École des Ponts ParisTech
Internship report - M2 Research
04/04/2022 - 30/09/2022

Nadège Polette
Engineering student

Improvement of the Bayesian resolution for the inverse problem related to seismic-acoustic tomography

CEA, DAM, DIF, F-91297 Arpajon, France

Supervisors: Pierre Sochala [CEA DAM], with

- Alexandrine Gesret [Mines ParisTech, Geosciences center, Fontainebleau],
- Olivier Le Maître [CNRS, CMAP]

Julien Reygner [CERMICS]

Contents

Acknowledgements

I would like to thank my supervisor, Pierre Sochala, who gave me the opportunity to realize this internship. I really enjoy our discussions and I am looking forward to continue this work during my thesis. I would also like to thank Alexandrine and Olivier for our fruitful exchanges.

Many thanks to the entire team for the warm atmosphere, our scientific and casual talks are a daily pleasure.

I would like to thank Julien Reygner for supervising this internship. I also greatly appreciated the caring and efficiency of Isabelle and Caroline as well as Aurélie, Alice and Sandrine regarding the practical aspects of this experience.

Many thanks to my friends and family too, who are always here for me and who help me to not forget what really matters.

Summary

The fight against nuclear proliferation consists in the first place in overseeing the respect of the international treaties. In that context, numerical tools are developed in order to monitor the environment and analyze geophysical events. In particular, when an earthquake occurs, the displacement of the ground due to the waves travelling inside the Earth's crust is registered by seismographs. With these data, information about the event can be reconstructed, for instance its location and its nature.

In this report, the modeling is based on a seismic tomography approach. The link between the waves arrival times detected by receptors and the waves velocity field in the traveled region is studied. A method is developed in order to infer a continuous velocity field from the data, without using a strong *a priori* on the field shape. The computational cost of this method is reduced by the use of surrogate modeling, which consists in replacing expensive and accurate representations by cheaper approximations.

This approach is tested on a realistic one dimensional case. To not use a stronger *a priori* shape for the initial distribution of the velocity field allows to avoid overfitting and to obtain a more realistic and general prediction.

The extension to a two-dimensional case will need to add supplementary reduction methods and a better quantification of the error sources. The knowledge of the velocity field and the quantification of uncertainty related to this knowledge are useful to improve the determination of the seismic sources location and therefore of their accurate characterization.

Keywords: seismic tomography, Bayesian inference, surrogate models.

Résumé

La lutte contre la prolifération nucléaire passe en premier lieu par le contrôle du respect des traités internationaux. Dans ce contexte, des outils numériques sont développés afin de surveiller l'environnement et d'analyser les évènements géophysiques. En particulier, lorsqu'un séisme se produit, le déplacement du sol est enregistré par des séismographes. La localisation et la nature de l'évènement peuvent être déduites de ces données.

Dans ce rapport, la modélisation s'appuie sur une approche de tomographie sismique. Le lien entre le temps d'arrivée des ondes enregistrées par un récepteur et le champ de vitesse de ces ondes dans la zone traversée est étudié. Une méthode est développée afin d'inférer un champ de vitesse continu à partir des données, sans utiliser d'*a priori* trop contraignant sur la forme du champ. L'efficacité computationnelle est améliorée grâce à l'utilisation de modèles de substitution, qui sont des approximations des modèles de référence précis, mais coûteux.

Cette approche est appliquée à un cas réaliste unidimensionnel. Ne pas forcer un *a priori* trop fort sur la distribution initiale du champ de vitesse permet d'éviter le surapprentissage et d'avoir une prédiction plus réaliste et généralisable.

L'extension à des champs à deux dimensions nécessitera l'utilisation de méthodes de réduction supplémentaires, et une meilleure quantification des différentes sources d'erreurs. La connaissance du champ de vitesse et des incertitudes associées font partie des moyens d'améliorer la localisation d'une source sismique et donc de mieux définir ses caractéristiques.

Mots-clés: tomographie sismique, inférence bayésienne, modèle de substitution.

List of Figures

List of Tables

List of Algorithms

Presentation of the company

The CEA and the DAM

The *Commissariat à l'Énergie Atomique et aux énergies alternatives* (CEA) is a French public government-funded research organization created in 1945 by Charles de Gaulle. Its areas of expertise are

- defense and security,
- low carbon energy (nuclear and other renewable energy),
- research for the industry,
- fundamental research (science of matter and life science).

Collaboration projects with numerous academic and industrial partners are led. The CEA has 9 sites across France:

- headquarters in Paris,
- 4 civil sites in Paris-Saclay, Grenoble, Marcoule and Cadarache,
- 4 centers for military applications in *Direction des Applications Militaires* (DAM) Île-de-France (DIF), Le Ripault, Valduc, Cesta and Gramat.

My internship took place in the DAM site of the CEA, which has to design, build, maintain in operating conditions and dismantle nuclear warheads that equip airborne and oceanic French nuclear forces. It develops and produces nuclear cores and reactors for the French Navy and is responsible for the nuclear supply for peacekeeping. Finally, it contributes to national and international security against nuclear proliferation, terrorism and deals with disarmament matters.
More specifically, I worked at the DIF center for 6 months, located in Bruyères-Le-Châtel (Essonne). This site focuses on studying weapon physics and numerical simulations while combating nuclear proliferation and monitoring the French territory in case of any nuclear, earthquake or tsunami emergencies.

Workforce and technology

To achieve these goals, the DIF is composed of 2000 employees, mainly researchers, and it hosts around 600 workers from foreign companies every day. Since 2003, the center hosts the scientific computing complex of the CEA, which gathers all the CEA's supercomputers, like the Tera1000 supercomputer which is used for the DAM Simulation program. It also sells computational power to external companies (mainly French industries). These supercomputers, which are among the most powerful ones in Europe, help the CEA contribute to European

projects. Prominent researchers make up the workforce of the DAM site, publishing numerous papers all year long in cooperation with other (French or foreign) academics.

The monitoring mission of the CEA

The DAM is involved in the fight against the nuclear proliferation and the terrorism. Among other things, it places its expertise at the service of the *International Atomic Energy Agency* (IAEA) and of the *Comprehensive Nuclear-Test-Ban Treaty Organization* (CTBTO). This contribution to strategic programs is based on the abilities of the CEA concerning the tracking of seismic events, the detection of radionuclides, the measure of atmospheric phenomena, and the conception of sensors and associated networks. With the experience of its teams in this field, the CEA is in charge of detecting all seismic event, locating it rapidly, computing its magnitude, and emitting an alert aimed at public authorities. Its activities also includes quantification of seismic risk for given geographic areas.

In that context, my mission at the CEA is part of the fundamental research and studies that are led to improve modeling and detecting tools. In order to be efficient and precise when emitting alerts, the CEA needs competitive numerical models. In my six-months internship, I develop and enhance numerical methods aimed at better characterize the environment. This internship is supervised by Pierre Sochala, researcher at the DIF, and my in-school tutor is Julien Reygner. Regular meetings with Alexandrine Gesret, researcher at Mines ParisTech (Geosciences center, Fontainebleau) and Olivier Le Maître (CNRS, LIMSI, Université Paris-Saclay) complement the geophysical and mathematical sides of this internship.

1 Introduction

This internship takes place in the context of the data analysis from the *International Monitoring System* (IMS), which is introduced in Section 1.1. In particular, data of interest are seismic waves records. The associated geophysical context is presented in Section 1.2. In Section 1.3, the modeling framework is detailed.

1.1 International Monitoring System (IMS)

The fight against nuclear proliferation consists in the first place in international treaties' monitoring: the *Non-Proliferation Treaty* (NTP) and the *Comprehensive Test Ban Treaty* (CTBT). In particular, the CTBT's verification system includes the IMS, which is currently being deployed, and which will be composed of 321 stations using several detection methods:

- hydroacoustic technology: 11 stations, to detect acoustic energy generated by underwater events, and to help climate research;
- infrasound technology: 60 stations, to detect low-frequency sound waves in the atmosphere;
- seismic technology: 170 stations, to detect shock waves traveling through the Earth;
- radionuclide technology: 80 stations, to detect radioactive particles and radioactive noble gases;

The repartition of these stations are shown in Appendix A. The recordings of these stations will be transmitted in real-time to an international center based in Vienna (Austria), that will analyze them. The objective of the IMS is to guarantee the detection of every nuclear explosion if its energy is superior to 1,000 tons of TNT whatever its location onto the globe, in the atmosphere or the oceans [7].

1.2 Seismic tomography

As a result of the plate tectonics, constraints accumulate in the earth's crust. When these constraints become superior to the resistance of the crust, the weakest zones (the faults) break and a seism is triggered: the fault suddenly slips and a rupture propagates, generating waves inside the earth. The larger the broken surfaces, the stronger the seismic magnitude, *i.e.* the larger the quantity of energy emitted during the rupture, and the more amplified the earth movements at the surface tend to be. If the seism is located in the ocean, it can cause a tsunami.

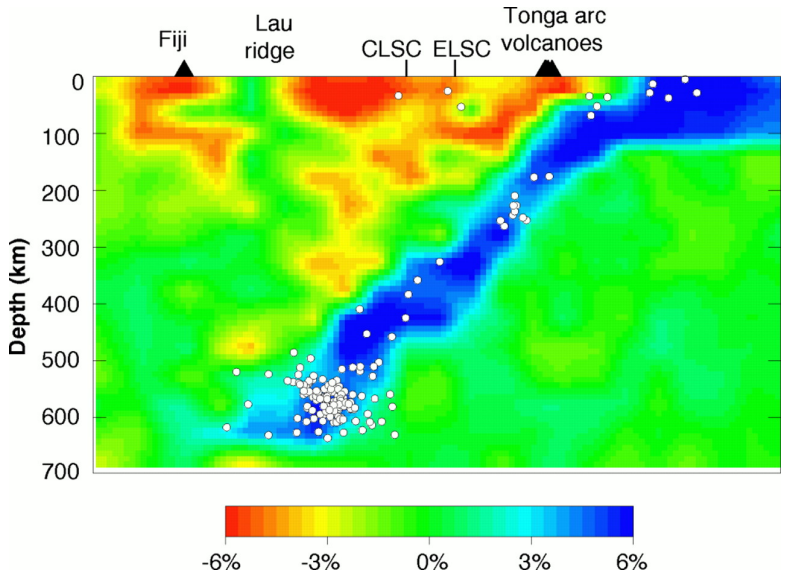


Figure 1: Anomaly map from seismic tomography: example of a subduction zone. Red and blue colors denote slow and fast velocities, respectively. The faster the velocity, the cooler the environment. The subduction phenomenon is highlighted: the cool crust part dives into the warmer sub-environment. Figure extracted from [24].

When a seismic event occurs, waves are registered by seismograms of the IMS. This information is used to cartography the earth and its properties: it is called seismic tomography. In particular, arrival times of the shear waves are considered. Considering the difference between the value of arrival time for typical values and wave behaviors versus real observations, anomaly maps as in Figure 1 are built. This allows to characterize the general Earth structure. More precisely, these arrival times are linked to the velocity field by the eikonal equation

$$|\nabla t(\boldsymbol{x})|^2 = \frac{1}{v^2(\boldsymbol{x})}, \tag{1}$$

where $\boldsymbol{x} \in \Omega \subset \mathbb{R}^2$ is the current spatial position, t is the arrival time, and $v : \Omega \to \mathbb{R}_+$ is the velocity field. Therefore, knowing the velocity field is sufficient to have a good approximation of the times of arrival. It also allows finding the location of the seism and its rupture mechanism.

This general analysis is valid for acoustic tomography too. The difficulty resides in the fact that the atmosphere is rapidly changing in time, in contrary to the Earth's crust, which enforces to model the environmental variations according to time at the considered time scale. The weather also need to be taken into account.

