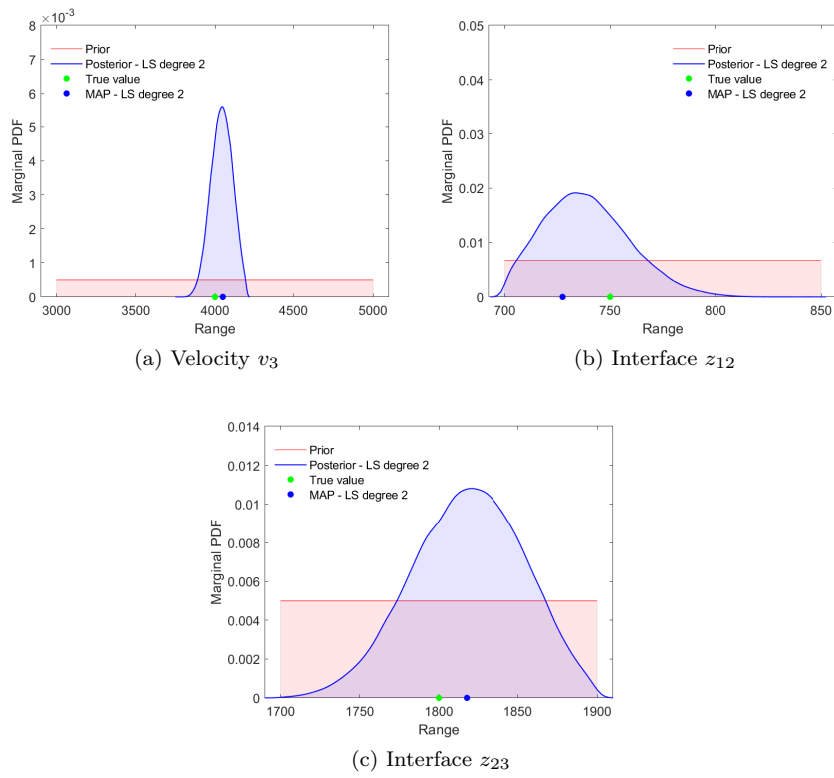


Figure 18: Refraction case: priors and posterior marginals of  $v_3$ ,  $z_{12}$  and  $z_{23}$  based on the polynomial surrogate constructed with  $d_2^o = 2$  on a sample set of 1,000 values drawn using the initial PC surrogates of the traveltimes with  $d^o = 5$ .

parameters compared to a direct Bayesian sampling using the eikonal solver. For instance, the second-order surrogates ( $d_2^o = 2$ ) can be estimated from a significantly smaller subset of samples than the one used in Fig. 18, such that the cost of the second step is just a fraction of the first. Similarly, one may prefer to start with lower-degree initial PC surrogates, constructed on a coarser sample set, and perform few adaptation iterations to correct any initial misfit, possibly retaining previous eikonal-based estimations of the traveltimes.

Future methodological developments could draw upon existing works on advanced strategies in spatial and parametric domains. For example, the two-stage MCMC [18] reduces the computational cost by predicting the acceptance of the proposed state using a coarse mesh (or a surrogate model) and relies on a fine-mesh (or the physical model) computation only if the state is found acceptable. Alternatively, one could consider using a data-driven posterior-oriented surrogate construction [15], where few samples of the detailed model correct the surrogate prediction along the MCMC chain.

## 6. Conclusion

This paper has investigated the use of Polynomial Chaos surrogates in seismic traveltime tomography. The numerical tests for canonical problems (simple layered media, microseismic and seismic refraction configurations) have illustrated that a moderate number of evaluations of the eikonal solver suffices to build traveltime surrogates with an error less than the typical observation noise. When constructed, the surrogates of the traveltimes enable the extensive sampling of the posterior distribution with Markov Chain Monte Carlo methods (even simple ones). Our experiments have shown that the posterior uncertainty in the velocity model parameters depends on the observations available and the structural dependencies between the parameters and the traveltimes. We have also demonstrated that the information content of a source-to-receiver traveltime measurement can be estimated, a priori, using global sensitivity analyses, which are straightforward to carry from the PC expansions. The last configuration tested (seismic refraction) also highlighted the potential of relying on advanced surrogate construction methods, suited to the posterior distribution, through an iterative adaptation. The adaptive approach can drastically improve the precision of the inference while reducing the overall construction cost [43].