1.3 Modeling framework

In this study, we consider an heterogeneous domain $\Omega \subset \mathbb{R}^2$, where the location of the sources and the receptors are fixed. For each of the couples source-receptor, the times of arrival are known. The objective is to solve the inverse problem associated to equation (1), *i.e.* to find the underlying velocity field according to these observations as presented in Section 2.1.

In order to be computationally efficient and accurate, various tools are introduced in Section 2. A new formulation of the velocity field is built in Section 2.2: hyperparameters are taken into account without increasing the dimension of the problem. Moreover, surrogates are used in Section 2.5 to diminish the computational cost.

An application is presented in Section 3. We consider a one dimensional field: the velocity of the waves varies just according to the depth. For this model, results show that the formulation of the velocity field through a reference basis leads to an accurate inference of this field, without a strong *a priori* on the shape of the solution.

The interest of having a better knowledge of the velocity field is multiple. First, the wave behavior prediction may be more accurate in future direct modeling. Second, in the context of the CEA mission, precisely and efficiently characterizing the velocity field leads to the possibility of doing the same for the location and the nature of a seismic event, which improves the quality of the alert. Finally, having an error estimate on our predictions thanks to uncertainty quantification is important to define the trust level of the alerts.

2 Modeling the problem

The objective of this section is to detail the problem and the mathematical tools used to solve it. The inverse problem associated to seismic tomography is presented in Section 2.1. A parametrization of the velocity field using a covariance function and a method to infer the hyperparameters of this covariance function are introduced in Section 2.2. The Bayesian formulation is then detailed in Section 2.3 along with the *Markov Chain Monte–Carlo* (MCMC) sampling in Section 2.4. Finally, *Polynomial Chaos* (PC) surrogate models are described in Section 2.5 in order to reduce the computational cost.

2.1 Definition of the inverse problem

Considering that N^{obs} observations are available, a relation between the times of arrival $\mathcal{T} = \{t_i\}_{1 \leqslant i \leqslant N^{\text{obs}}}$ and the velocity field v is given in equation (1). Knowing the velocity field v, there

exists a way to explicitly compute the time of arrival from a source s to a receptor r. This is the direct problem, solved by an eikonal solver M:

$$M(v) = t : \Omega \times \Omega \to \mathbb{R}_+.$$

The quantity $t(s, r)$ characterizes the time of arrival of seismic waves for a source located in $s \in \Omega$ and a receptor located in $r \in \Omega$, as illustrated in Figure 2. In our case, a set of times of arrival $\mathcal{T}$ is known, but not the velocity field. This leads to consider the inverse problem: we seek the field v^* minimizing the error between the observations t^{obs} and the model prediction,

$$v^* = \underset{v \in \mathcal{V}}{\text{argmin}} \sum_{i=1}^{N} |t_i^{\text{obs}} - t_i(v)|, \tag{2}$$

where the search space $\mathcal{V}$ is discussed in Section 2.2. We are searching for the field v that is the most likely according to the observations, and not the one that is the nearest to the true unknown field. The objective is to find a *predictive* field, that can explain and generate realistic observations, contrary to an *interpretative* field, that accounts for the real characteristics of the crust.

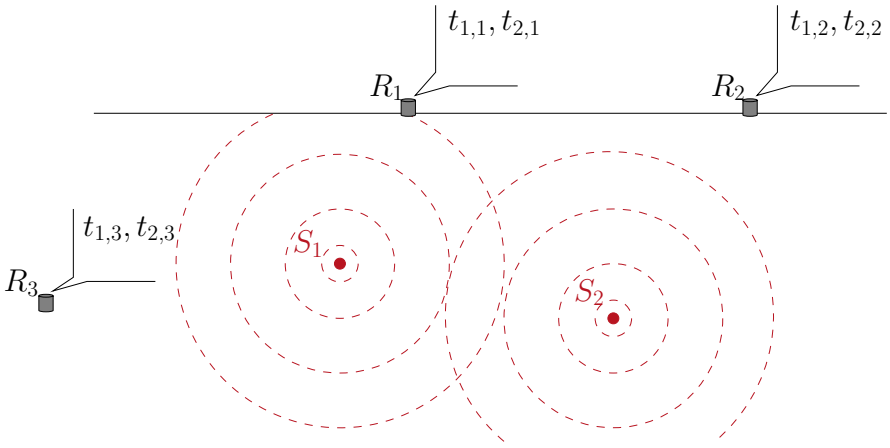


Figure 2: Principle of seismic tomography: seismic waves are emitted by sources $\mathcal{S} = (S_i)_i$ and registered by receptors $\mathcal{R} = (R_j)_j$. The data $\mathcal{T} = (t_{i,j}^{\text{obs}})_{i,j}$ are obtained.

2.2 Representation of the velocity field

This section aims at defining the search space $\mathcal{V}$ of the seismic wave velocity field. The representation of this field by its value at each point of a grid as done in full waveform inversion is discarded from the outset because of its intractable computational cost in a Bayesian framework. Given this, a first option is to use a layered model meaning that the seismic wave velocity

is a piecewise constant (or linear) field [20]. An explicit parametrization must then be introduced including the number of layers, their widths and dips as well as a type of velocity profile in each layer (e.g. constant or linear). Although attractive since it relies on a moderate number of parameters, this approach is restricted to simple profiles and suffers from a strong *a priori* knowledge about the field parametrization and a high dependence of this knowledge onto the optimization procedure.
A second option is to use a continuous model by means of a spatial covariance function [21]. Although more general, this approach requires to specify a type of covariance function (e.g. squared-exponential or Matérn) and the set of associated hyperparameters such as the amplitude and correlation length. In this section, we introduce the *Karhunen–Loève* (K–L) decomposition and detail a method, called *change of measure*, implemented to infer the hyperparameters of a given covariance function in order to decrease the a priori knowledge.

2.2.1 Framework

We are interested in the inference of a scalar velocity field $v(\boldsymbol{x})$ on a spatial domain $\Omega \subset \mathbb{R}^2 \to \mathbb{R}^+$. To ensure the positivity of $v(\boldsymbol{x})$, the logarithmic transformation $u(\boldsymbol{x}) := \log(v(\boldsymbol{x}))$ is performed. In addition, the transformed velocity field $u(\boldsymbol{x})$ is assumed to be a particular realization of a Gaussian random field $u(\boldsymbol{x}, \omega)$, ω denoting a random event belonging to a sample space $\mathcal{W}$ of an abstract probability space $(\mathcal{W}, \mathfrak{S}, P)$. Separating the mean profile μ, we have

$$u(\boldsymbol{x}, \omega) = \mu + U_q(\boldsymbol{x}, \omega), \quad U_q \sim \mathcal{N}(0, k(\cdot, \cdot, q)), \tag{3}$$

where $\mu \in [\mu_-, \mu_+] \in \mathbb{R}$ and the autocovariance function (or kernel) $k(\cdot, \cdot, q)$ depends on a set of hyperparameters $q \in \mathbb{H}$. The kernel is defined as

$$\forall q \in \mathbb{H},\ \forall \boldsymbol{x}, \boldsymbol{x}' \in \Omega, \quad k(\boldsymbol{x}, \boldsymbol{x}', q) := \mathbb{C}\mathrm{ov}(U_q(\boldsymbol{x}), U_q(\boldsymbol{x}')) := \mathbb{E}_\omega(U_q(\boldsymbol{x}, \cdot)U_q(\boldsymbol{x}', \cdot)), \tag{4}$$

where $\mathbb{C}\mathrm{ov}(\cdot, \cdot)$ and $\mathbb{E}_\omega(\cdot)$ respectively are the covariance and expectation operators. The expectation operator is defined as

$$\mathbb{E}_\omega(u(\boldsymbol{x}, \cdot)) := \int_{\mathcal{W}} u(\boldsymbol{x}, \omega) dP.$$

Finding the true logarithmic field is then equivalent to finding the values μ, q, ω such that for any $\boldsymbol{x} \in \Omega$, $u(\boldsymbol{x}) = \mu + U_q(\boldsymbol{x}, \omega)$.

2.2.2 Karhunen–Loève decomposition

For a given $q \in \mathbb{H}$, the K–L decomposition of $U_q(\boldsymbol{x}, \omega)$ reads

$$\forall \boldsymbol{x} \in \Omega, \qquad U_q(\boldsymbol{x}, \omega) := \sum_{i=1}^{+\infty} \lambda_i^{1/2}(q) u_i(\boldsymbol{x}, q) \eta_i(\omega),$$

where the couples $(u_i(q), \lambda_i(q))_{i \in \mathbb{R}_+^\star \; i \in \mathbb{R}_+^\star}$ are the eigenelements of the kernel and are obtained by solving the Fredholm equation of the second kind,

$$\forall \boldsymbol{x} \in \Omega, \forall i \in \mathbb{R}_+^*, \qquad \int_\Omega k(\boldsymbol{x}, \boldsymbol{x}', q) u_i(\boldsymbol{x}', q) d\boldsymbol{x}' = \lambda_i(q) u_i(\boldsymbol{x}, q). \tag{5}$$

As $U_q(\boldsymbol{x}, \cdot)$ is a Gaussian random field, the K–L coordinates are *independent and identically distributed* (i.i.d.) and follow a standardized Gaussian law. In addition, the K–L decomposition is truncated into a finite number of terms r, leading to

$$U_q(\boldsymbol{x}, \omega) \simeq U_q^r(\boldsymbol{x}, \omega) := \sum_{i=1}^{r} \lambda_i^{1/2}(q) u_i(\boldsymbol{x}, q) \eta_i(\omega), \quad \boldsymbol{\eta} := (\eta_i)_{1 \leqslant i \leqslant r} \sim \mathcal{N}(0, \mathrm{I}_r), \tag{6}$$

where I_r is the identity matrix of size r.

Remark 1. The K–L decomposition is bi-orthonormal, in the sense that

(i) the eigenfunctions $\{u_i(\boldsymbol{x})\}_{i \in \mathbb{R}_+^\star}$ are orthonormal,

$$\forall i, j \in \mathbb{R}_+^*, \qquad \langle u_i, u_j \rangle_\Omega := \int_\Omega u_i(\boldsymbol{x}) u_j(\boldsymbol{x}) d\boldsymbol{x} = \delta_{i,j}, \tag{7}$$

(ii) the coefficients (or coordinates) $\{\eta_i\}_{i \in \mathbb{R}_+^\star}$ are centered, non-correlated and of unitary variance:

$$\forall i, j \in \mathbb{R}_+^*, \qquad \mathbb{E}_\omega(\eta_i) = 0 \quad \text{and} \quad \mathbb{E}_\omega(\eta_i \eta_j) = \delta_{i,j}.$$

□

Remark 2. Any kernel admits the following (truncated) spectral decomposition,

$$k^r(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^{r} \lambda_i^{1/2} u_i(\boldsymbol{x}) u_i(\boldsymbol{x}'). \tag{8}$$

□

Remark 3. In this study, the squared exponential covariance kernel is used. This kernel is also known as the Gaussian kernel but this name is not used here to avoid confusion with the Gaussian distribution of $U_q(\boldsymbol{x}, \omega)$. The squared exponential covariance kernel is

$$k\left(\boldsymbol{x}, \boldsymbol{x}', \{A, \ell\}\right) := A \exp\left(\frac{-\left\|\boldsymbol{x} - \boldsymbol{x}'\right\|^2}{2\ell^2}\right),$$

where $\|\cdot\|$ is the Euclidean norm. This kernel has two positive hyperparameters, the amplitude (or variance) $A \in \mathbb{R}_+^\star$ and the correlation length $\ell \in \mathbb{R}_+^\star$,

$$q = \{A, \ell\} \in \mathbb{H} = [A_-, A_+] \times [\ell_-, \ell_+] \in \left(\mathbb{R}_+^\star\right)^2,$$

where A_- and A_+ (resp. ℓ_- and ℓ_+) are the minimum and maximum values of A (resp. ℓ). □

In order to find the best approximation of v, we want to find the optimal values of $\boldsymbol{\eta}$, μ and q.