Two methods, the sparse pseudo-spectral projection and the least-squares regression, have been used to build the surrogates in the numerical experiments but other approaches providing sparse Polynomial Chaos expansions, such as basis pursuit and matching pursuit, could be implemented in case of a higher number of model parameters (a few dozen). Although not illustrated in the present work, the resulting surrogates can serve testing assumptions, without significant computational overhead, and discriminate, for instance, different likelihood forms (i.e., different observation noise models). This capability may be especially relevant when in situ measurements can not be well reproduced using the considered velocity model, such that several discrepancy models must be tested and discriminated to account for model inadequacy [36]. Similarly, the design of new experiments to optimally reduce the posterior uncertainty and optimize the acquisition campaigns can exploit the surrogates.

Although the present article has demonstrated the potential of using surrogates in Bayesian traveltime tomography, many developments remain to apply this technique to the inference of realistic velocity models. Our researches currently focus on this aspect, including the case of continuous velocity models. Specifically, we are considering the introduction of parsimonious parametrizations of the velocity field, distinguishing the treatment of the components informed by the observations from the rest of the field. Another source of improvement currently explored is the extension of the inference problem to include other attributes, such as the direction of propagation, amplitude, and energy of the seismic wave at the receiver.

## References

- [1] M. Abramowitz and I. A. Stegun. Orthogonal polynomials. In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, number 55, chapter 22, pages 771–802. U.S. Government Printing Office, Washington, D.C., 9 edition, 1972.