2.2.3 Change of measure

The dependency of the eigenelements $\{u_i(\boldsymbol{x}, q), \lambda_i(q)\}_{1 \leqslant i \leqslant r}$ on the hyperparameters in the decomposition (6) is problematic in a Bayesian inference context for two reasons. First, it induces to solve the eigenvalue problem (5) for each new evaluation of the likelihood. Second, it requires to introduce a change of coordinates to a fixed reference basis [21]. To avoid these pitfalls, we propose to use the same decomposition basis regardless of the choice of hyperparameters; carrying the dependency of hyperparameters over the K–L coordinates. This decomposition, say $\overline{U}_q^r(\boldsymbol{x}, \omega)$, reads

$$U_q^r(\boldsymbol{x}, \omega) \simeq \overline{U}_q^r(\boldsymbol{x}, \omega) := \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x}) \xi_i(\omega), \qquad \boldsymbol{\xi} := (\xi_i)_{1 \leqslant i \leqslant r} \sim \mathcal{N}(0, \Sigma(q)), \tag{9}$$

where $\left(\overline{u}_i, \overline{\lambda}_i\right)_{1 \leqslant i \leqslant r}$ are the elements of the reference basis and $\Sigma(q)$ is the covariance matrix of the random coordinates $\boldsymbol{\xi}$. The decomposition basis and the covariance matrix are now introduced in definitions 1, 2 by using the mean $\mathbb{E}_q(\cdot)$ over the hyperparameters domain $\mathbb{H}$ and the inner product $\langle \cdot, \cdot \rangle_\Omega$ defined in (7).

Definition 1 (Reference basis)
The reference covariance kernel $\overline{k}$ is the covariance kernel $k(\cdot, \cdot, q)$ averaged with respect to q, namely

$$\forall \boldsymbol{x}, \boldsymbol{x}' \in \Omega, \qquad \overline{k}(\boldsymbol{x}, \boldsymbol{x}') := \mathbb{E}_q(k(\boldsymbol{x}, \boldsymbol{x}', \cdot)) := \int_{\mathbb{H}} k(\boldsymbol{x}, \boldsymbol{x}', q) \pi_q(q) dq, \tag{10}$$

where $\pi_\cdot$ denotes the *a priori* probability distribution of the quantity $\cdot$. The reference eigenelements $\{\overline{u}_i(\boldsymbol{x}), \overline{\lambda}_i\}_{i \in \mathbb{R}_+^\star}$, independent from q, are obtained by solving the eigenvalue problem for the reference kernel, that is

$$\forall \boldsymbol{x} \in \Omega, \ \forall i \in \mathbb{R}_+^*, \qquad \int_\Omega \overline{k}(\boldsymbol{x}, \boldsymbol{x}') \overline{u}_i(\boldsymbol{x}') d\boldsymbol{x} = \overline{\lambda}_i \overline{u}_i(\boldsymbol{x}). \tag{11}$$

Definition 2 (Covariance matrix)
The covariance matrix $\Sigma(q)$ of the K–L coordinates $\boldsymbol{\xi}$ is defined as

$$\forall 1 \leqslant i, j \leqslant r, \ \forall q \in \mathbb{H}, \qquad \Sigma_{i,j}(q) := (\overline{\lambda}_i \overline{\lambda}_j)^{-1/2} \left\langle \left\langle k(\boldsymbol{x}, \boldsymbol{x}', q), \overline{u}_j \right\rangle_\Omega, \overline{u}_i \right\rangle_\Omega. \tag{12}$$

Using this representation ensures that the error is minimized as shown in the following proposition.

Proposition 1. *Given the reference basis computed from Definition 1, then the covariance matrix introduced in Definition 2 minimizes the error of representation in the L^2-sense [21],*

$$\forall \boldsymbol{x}, \boldsymbol{x}' \in \Omega, \qquad \mathbb{E}_q\left[\mathbb{E}_\omega(\overline{U}_q^r(\boldsymbol{x},\cdot)\overline{U}_q^r(\boldsymbol{x}',\cdot)) - \mathbb{E}_\omega(U_q^r(\boldsymbol{x},\cdot)U_q^r(\boldsymbol{x}',\cdot))\right] = 0.$$

Proof of Proposition 1. In order to best approximate the Gaussian process U_q, the covariance kernel of $\overline{U}_q^r$ must be close to $k(\cdot,\cdot,q)$. Let $\boldsymbol{x}, \boldsymbol{x}'$ be in Ω. The autocovariance of $\overline{U}_q^r$ writes

$$\begin{aligned}\mathbb{E}_\omega(\overline{U}_q^r(\boldsymbol{x},\cdot)\overline{U}_q^r(\boldsymbol{x}',\cdot)) &= \sum_{i,j=1}^{r} \overline{\lambda_i}^{1/2}\overline{u_i}(\boldsymbol{x})\mathbb{E}_\omega(\xi_i(\cdot)\xi_j(\cdot))\overline{\lambda_j}^{1/2}\overline{u_j}(\boldsymbol{x}'),\\ &\overset{(9)}{=} \sum_{i,j=1}^{r} \overline{\lambda_i}^{1/2}\overline{u_i}(\boldsymbol{x})\Sigma_{i,j}(q)\overline{\lambda_j}^{1/2}\overline{u_j}(\boldsymbol{x}'). \end{aligned} \tag{13}$$

Applying the mean operator $\mathbb{E}_q(\cdot)$ to a covariance matrix term $\Sigma_{i,j}(q)$, we have

$$\begin{aligned}\mathbb{E}_q(\Sigma_{i,j}(\cdot)) &\overset{(12)}{=} (\overline{\lambda}_i\overline{\lambda}_j)^{-1/2}\left\langle\langle\mathbb{E}_q(k(\boldsymbol{x},\boldsymbol{x}',\cdot)),\overline{u}_j\rangle_\Omega,\overline{u}_i\right\rangle_\Omega, \qquad (14)\\ &\overset{(10)}{=} (\overline{\lambda}_i\overline{\lambda}_j)^{-1/2}\left\langle\left\langle\overline{k}(\boldsymbol{x},\boldsymbol{x}'),\overline{u}_j\right\rangle_\Omega,\overline{u}_i\right\rangle_\Omega,\\ &\overset{(11)}{=} \sqrt{\frac{\overline{\lambda_j}}{\overline{\lambda_i}}}\left\langle\overline{u}_j,\overline{u}_i\right\rangle_\Omega,\\ &\overset{(7)}{=} \delta_{i,j}. \qquad (15)\end{aligned}$$

Applying the mean operator $\mathbb{E}_q(\cdot)$ to equation (13), we get

$$\mathbb{E}_q\left[\mathbb{E}_\omega(\overline{U}_q^r(\boldsymbol{x},\cdot)\overline{U}_q^r(\boldsymbol{x}',\cdot))\right] = \sum_{i,j=1}^{r}\overline{\lambda}_i^{1/2}\overline{u_i}(\boldsymbol{x})\mathbb{E}_q(\Sigma_{i,j}(\cdot))\overline{\lambda}_j^{1/2}\overline{u_j}(\boldsymbol{x}') \overset{(15)}{=} \sum_{i=1}^{r}\overline{\lambda}_i\overline{u_i}(\boldsymbol{x})\overline{u_i}(\boldsymbol{x}').$$

In addition, we have

$$\mathbb{E}_q\left[\mathbb{E}_\omega(U_q^r(\boldsymbol{x},\cdot)U_q^r(\boldsymbol{x}',\cdot))\right] \overset{(4)}{=} \mathbb{E}_q\left[k^r(\boldsymbol{x},\boldsymbol{x}')\right] \overset{(10)}{=} \overline{k}^r(\boldsymbol{x},\boldsymbol{x}') \overset{(8)}{=} \sum_{i=1}^{r}\overline{\lambda}_i\overline{u_i}(\boldsymbol{x})\overline{u_i}(\boldsymbol{x}').$$

The linearity of the operator $\mathbb{E}_q(\cdot)$ concludes the proof. □

2.3 Bayesian formulation

2.3.1 Bayes' rule

The parameters set of the model defined by equations (3) and (9) includes the mean μ, the r coordinates ξ_i and the hyperparameters q. A noise level parameter $\alpha \in \mathbb{R}_+^\star$ is also introduced

to take into account the error measurements. Therefore, the total number of parameters is $d = r+|q|+2$ and the objective is to infer the distribution of $Y = \{\mu, \boldsymbol{\xi}, q, \alpha\} \in \mathbb{R}^d$. The Bayes' formula, used to compute the posterior of the parameters $\mu, \boldsymbol{\xi}$ and the hyperparameters q as well as the noise level parameter α, reads

$$p_{\text{post}}(Y|t^{\text{obs}}) \propto \mathcal{L}(t^{\text{obs}}|Y)\pi_Y(Y), \tag{16}$$

where $\mathcal{L}(t^{\text{obs}}|Y)$ denotes the likelihood and $\pi_Y(Y)$ the prior distribution of Y.

2.3.2 Likelihood function

The measurements are assumed to be corrupted by an observation noise having independent Gaussian distribution with mean zero and covariance matrix $\alpha \mathrm{I}, \alpha \in \mathbb{R}_+$. The misfit function χ used to quantify the quality of a proposition is then the standard l_2-norm,

$$\chi(Y|t^{\text{obs}}) := \frac{1}{\alpha^2} \sum_{i=1}^{N} \left(t_i^{\text{obs}} - t_i(\mu, \boldsymbol{\xi})\right)^2, \tag{17}$$

and its associated Gaussian likelihood writes

$$\mathcal{L}(t^{\text{obs}}|Y) := \frac{1}{\sqrt{(2\pi\alpha^2)^N}} \exp\left(-\frac{1}{2}\chi(Y|t^{\text{obs}})\right). \tag{18}$$

We remark that maximizing the likelihood is equivalent to minimizing the misfit function.

2.3.3 Prior distributions

Since the components of Y are independent, except for the K–L coordinates $\boldsymbol{\xi}$ that depend on q, the prior distribution of Y is

$$\pi_Y(Y) = \pi_\mu(\mu)\pi_{\boldsymbol{\xi}}(\boldsymbol{\xi}|q)\pi_q(q)\pi_\alpha(\alpha).$$

An uniform prior is used for $\mu \in [\mu_-, \mu_+] \subset \mathbb{R}$,

$$\pi_\mu(\mu) = \frac{1}{\mu_+ - \mu_-}.$$

As $\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma(q))$, its prior is

$$\pi_{\boldsymbol{\xi}}(\boldsymbol{\xi}|q) = \frac{1}{\sqrt{(2\pi)^r \det(\Sigma(q))}} \exp\left(-\frac{\boldsymbol{\xi}^\top \Sigma(q)^{-1}\boldsymbol{\xi}}{2}\right).$$

Uniform priors are also used for the hyperparameters $q \in [A_-, A_+] \times [\ell_-, \ell_+] \subset \mathbb{R}_+^2$,

$$\pi_q(q) = \frac{1}{\ell_+ - \ell_-} \times \frac{1}{A_+ - A_-}.$$

A truncated Jeffreys prior is chosen for the noise parameter $\alpha \in [\alpha_-, \alpha_+] \subset \mathbb{R}_+$,

$$\pi_\alpha(\alpha) \propto \frac{1}{\alpha}.$$

2.4 Markov chain Monte–Carlo sampling

The above mentioned posterior distribution does not have a closed-form expression and is sampled using the *Metropolis–Hastings* (MH) algorithm [5, 12, 15]. The principle is to build a MCMC $(Y^{(i)})_{i \in \mathbb{R}_+}$ that converges to a stationary distribution of parameters corresponding to their most probable values. The different steps of the procedure are listed in Algorithm 1. The

Algorithm 1 Metropolis–Hastings algorithm.

Input: $Y^{(0)} \sim \pi_Y(Y), N$ ▷ Initialization

1: **for** $n = 1, \dots, N$ **do** ▷ Iterations

2: $Y_{\text{new}} \sim \pi_{\text{tr}}(\cdot, Y^{(n)})$, ▷ New state proposal

3: $r_{\text{MH}} = \min\left(1, \dfrac{p_{\text{post}}(Y_{\text{new}}|t^{\text{obs}})\pi_{\text{tr}}(Y_{\text{new}}, Y^{(n)})}{p_{\text{post}}(Y^{(n)}|t^{\text{obs}})\pi_{\text{tr}}(Y^{(n)}, Y_{\text{new}})}\right)$, ▷ MH ratio

4: $Y^{(n+1)} = \begin{cases} Y_{\text{new}} & \text{if } u \sim \mathcal{U}(0,1) < r_{\text{MH}}, \\ Y^{(n)} & \text{else.} \end{cases}$ ▷ Acceptance or rejection

5: **end for**

Output: $\{Y^{(n)}\}$

use of the MH criterion in line 3 guarantees the respect of the detailed balance principle. Indeed, if the new state is more probable than the existing point, it is always accepted. However, if it is a less probable state, the larger the relative drop in probability, the more likely the state is rejected. Therefore, the chain tends to stay in high-probability regions of $p_{\text{post}}(Y)$, while only occasionally visiting low-probability regions.

The proposal distribution used in line 2 is assumed to be Gaussian, that is

$$Y_{\text{new}} \sim \pi_{\text{tr}}(\cdot, Y^{(n)}) := \mathcal{N}(Y^{(n)}, C^{(n)}), \tag{19}$$

where the covariance matrix $C^{(n)} \in \mathbb{R}^{d\times d}$ is adapted every N_{adapt} iterations as done in [11] as follows,

$$C^{(n)} = \begin{cases} C^{(0)} \text{ given} & \text{if } n = 0, \\ s_d \mathbb{C}\text{ov}\left(Y^{(0)}, \cdots Y^{(n)}\right) & \text{if } \text{mod}(n, N_{\text{adapt}}) = 0 \text{ and } N \leq N_{\text{burn-in}}, \\ C^{(n-1)} & \text{else.} \end{cases}$$

The initial covariance matrix $C^{(0)}$ is chosen to be a diagonal matrix with entries (corresponding to the variances of each parameters) proportional to the prior intervals. The factor s_d only depends on the dimension d, and its value is chosen to be $s_d = 2.4^2/d$ that corresponds to the optimal mixing properties in the case of Gaussian targets and proposals [8]. Drawing a new proposal Y_{new} is obtained from

$$Y_{\text{new}} = Y^{(n)} + U\Lambda^{1/2}U^\top X_{\text{new}}, \quad X_{\text{new}} \sim \mathcal{N}(0, \mathrm{I}_d),$$

where the unitary matrix U and the diagonal matrix Λ come from the *singular value decomposition* (SVD) $C^{(n)} = U\Lambda U^\top$.
The posterior distribution (16) needs to be computed.

Proposition 2. *Considering that the parameters variables are independent, and that the likelihood is based on the squared Mahalanobis distance as in* (18)*, it holds*

$$\pi(Y|t^{\text{obs}}) \propto \frac{1}{\alpha^{N+1}\det(\Sigma(q))^{1/2}} \exp\left(-\frac{\boldsymbol{\xi}^\top\Sigma(q)^{-1}\boldsymbol{\xi}}{2}\right) \prod_{i=1}^{N} \exp\left(-\frac{1}{2}\left(\frac{t_i^{\text{obs}} - t_i(\mu, \boldsymbol{\xi})}{\alpha}\right)^2\right), \quad (20)$$

where R is a constant independent from Y.

Proof of Proposition 2. Using the prior distributions defined in Section 2.3.3, it holds

$$\pi_Y(Y) = \frac{1}{\mu_+ - \mu_-}\frac{1}{\ell_+ - \ell_-}\frac{1}{A_+ - A_-}\frac{1}{\alpha}\frac{1}{\sqrt{(2\pi)^r \det\left(\Sigma(q)\right)}} \exp\left(-\frac{\boldsymbol{\xi}^\top\Sigma(q)^{-1}\boldsymbol{\xi}}{2}\right).$$

As for the likelihood, using the expression of the misfit function (17) inside the definition of the likelihood (18), we have

$$\begin{aligned} \mathcal{L}(t^{\text{obs}}|\mu, \boldsymbol{\xi}, \alpha) &= \frac{1}{(2\pi)^{N/2}\alpha^N} \exp\left(-\frac{\sum_{i=1}^{N}\left(t_i^{\text{obs}} - t_i(\mu, \boldsymbol{\xi})\right)^2}{2\alpha^2}\right), \\ &= \frac{1}{(2\pi)^{N/2}\alpha^N} \prod_{i=1}^{N} \exp\left(-\frac{1}{2}\left(\frac{t_i^{\text{obs}} - t_i(\mu, \boldsymbol{\xi})}{\alpha}\right)^2\right). \end{aligned}$$

Finally, according to the Bayes' formula (16), the posterior distribution writes

$$\begin{aligned} p_{\text{post}}(Y|t^{\text{obs}}) &\propto \frac{1}{(2\pi)^{N/2}\alpha^N}\prod_{i=1}^{N}\exp\left(-\frac{1}{2}\left(\frac{t_i^{\text{obs}}-t_i(\mu,\boldsymbol{\xi})}{\alpha}\right)^2\right) \\ &\quad\times\ \frac{1}{\alpha}\frac{1}{\sqrt{(2\pi)^r\det(\Sigma(q))}}\exp\left(-\frac{\boldsymbol{\xi}^\top\Sigma(q)^{-1}\boldsymbol{\xi}}{2}\right), \\ &= \frac{1}{(2\pi)^{(N+r)/2}\alpha^{N+1}\det(\Sigma(q))^{1/2}}\prod_{i=1}^{N}\exp\left(-\frac{1}{2}\left(\frac{t_i^{\text{obs}}-t_i(\mu,\boldsymbol{\xi})}{\alpha}\right)^2\right) \\ &\quad\times\exp\left(-\frac{\boldsymbol{\xi}^\top\Sigma(q)^{-1}\boldsymbol{\xi}}{2}\right), \end{aligned}$$

since the constant factor due to the uniform priors of hyperparameters does not need to be considered. The expression (20) therefore holds. □

Numerically, it is better to consider the logarithm of r_{MH}. The MH criterion thus becomes: "Y_{new} is accepted if $-\log(r_{\text{MH}}) < -log(\boldsymbol{s})$" with $\boldsymbol{s}$ being the realization of an uniform law on $[0,1]$. Since $\pi_{\text{tr}}\left(Y_{\text{new}}, Y^{(n-1)}\right) = \pi_{\text{tr}}\left(Y^{(n-1)}, Y_{\text{new}}\right)$ according to the transition proposal defined in (19), this logarithm writes, in the case where the distribution ratio is inferior to 1 (else, this is trivial):

$$-\log(r_{MH}) = \log(p_{\text{post}}(Y^{(n-1)}|t^{\text{obs}})) - \log(p_{\text{post}}(Y_{\text{new}}|t^{\text{obs}})),$$

and using the Proposition 2, we obtain

$$\begin{aligned} -\log(p_{\text{post}}(Y|t^{\text{obs}})) &= R + (N+1)\log(\alpha) \\ &\quad + \frac{1}{2}\left[\log\det(\Sigma(q)) + \sum_{i=1}^{N}\left(\frac{t_i^{\text{obs}}-t_i(\mu,\boldsymbol{\xi})}{\alpha}\right)^2 + \boldsymbol{\xi}^\top\Sigma(q)^{-1}\boldsymbol{\xi}\right], \end{aligned}$$

where the constant R, independent from Y, will not appear after computing the difference.

Remark 4. The empirical mean of the chain states converges almost surely to the target distribution provided that the search space is bounded as proven in [11], that is why we set $\boldsymbol{\xi} \in [-20, 20]^r$ and $\alpha \in [1.10^{-6}, 0.1]$. □

2.5 Surrogate models

Surrogate models consist in replacing expensive and accurate representations by cheaper approximations. In this work, these estimates are written with the tilde symbol $\widetilde{\cdot}$.

2.5.1 Traveltimes computation

In order to reduce the computational cost due to the extensive MCMC sampling, PC approximations $\{\widetilde{t}_i\}_{1\leqslant i\leqslant N^{\text{obs}}}$ of the N^{obs} traveltimes $\{t_i\}_{1\leqslant i\leqslant N^{\text{obs}}}$ are computed,

$$\forall 1 \leqslant i \leqslant N^{\text{obs}}, \quad t_i(\mu, \boldsymbol{\xi}) \simeq \widetilde{t}_i(\mu, \boldsymbol{\xi}) := \sum_{\kappa\in\mathcal{K}} t_{i,\kappa}\phi_\kappa(\mu, \boldsymbol{\xi}), \tag{21}$$

where $\mathcal{K}$ is a set of multi-indices, $\{t_{i,\kappa}\}_{1\leqslant i\leqslant N^{\text{obs}},\kappa\in\mathcal{K}}$ are coefficients to determine, and $\{\phi_\kappa\}_{\kappa\in\mathcal{K}}$ denotes the polynomial basis functions [10, 23]. The standard PC are multivariate polynomials orthogonal with respect to the uncertain parameters. As an example, Hermite and Legendre polynomials correspond to normal and uniform distributions, respectively.

In our case, the K–L coordinates $\boldsymbol{\xi}$ follow $\mathcal{N}(0, \Sigma(q))$, which depends on q. Yet, the goal is that the likelihood, and consequently the surrogate, does not depend on the hyperparameters in order to reduce the number of variables to $r+1$ instead of $r+1+|q|$. We choose $\boldsymbol{\xi} \sim \mathcal{N}(0, I_r)$ that is satisfied by the coordinates $\boldsymbol{\xi}$ on average over the hyperparametric domain since $\mathbb{E}_q(\Sigma(\cdot)) = \mathrm{I}_r$ from equation (14). The spectral coefficients $\{t_{i,\kappa}\}_{1\leqslant i\leqslant N^{\text{obs}},\kappa\in\mathcal{K}}$ in equation (21) are computed using the sparse *pseudo spectral projection* (PSP) [17], see Appendix B for more details. In this method, N_s *sparse grid* (SG) nodes are used to sample the standardized parametric domain $[0, 1]^{r+1}$, say

$$\left(\mu_{SG}^{(s)}, \boldsymbol{\xi}_{SG}^{(s)}\right)_{1\leqslant s\leqslant N_s}.$$

The physical scaling are

$$\mu^{(s)} = \mu_- + (\mu_+ - \mu_-)\mu_{SG}^{(s)}, \quad \text{and} \quad \boldsymbol{\xi}^{(s)} = F_{\mathcal{N}}^{-1}\left(\boldsymbol{\xi}_{SG}^{(s)}\right),$$

where $F_{\mathcal{N}}^{-1}$ is the inverse of the cumulative distribution function associated to a standard Gaussian variable. In order to obtain the final surrogate, the steps presented in Algorithm 2 are performed.

2.5.2 Covariance matrix of the change of measure

We recall that Σ is defined in equation (12) as the projection of the parametrized covariance kernel on the reference basis:

$$\Sigma_{i,j}(q) = (\overline{\lambda_j\lambda_i})^{-1/2}\left\langle \langle k(\boldsymbol{x}, \cdot, q), \overline{u}_j\rangle_\Omega, \overline{u}_i\right\rangle_\Omega.$$

The prior probability of $\boldsymbol{\xi}$ requires to compute $\Sigma(q)^{-1}$ and $\det(\Sigma(q))$. In order to guarantee the positivity of the surrogate $\widetilde{\Sigma}(q)^{-1}$, the PC method approximates rather $\Sigma(q)^{-1/2}$ using its

Algorithm 2 Computation of the traveltime surrogate.