- [2] T. Bayes. An essay towards solving a problem in the doctrine of chances. Phil. Trans. of the Royal Soc. of London, 53:370–418, 1763.
- [3] J. Belhadj, T. Romary, A. Gesret, M. Noble, and B. Figliuzzi. New parameterizations for bayesian seismic tomography. Inverse Problems, 34(6):065007, may 2018.
- [4] J. E. Besag. Comments on ‘representations of knowledge in complex systems’ by u. grenander and m.i. miller. J. Roy. Statist. Soc. Ser. B, 56:591–592, 1994.
- [5] S. D. Billings, M. S. Sambridge, and Kennett B. L. N. Errors in hypocenter location: Picking, model, and magnitude dependence. Bull. Seism. Soc. Am., 84(6):1978–1990, 1994.
- [6] T. Bodin and M. Sambridge. Seismic tomography with the reversible jump algorithm. Geophys. J. Int., 178(3):1411–1436, 2009.
- [7] A. Bottero, A. Gesret, T. Romary, M. Noble, and C. Maisons. Stochastic seismic tomography by interacting Markov chains. Geophys. J. Int., 207(1):374–392, 2016.
- [8] J. Brynjarsdóttir and A. O’Hagan. Learning about physical parameters: the importance of model discrepancy. Inverse Problems, 30(11):114007, oct 2014.
- [9] R.H. Cameron and W.T. Martin. The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals. Ann. Math., 48:385–392, 1947.
- [10] A. Cohen and G. Migliorati. Optimal weighted least-squares methods. SMAI J. Comp. Math., 3(3716755):181–203, 2017.
- [11] P. R. Conrad and Y. M. Marzouk. Adaptive Smolyak pseudospectral approximations. SIAM J. Sci. Comp., 35(6):A2643–A2670, 2013.
- [12] P. G. Constantine, M. S. Eldred, and E. T. Phipps. Sparse pseudospectral approximation method. Comput. Methods Appl. Mech. Engrg., 229:1–12, 2012.
- [13] T. Crestaux, O Le Maître, and J-M. Martinez. Polynomial Chaos Expansion for Sensitivity Analysis. Reliab. Eng. and Syst. Saf., 94(7):1161–1172, 2009.
- [14] H. Cruz-Jiménez, G. Li, P. Mai, I. Hoteit, and O. Knio. Bayesian inference of earthquake rupture models using polynomial chaos expansion. Geosci. Model Dev., 11(7):3071–3088, 2018.
- [15] T. Cui, Y. M. Marzouk, and K. E. Willcox. Data-driven model reduction for the bayesian solution of inverse problems. International Journal for Numerical Methods in Engineering, 102(5):966–990, 2015.
- [16] S. Duane, A.D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid monte carlo. Physics Letters B, 195(2):216 – 222, 1987.
- [17] S. Earp and A. Curtis. Probabilistic neural network-based 2d travel-time tomography. Neural Computing and Applications, 32:216 – 222, 2020.
- [18] Y. Efendiev, A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallick. An efficient two-stage markov chain monte carlo method for dynamic data integration. Water Resour. Res., 41(12), 2005.
- [19] Leo Eisner, Peter M Duncan, Werner M Heigl, and William R Keller. Uncertainties in passive seismic monitoring. The Leading Edge, 28(6):648–655, 2009.
- [20] A. H. Elsheikh, I. Hoteit, and M. F. Wheeler. Efficient bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates. Comput. Methods Appl. Mech. Engrg., 269:515 – 537, 2014.
- [21] A. Fichtner, A. Zunino, and L. Gebraad. Hamiltonian monte carlo solution of tomographic inverse problems. Geophys. J. Int., 216(2):1344–1363, 11 2018.
- [22] L. Formaggia, A. Guadagnini, I. Imperiali, V. Lever, G. Porta, M. Riva, A. Scotti, and L. Tamellini. Global sensitivity analysis through polynomial chaos expansion of a basin-scale geochemical compaction model. Comput. Geosci., 17(1):25–42, 2013.
- [23] T. Gerstner and M. Griebel. Numerical integration using sparse grids. Numerical Algorithms, pages 209–232, 1998.
- [24] R. G. Ghanem and S. D. Spanos. Stochastic Finite Elements: a Spectral Approach. Springer Verlag, 1991.
- [25] L. Giralaldi, O. Le Maître, K. Mandli, C. Dawson, I. Hoteit, and O. Knio. Bayesian inference of earthquake parameters from buoy data using a polynomial chaos-based surrogate. Comput. Geosci., 21(4):683–699, 2017.
- [26] H. Haario, E. Saksman, and J. Tamminen. An adaptive metropolis algorithm. Bernoulli, 7(2):223–242, 2001.
- [27] J. Hampton and A. Doostan. Coherence motivated sampling and convergence analysis of least squares polynomial chaos regression. Comput. Methods Appl. Mech. Engrg., 290:73–97, 2015.
- [28] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika, 57(1):97–109, 1970.
- [29] R. Hawkins, T. Bodin, M. Sambridge, G. Choblet, and L. Husson. Trans-Dimensional Surface Reconstruction With Different Classes of Parameterization. Geochemistry, Geophysics, Geosystems, 20(1):505–529, 2019.
- [30] R. Hawkins and M. Sambridge. Geophysical imaging using trans-dimensional trees. Geophys. J. Int., 203(2):972–1000, 09 2015.
- [31] W. Hoeffding. A class of statistics with asymptotically normal distribution. The annals of Mathematical Statistics, 19:293–325, 1948.
- [32] M. Iskandarani, S. Wang, A. Srinivasan, W. C. Thacker., J. Winokur, and O. Knio. An overview of uncertainty quantification techniques with application to oceanic and oil-spill simulations. J. Geophys. Res.: Oceans, 121(4):2789–2808, 2016.
- [33] H. Jeffreys. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 186(1007):453–461, 1946.
- [34] H. Jeffreys. Theory of Probability. Oxford, third edition, 1961.
- [35] K. Karhunen. Über lineare methoden in der wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys., 37:1–79, 1947.
- [36] Marc C. Kennedy and Anthony O’Hagan. Bayesian calibration of computer models. Journal of the Royal Statistical