Input: $\left(\mu_{SG}^{(s)}, \boldsymbol{\xi}_{SG}^{(s)}\right)_{1\leqslant s\leqslant N_s}$ ▷ Initialization

1: **for** $s = 1, \ldots, N_s$ **do** ▷ Iterations

2: $\left(\mu^{(s)}, \boldsymbol{\xi}^{(s)}\right)$, ▷ Physical scaling

3: $v(\boldsymbol{x}) = \exp\left(\mu^{(s)} + \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x}) \xi_i^{(s)}\right)$, ▷ Associated velocity field

4: $t^{(s)} = (t_i(v))_{1\leqslant i\leqslant N^{\text{obs}}}$. ▷ Eikonal solver

5: **end for**

6: $\{t_{i,\kappa}\}_{1\leqslant i\leqslant N^{\text{obs}}, \kappa\in\mathcal{K}}$, ▷ PSP matrix computation

Output: $(\widetilde{t}_i(\mu, \boldsymbol{\xi}))_{1\leqslant i\leqslant N^{\text{obs}}}$

SVD. The PC surrogate of the coefficient $\Sigma(q)_{i,j}^{-1/2}$ has the following form:

$$\Sigma(q)_{i,j}^{-1/2} \simeq \widetilde{\Sigma}(q)_{i,j}^{-1/2} = \frac{1}{\sqrt{A}} \sum_{\kappa\in\mathcal{K}} s_{i,j,\kappa} \psi_\kappa(\ell),$$

where $s_{i,j,\kappa}$ are the spectral coefficients associated to the multi-index κ. Indeed, it is equivalent to consider only $q \setminus \{A\}$ as input for the surrogate and to scale the approximation by $1/\sqrt{A}$ as explained in Appendix D. The method used to compute this surrogate is detailed in Algorithm 3. The regularity of Σ in ℓ needs to be analyzed in order to define the maximal order of polynomials.

Algorithm 3 Computation of the covariance matrix surrogate.

Input: $\{\ell^{(1)}, \cdots, \ell^{(N_s)}\}$ ▷ Initialization with *Latin hypercube sampling* (LHS) method [13]

1: **for** $s = 1, \ldots, N_s$ **do** ▷ Iterations

2: $C = C(1, \ell^{(s)})$, ▷ Covariance computation

3: $\text{Sig} = \Sigma(1, \ell^{(s)})$, ▷ Projection on the normalized reference basis

4: $\mathfrak{s}^{(s)} = U^\top \Lambda^{-1/2} U$. ▷ By SVD of Sig

5: **end for**

6: $S = \left[\psi_i(\ell^{(s)})\right] \in \mathbb{R}^{|\mathcal{K}|, N_s}$ ▷ Matrix of PC basis functions evaluated at the LHS points

7: $\boldsymbol{s}_{i,j} = (S^\top S)^{-1} S^\top \mathfrak{s}_{i,j}$ ▷ Vector of spectral coefficients, using *ordinary least-square* (OLS) method

Output: $\widetilde{\Sigma}^{-1/2}(A, \ell) = \frac{1}{\sqrt{A}} \widetilde{\Sigma}^{-1/2}(1, \ell)$ ▷ Final surrogate

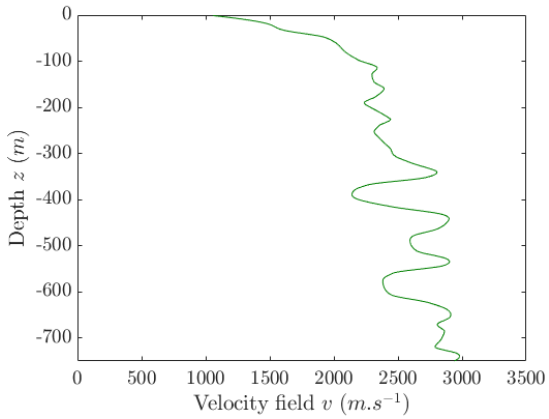
Concerning the determinant, no surrogate is built since it does not improve the accuracy, neither the computational efficiency.

3 Application to a one dimensional field

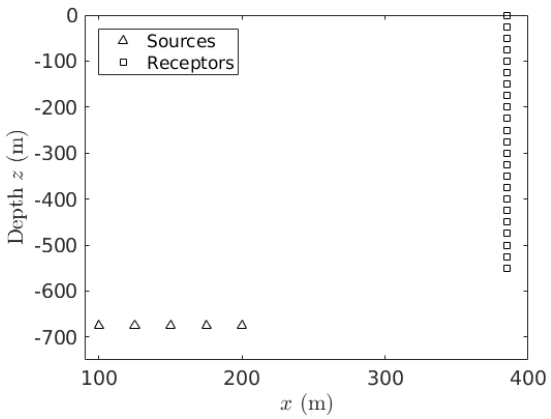
The method studied in the previous section is applied to an example defined in Section 3.1. The different choices for the implementations are presented in Section 3.2 and the reference basis and its interest in this modeling are analyzed. Then, the same work is led concerning the PC surrogates in Section 3.3. Finally, results of the MCMC inference are presented and discussed in Section 3.4.

3.1 Case presentation

We consider a continuous one dimensional velocity field v_{true} which is represented in Figure 3a. This field is obtained from the realistic model Amoco [16], however, the associated observations are computed thanks to a resolution of equation (1) by the eikonal solver (direct problem associated to the inverse problem we consider) and then a Gaussian noise of 0.001s is added[1]. This construction allows us to verify our algorithm on an example for which we control the observations. The set of observations is composed of the times of arrival in 23 receptors from 5 sources, whose locations are detailed in Figure 3b. It is worth to note that the region of inference depends largely of the geometry of the equipments: if no waves travel through an area, it is a shaded zone and there is no information on it given by the objective function (2) used in the inference. This remark is illustrated by the results of a *sensitivity analysis* (SA) in Appendix C.



(a) True velocity field v_{true}.



(b) Sources and receptors locations

Figure 3: Continuous 1-D case: field and observations.

[1]For each observation, its value is set to $t_i^{\text{obs}} = t_i(v_{\text{true}}) + X$, with $X \sim \mathcal{N}(0, 0.001^2)$.

3.2 Choice for implementations

Choice of parameters' space. Even if the inference of parameters is a method used to not define *a priori* the value of μ, A and ℓ, it is necessary to set *a priori* bounds in order to compute the associated reference basis and to have a bounded search space for the MCMC algorithm. The choice is based on a physical *a priori*. The search space must be composed of velocity field with values that are between 500 and 6000 m.s^{-1}, based on the realistic velocities P waves can achieve in the first kilometer of the Earth's crust.

In order to assess the pertinence of different values, and to see the field dependency to the parameters, several Gaussian processes are generated and the corresponding velocity fields are plotted in Figure 4. The role of each parameter is clear:

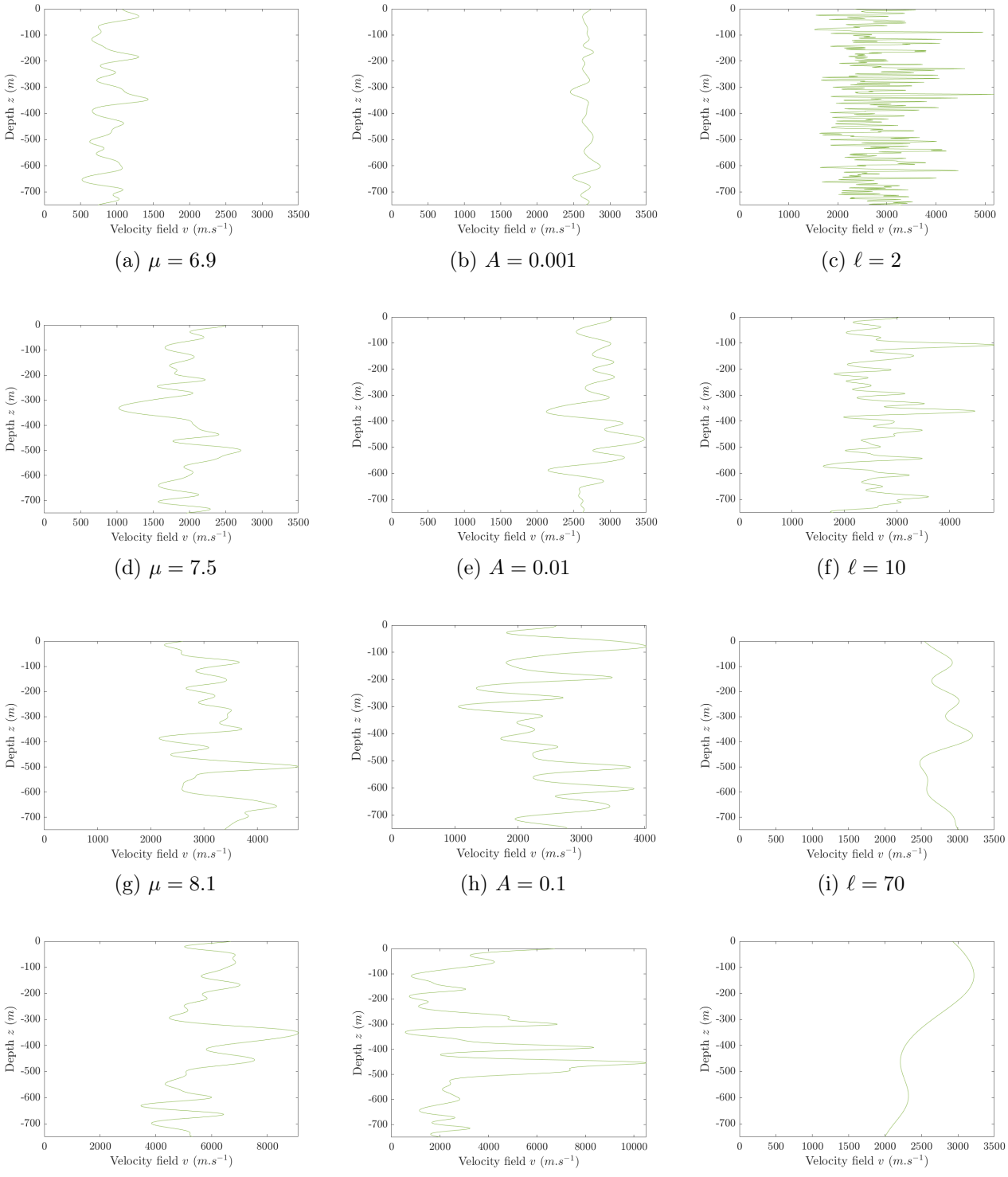
- the mean μ defines the mean range of the values. Since the field is obtained by computing the exponential, it is a factor term, which also explains the scale of the variation obtained when increasing μ,
- the amplitude A defines as expected the amplitude of the values the velocity can take. The larger A, the larger the spatial variance of the field. Since we consider here a continuous case, too large amplitudes do not interest us,
- the length of correlation ℓ defines the length scale needed for the field to vary, and thus its homogeneity. Here, very small length of correlation are not pertinent, the length of correlation is around 1% to 20% of the spatial domain size.

The priors of these three parameters are assumed to be uniform on the intervals given in Table 1.

Parameter	Bounds
μ	$[6.9, 8.1]$
A	$[1.10^{-6}, 1.10^{-1}]$
ℓ	$[10, 150]$

Table 1: Bounds used for shape parameters.

Computation of the reference basis. Thanks to the priors that we defined, the reference basis is computed as defined in equations (10)-(11), considering that the reference covariance kernel $\overline{k}$ is the limit of an empirical mean of random realizations. The convergence of this method is shown in Figure 5. After this computation, it is possible to verify that the projection



(a) $\mu = 6.9$ (b) $A = 0.001$ (c) $\ell = 2$

(d) $\mu = 7.5$ (e) $A = 0.01$ (f) $\ell = 10$

(g) $\mu = 8.1$ (h) $A = 0.1$ (i) $\ell = 70$

(j) $\mu = 8.7$ (k) $A = 1$ (l) $\ell = 150$

Figure 4: Velocity field v for different values of parameters μ, A and ℓ. Obtained considering $v = \exp(u_{\mathrm{KL}})$, where u_{KL} is the realization of a Gaussian process, with by default $\mu = 7.9$, $A = 0.05$ and $\ell = 20$.

on this basis is a good approximation of the true field. This approximation writes as

$$v = \exp(u) = \exp\left(\mu + \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i \xi_i\right), \text{ with } \xi_i = \overline{\lambda}_i^{-1/2} u_{\text{true}}^{\top} \overline{u}_i.$$

We compare the true field to the projected one in Figure 6. The field and its approximation match almost exactly. When comparing the error on times of arrival with the eikonal solver, it is clear that the error due to the projection is dominated by the noise present in the data used for the inference. The error represented in Figure 6b is the absolute difference between true times of arrival and computed ones, which corresponds to L^1-norm errors. The projection error is in the order of 10^{-4}s while the noise of observation is around 10^{-3}s.

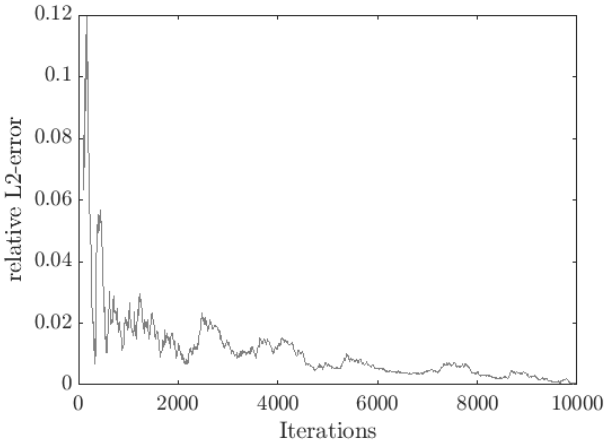
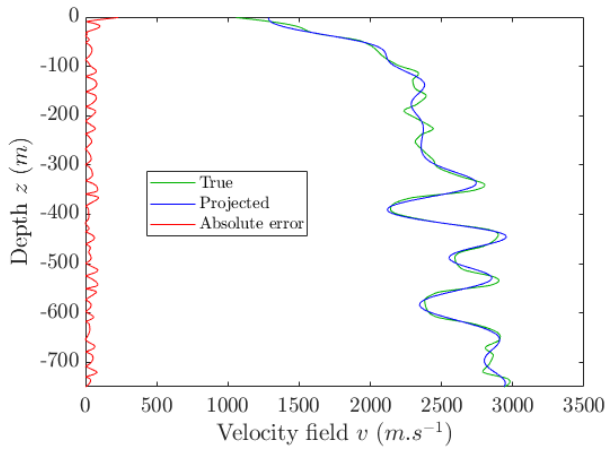


Figure 5: Convergence of the empirical mean when computing the reference kernel $\overline{k}$.

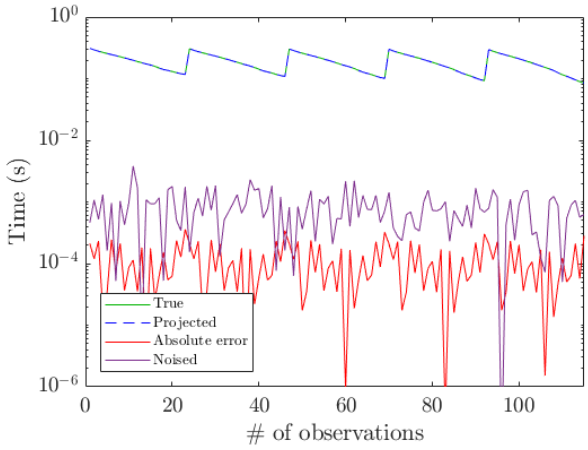
Interest of reference basis. The first advantage of the reference basis is to avoid a strong *a priori* on the shape of the searched velocity field without increasing a lot the computational cost. Moreover, the following analysis highlights the adaptation to various fields, in contrary to parametrized basis, and the independence of the reference basis to the number of K–L modes used.

In order to assess the pertinence of the reference basis when the shape of the field is not known, the study introduced in Algorithm 4 is led. The mean normalized projection error is then compared for the different bases. This comparison is shown in Figure 7. More precisely, in Figure 7a, the mean normalized projection error is plotted according to the number of K–L terms r and the used basis:

$$E_{\mathcal{B}}(r) = \mathbb{E}_{\ell}(\varepsilon_{\mathcal{B}}(r, \ell)),$$



(a) True field (green), projected field (blue), absolute error (red)



(b) Comparison between t^{true} and $t(v_{\text{proj}})$, and t^{obs}

Figure 6: Comparison of the field projected in the reference basis with the true field. $r = 20$.

Algorithm 4 Study of reference basis interest.

Input: $L_\ell = [10, 150]$, N_g number of Gaussian processes, ▷ Initialization

1: **for** $\ell \in L_\ell$ **do** ▷ Iterations

2: $\quad L_{\mathcal{B}}(\ell) = (u_i(\ell), \lambda_i(\ell))_{1 \leqslant i \leqslant r}$, ▷ Associated basis

3: $\quad$ **for** $i = 1 \cdots N_g$ **do**

4: $\quad\quad g_i(\ell) = U_{\{1,\ell\}}(\cdot, \omega)$, ▷ Generation of Gaussian processes

5: $\quad$ **end for**

6: **end for**

7: **for** $\ell \in L_\ell,\ i = 1 \cdots N_g,\ \mathcal{B} \in L_{\mathcal{B}}$ **do**

8: $\quad g_{\text{KL}} = \langle g_i(\ell), \mathcal{B} \rangle$, ▷ Projection on bases

9: $\quad \text{err}(g_i(\ell), \mathcal{B}) = \|g_i(\ell) - g_{\text{KL}}\|^2_{L^2}$, ▷ Projection error

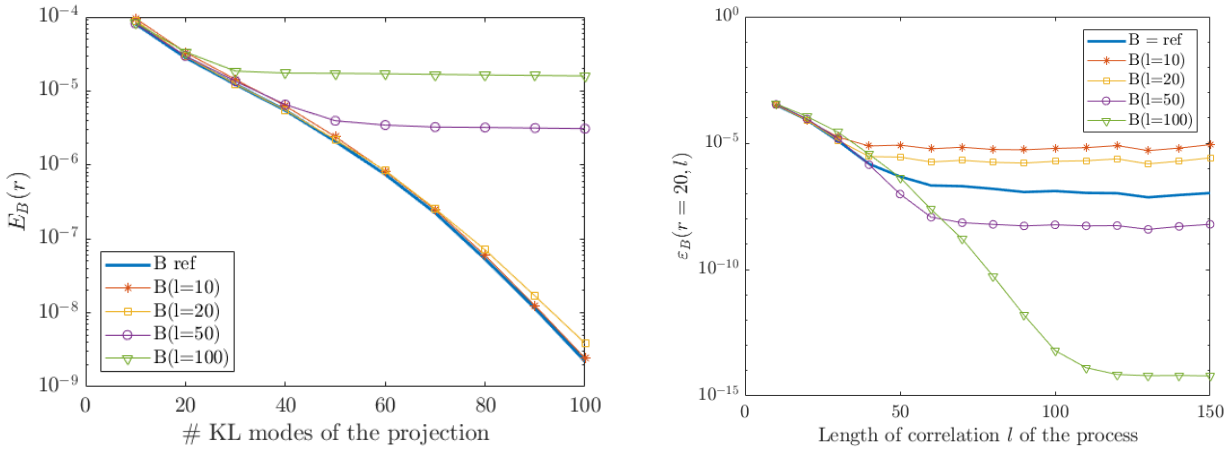
10: **end for**

Output: $\{\text{err}(g_i(\ell), \mathcal{B})\}_{1 \leqslant i \leqslant N_g, \ell \in L_\ell, \mathcal{B} \in L_{\mathcal{B}}}$

where $\mathbb{E}_\ell$ is approximated by Monte–Carlo sampling over ℓ, and

$$\varepsilon_{\mathcal{B}}(r, \ell) = \frac{\sum_i \text{err}(g_i(\ell), \mathcal{B})}{\sum_i \|g_i(\ell)\|^2_{L^2}}$$

is the normalized error for a fixed number of K–L terms, basis, and length of correlation for the Gaussian processes. In Figure 7b, r is set to 20 and the errors $\varepsilon_{\mathcal{B}}(r, \ell)$ are compared according to ℓ and $\mathcal{B}$. The reference basis is thus a way to minimize the error of representation when not



(a) Mean (over ℓ) normalized error according to r and $\mathcal{B}$.(b) Normalized error according to ℓ and $\mathcal{B}$ for $r = 20$.

Figure 7: Comparison of projection error for different bases.

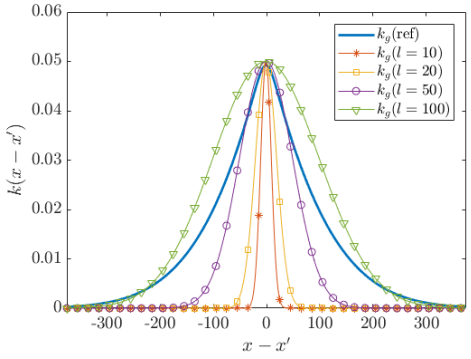
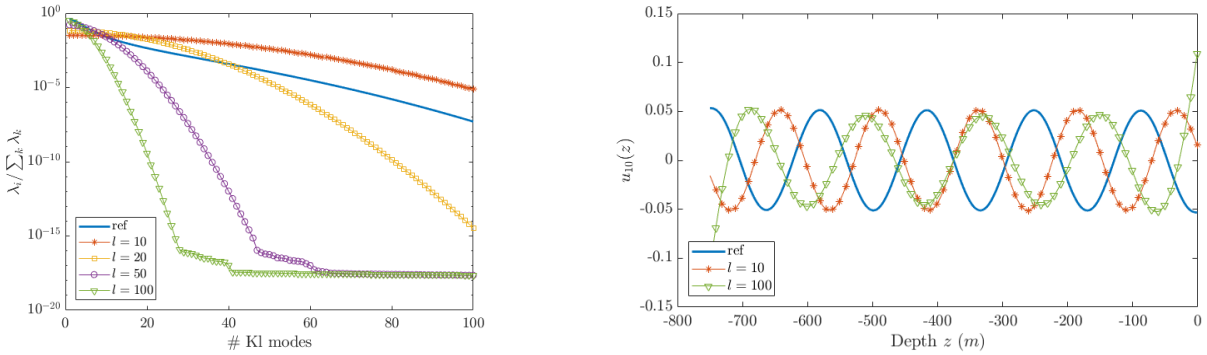


Figure 8: Shapes of different covariance kernels $k(\ell)$.

knowing the true parameters of the field as explained in [21]: this basis is not optimal for any ℓ, since the best in that case is the basis associated to ℓ, but it is the best choice on average over ℓ. Moreover, in the case where r is fixed, it also limits the maximal error to roughly 0.1% whatever the length of correlation of the process.