- Society: Series B (Statistical Methodology), 63(3):425–464, 2001.
- [37] J. K. Kruschke. Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Academic Press, Inc., Orlando, FL, USA, 1st edition, 2010.
- [38] P. Le Bouteiller, M. Benjema, L. Métivier, and J. Virieux. An accurate discontinuous Galerkin method for solving point-source Eikonal equation in 2-D heterogeneous anisotropic media. Geophys. J. Int., 212(3):1498–1522, 10 2017.
- [39] O. P. Le Maître and O. M. Knio. Spectral Methods for Uncertainty Quantification. Scientific Computation. Springer, 2010.
- [40] C. Ley, G. Reinert, and Y. Swan. Distances between nested densities and a measure of the impact of the prior in bayesian statistics. Ann. Appl. Probab., 27(1):216–241, 2017.
- [41] G. Li, M. Iskandarani, M. Le Hénaff, J. Winokur, O. P. Le Maître, and O. M. Knio. Quantifying initial and wind forcing uncertainties in the Gulf of Mexico. Comput. Geosci., 20(5):1133–1153, 2016.
- [42] M. Loève. Probability Theory. The university series in higher mathematics. affiliated East-West-Press Pvt. Limited, 1968.
- [43] D. Lucor and O. P. Le Maître. Cardiovascular modeling with adapted parametric inference. ESAIM: ProcS, 62:91–107, 2018.
- [44] A. Malinverno. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. Geophys. J. Int., 151(3):675–688, 2002.
- [45] Y. M. Marzouk, Najm H. N., and L. A. Rahn. Stochastic spectral methods for efficient Bayesian solution of inverse problems. J. Comput. Phys., 224:560–586, 2007.
- [46] Y. M. Marzouk and D. Xiu. Stochastic spectral methods for efficient Bayesian solution of inverse problems. Commun. Comput. Phys., 6(4):826–847, 2009.
- [47] Shawn Maxwell. Microseismic: Growth born from success. The Leading Edge, 29(3):338–343, 2010.
- [48] M.D. McKay, W.J. Conover, and R.J. Beckman. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21:239–245, 1979.
- [49] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics, 21(6):1087–1092, 1953.
- [50] W. J. Morokoff and R. E. Caffisch. Quasi-Monte Carlo Integration. J. Comput. Phys., 122(2):218–230, 1995.
- [51] K. Mosegaard and A. Tarantola. Monte Carlo sampling of solutions to inverse problems. J. Geophys. Res.: Solid Earth, 100(B7):12431–12447, 1995.
- [52] M. Navarro, O. Le Maître, I. Hoteit, D. George, K. Mandli, and O. Knio. Surrogate-based parameter inference in debris flow model. Comput. Geosci., 22(6):1447–1463, 2018.
- [53] M. Noble, A. Gesret, and N. Belayouni. Accurate 3-d finite difference computation of traveltimes in strongly heterogeneous media. Geophys. J. Int., 199(3):1572–1585, 2014.
- [54] E. Parzen. On estimation of a probability density function and mode. Ann. Math. Statist., 33(3):1065–1076, 1962.
- [55] N. Rawlinson and M. Sambridge. Seismic traveltime tomography of the crust and lithosphere. Adv. Geophys., 46:81–198, 2003.
- [56] A.W. Roberts, R.W. Hobbs, M. Goldstein, M. Moorkamp, M. Jegen, and B. Heincke. Crustal constraint through complete model space screening for diverse geophysical datasets facilitated by emulation. Tectonophysics, 572-573:47 – 63, 2012.
- [57] G. O. Roberts and J. S. Rosenthal. Examples of Adaptive MCMC. Journal of Computational and Graphical Statistics, 18(2):349–367, 2009.
- [58] J. A. Sethian. A Fast Marching Level Set Method for Monotonically Advancing Fronts. In Proc. Nat. Acad. Sci., pages 1591–1595, 1995.
- [59] S.A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. Dokl. Akad. Nauk SSSR, 4(240-243):123, 1963.
- [60] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. Math. Model. Comput. Exp., 1:407–414, 1993.
- [61] P. Sochala and F. De Martin. Surrogate combining harmonic decomposition and polynomial chaos for seismic shear waves in uncertain media. Comput. Geosci., 2017.
- [62] Ihab Sraj, Olivier P. Le Maître, Omar M. Knio, and Ibrahim Hoteit. Coordinate transformation and polynomial chaos for the bayesian inference of a gaussian process with parametrized prior covariance function. Comput. Methods Appl. Mech. Engrg., 298:205 – 228, 2016.
- [63] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. Reliab. Eng. and Syst. Saf., 93(7):964 – 979, 2008.
- [64] A. Tarantola. Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.
- [65] J. Vidale. Finite-difference calculation of travel times. Bull. Seism. Soc. Am., 78(6):2062–2076, 12 1988.
- [66] D. J. Wilkinson. Parallel Bayesian computation. Statistics Textbooks and Monographs, 2006.
- [67] F. Zhang, R. Dai, and H. Liu. Seismic inversion based on L1-norm misfit function and total variation regularization. J. Appl. Geophys., 109:111 – 118, 2014.
- [68] H. Zhao. A fast sweeping method for Eikonal equations. Math. Comput., 74(250):603–627, 2005.