This result is linked to the analysis of the basis eigenelements and corresponding kernel shapes. These shapes are plotted in Figure 8. The larger the length of correlation, the larger the covariance kernel. The reference kernel is not gaussian. In Figure 9a, the decay rate $\lambda_i / \sum_k \lambda_k$ is represented for different bases. The smaller the length of correlation, the slower the decay rate. When the length of correlation is large, the information is concentrated on the first modes. In contrast, the decay rate for the reference basis is asymptotically similar (but with a smaller magnitude) to the smallest ℓ, while having the same magnitude as the



(a) Decay rate of eigenvalues $\lambda_i / \sum_k \lambda_k$

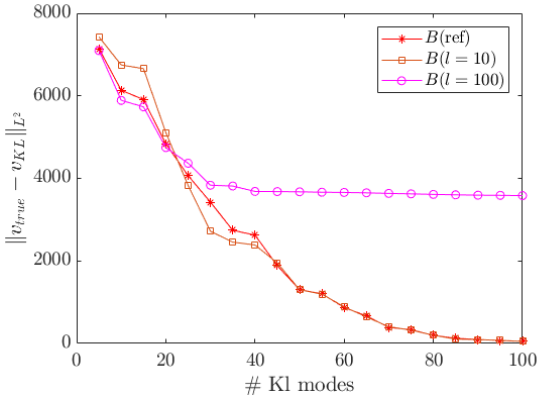
(b) Eigenvector u_{20}

Figure 9: Comparison of eigenelements for different bases.

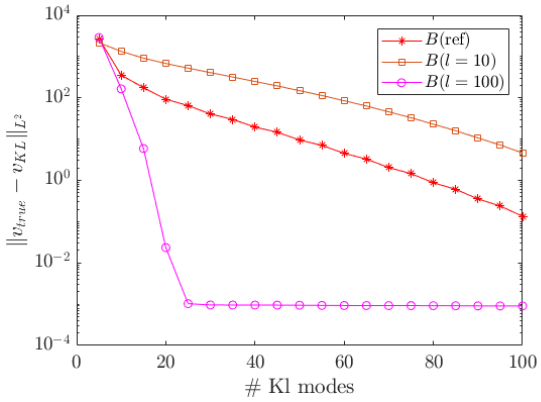
largest ℓ for the first modes. As for the eigenelements, represented in Figure 9b for $r = 20$, a smaller length of correlation leads to smaller scales of variations as expected, but also to smaller variations on bounds of the space. The reference eigenvectors seems to be a trade-off between the two extreme shapes. From a tomographic point of view, it can be interpreted by the fact that large lengths of correlation are good to infer with few modes the general variations of the fields, but that it cannot be accurate on small variations, even with increasing dimensions. In contrast, small lengths of correlation are good to infer small variations, but need numerous terms to infer the general shape, and cannot converge to a smooth functions, particularly at the boundaries. These extreme cases are shown in Figures 10,12,11.

For a field with a small length of correlation $\ell = 10$, the projection error on a projection basis with a large value of ℓ stalls when increasing dimension as shown in Figure 10a. The choice of a too large length of correlation leads to the impossibility to infer variations at small scales as shown in Figure 12c in comparison with Figures 12b-12a. For a field with a large length of correlation, the projection error is less important on both basis (Figure 10b), however, the choice of a too small value of ℓ prevents the projection to converge to the true field, particularly on the boundaries of the domain due to the shape of the eigenelements (Figure 11).

Note that the previous analysis is only led for ℓ, which affects the shape of the eigenvectors, and not for A, which is simply a scaling factor and thus integrated in the projection coordinates in that study, see Appendix D for details.



(a) Projection errors for a field with $\ell = 10$



(b) Projection errors for a field with $\ell = 100$

Figure 10: Influence of the basis on the projection error.

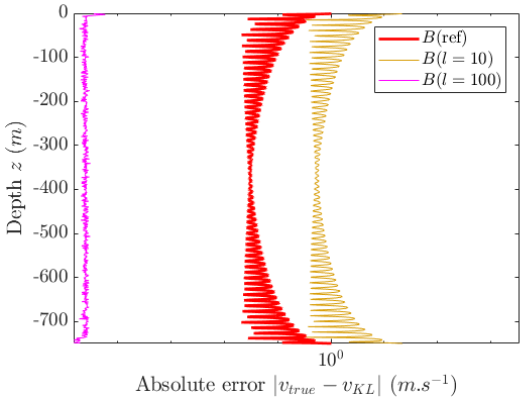
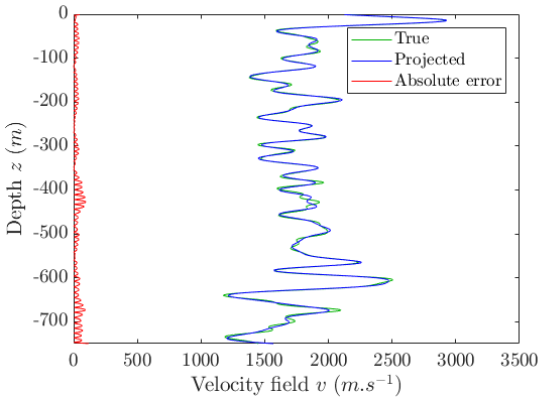
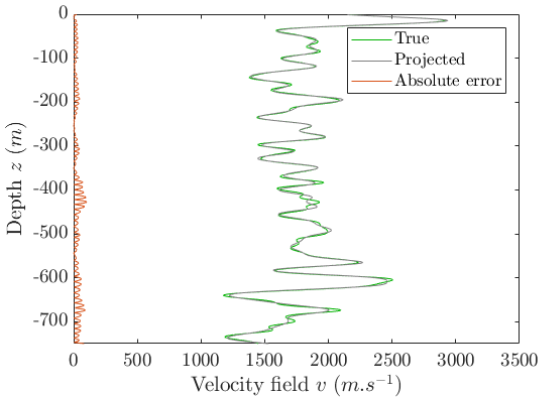


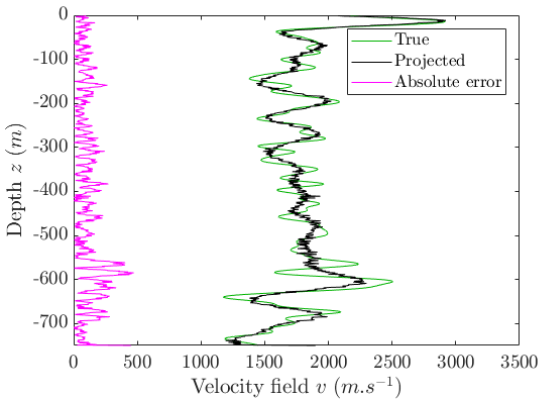
Figure 11: Absolute projection error for a large length of correlation field. $r = 60$.



(a) Projection on $\mathcal{B}$(ref)



(b) Projection on $\mathcal{B}(\ell = 10)$



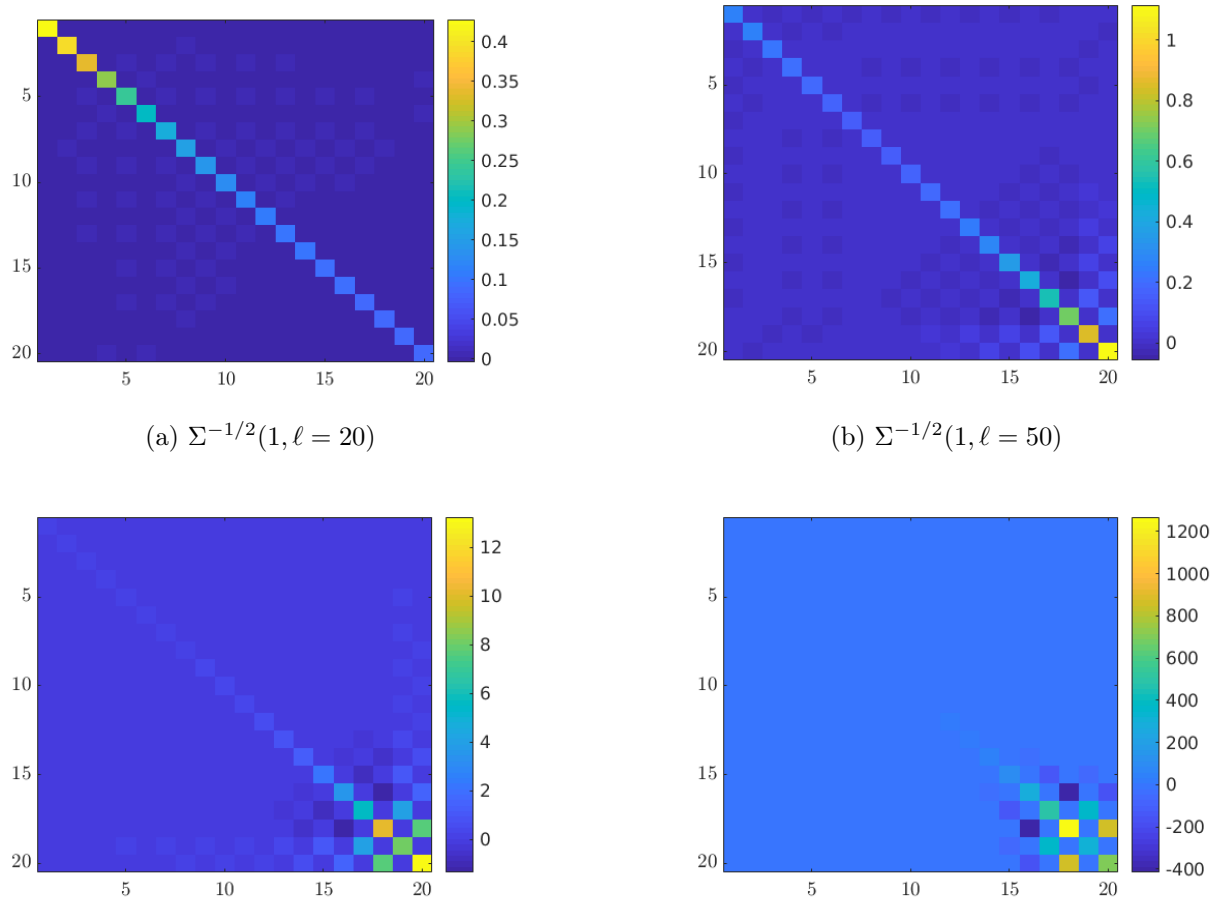
(c) Projection on $\mathcal{B}(\ell = 100)$

Figure 12: Projection for a small length of correlation field. $r = 60$.

3.3 Surrogate building and analysis

3.3.1 Surrogate for the change of measure

The surrogate of Σ is built as presented in Section 2.5.2. In Figure 13 are plotted different values for $\Sigma^{-1/2}(1, \ell)$.



(a) $\Sigma^{-1/2}(1, \ell = 20)$

(b) $\Sigma^{-1/2}(1, \ell = 50)$

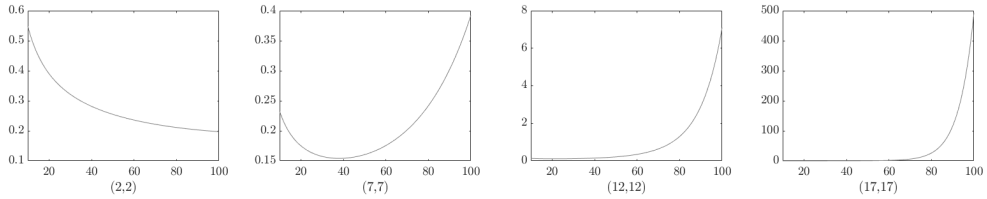
(c) $\Sigma^{-1/2}(1, \ell = 70)$

(d) $\Sigma^{-1/2}(1, \ell = 100)$

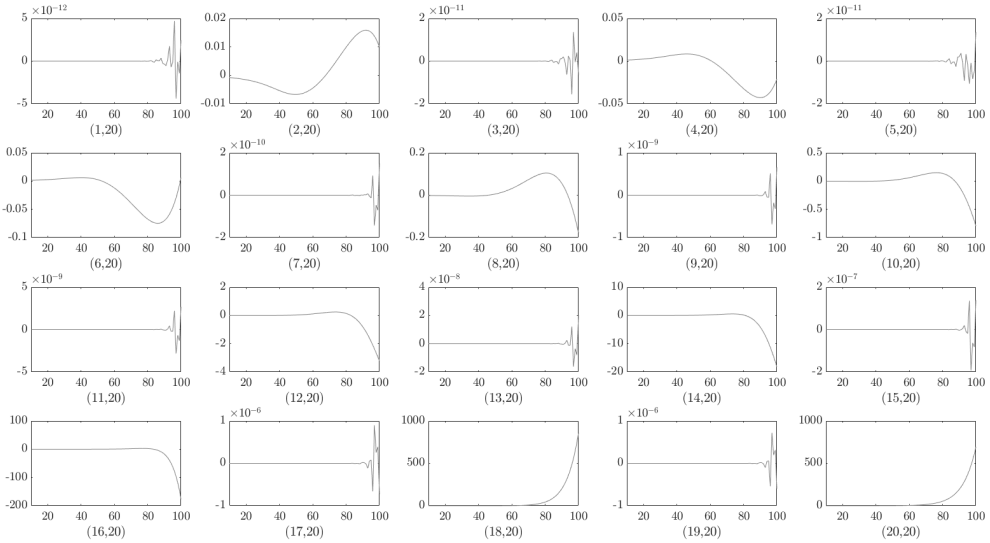
Figure 13: Values of $\Sigma^{-1/2}$ according to ℓ. $A = 1$.

Analysis of the coefficients. In order to better analyze the smoothness of the coefficients, they are plotted according to ℓ in Figure 14. These plots highlight two points:

1. the values of the diagonal terms are exploding when increasing ℓ;
2. the coefficients are either very smooth according to ℓ, either their values are negligible.



(a) Some diagonal terms of Σ according to ℓ



(b) Last colum terms of Σ according to ℓ

Figure 14: Coefficients of Σ according to ℓ. $A = 1$.

For the point 1, this is due to the behavior of $k(\cdot, \cdot, q)$ when ℓ is increasing: its limits is $1_{r\times r}$, which leads to degenerated behavior for the length of correlation we consider. In order to avoid that problem, partly induced by numerical inaccuracies, the superior bound for ℓ, which will be used to compute the surrogate of Σ and as bound for the search space in the MCMC algorithm, is determined according to the number of K–L terms used. The defined criterion is the following:

$$\frac{\lambda_r(\ell_+)}{\lambda_1(\ell_+)} > 10^{-9}.$$

This leads to consider $\ell_+ = 100$ for $r = 20$. As for the point 2, in order to reduce the computational cost and saving, a study is led to see if it is possible to disregard PC coefficients that are inferior to 10^{-6}, since small values in the matrix are associated to small values of the surrogate coefficients. For several values of A and ℓ, $\widetilde{\Sigma}^{-1/2}(q)$ is computed and compared to its true value via the computation of the *relative root mean squared error* (RRMSE) of $\pi_{\boldsymbol{\xi}}(\boldsymbol{\xi}|q)$,

where the RRMSE of a quantity is defined by

$$\mathrm{RRMSE}(Q_{\mathrm{true}}, Q_{\mathrm{approx}}) = \sqrt{\frac{\sum_{i=1}^{N}(Q_{\mathrm{true},i} - Q_{\mathrm{approx},i})^2}{\sum_{i=1}^{N} Q_{\mathrm{true},i}^2}}. \tag{22}$$

Here, the true quantities of interest $Q_{\mathrm{true},i}$ are $\pi_{\boldsymbol{\xi}}(\boldsymbol{\xi}^{(i)}|q)$ computed with $\widetilde{\Sigma}(q)$ and the approximated quantities of interest $Q_{\mathrm{approx},i}$ are $\pi_{\boldsymbol{\xi}}(\boldsymbol{\xi}^{(i)}|q)$ computed with $\widetilde{\Sigma}(q)$, setting the small coefficients to zero. The set of coordinates $(\boldsymbol{\xi}^{(i)})_{1\leqslant i\leqslant N_{\mathrm{VP}}}$ is obtained by drawing N_{VP} samples following a standardized Gaussian law[2]. The obtained RRMSEs are plotted in Figure 15. The value being close to the threshold of numerical precision, this confirms the possibility to disregard these coefficients.

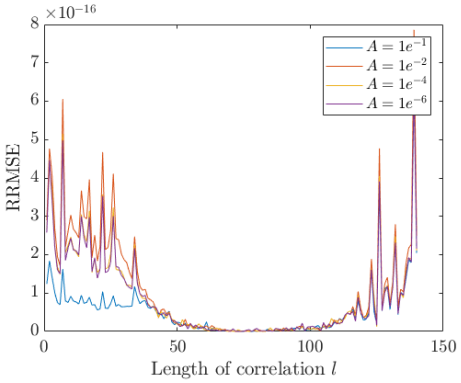
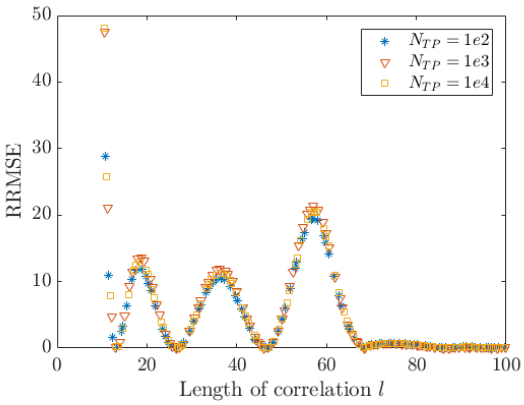


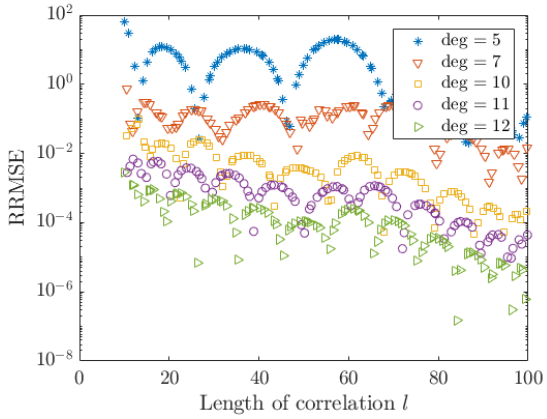
Figure 15: RRMSE for Σ surrogate when disregarding small surrogate coefficients.

Influence of the surrogate degree. The influence of the maximal degree for the polynomial function basis of the PC surrogates and the number of training point used to compute the PC coefficients is analyzed. As shown in Figures 16a-16b, the choice of the degree is much more important than the number of training point to approximate the values of interest. This can be explained by the general smoothness of the functions: several points are sufficient to describe them, however, the explosion in ℓ needs a high polynomial degree to be represented. Since the number of training points is therefore responsible for the increase of the time of computation, it is set to 1000. The last step is the choice of the maximal degree for the polynomial function basis of the PC surrogates. As shown in Figure 14, the dependence is rather smooth, however, some coefficients do have exponential behavior. In order to find a compromise between cost and accuracy, the degree of approximation is iteratively computed as presented in Algorithm 5.

[2]The number of samples N_{VP} is set to 100 in this study.



(a) RRMSE for $\widetilde{\Sigma}$ according to the number of training points N_{TP}



(b) RRMSE for $\widetilde{\Sigma}$ according to the surrogate degree

Figure 16: RRMSE for Σ surrogate according to the computations parameters. $r = 20$.

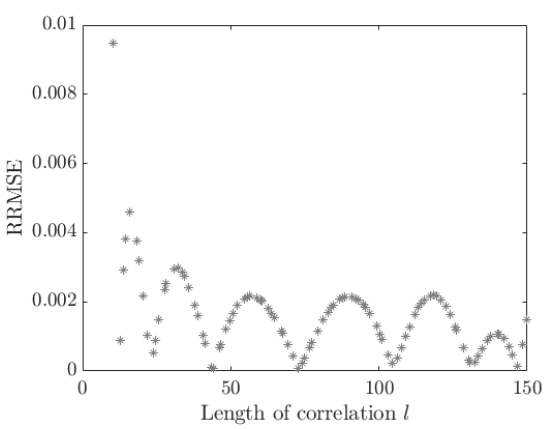
Algorithm 5 Choice of the PC surrogate degree.

Input: deg $= 5$ ▷ Initialization of the degree.
  $S = \widetilde{\Sigma}(\text{deg})$
  $R = \text{RRMSE}(S)$ ▷ Computed by (22).
  if $R > 0.01$ **then**
    deg$+ = 1$ ▷ Incrementation of the degree.
    $S = \widetilde{\Sigma}(\text{deg})$
    $R = \text{RRMSE}(S)$
  end if
Output: deg, S

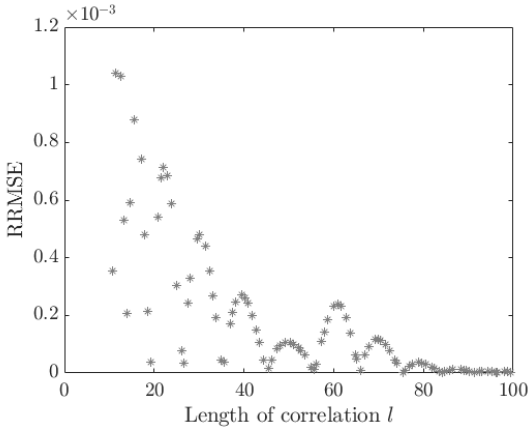
Results. In Figure 17 are presented the results for $r \in \{10, 20\}$. It is worth to note that the exponential explosion of the $\Sigma^{-1/2}$ coefficients leads directly to an explosion of the necessary degree and a fast decreasing of ℓ_+ when increasing the number of K–L terms r. For 1000 exact values with $r = 20$, the surrogate provides a $50\times$ speedup over the exact computation: the latter takes 0.82s versus 0.015s for the surrogate one.

3.3.2 Traveltime surrogate

The surrogate of the eikonal solver is computed as explained in Section 2.5.1. The polynomial basis functions level is set to 3, which limits the degree of monomial at 7 and the sum of the polynomials degree at 4. The input of the eikonal surrogate has dimension $r + 1$ (mean μ and



(a) RRMSE for $\widetilde{\Sigma}$ with $r = 10$. $\ell_+ = 150$, degree $= 6$



(b) RRMSE for $\widetilde{\Sigma}$ with $r = 20$. $\ell_+ = 100$, degree $= 12$

Figure 17: Results after building Σ surrogate for different number of K–L terms.

K–L coordinates $\boldsymbol{\xi}$). Choosing $r = 20$, the SG used to sample the training points contains 15135 points. Since in this study the MCMC algorithm is composed of 2×10^6 steps, this reduces the number of exact computations to less than 1% of the initial number.

The accuracy of the surrogate is assessed by computing the RRMSE defined in (22) with the quantity of interest being the times of arrival. The result is plotted in Figure 18. In average, the surrogate introduces an error of 1%. The speedup factor is even more important than the one for the Σ surrogate, since 1000 exact computations can take up to 30 minutes with the non-parallel current implementation, versus less than a second for the surrogate one.

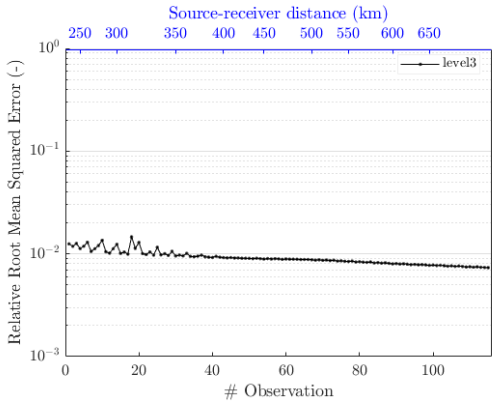


Figure 18: RRMSE of the eikonal surrogate $\widetilde{t}$ according to the observations.

Parameters	Value/Range
μ, A, ℓ	see Table 1
$\boldsymbol{\xi}$	$[-20, 20]^r$
$\log(\alpha)$	$[-10, -3]$
r	20
$N_{\text{burn-in}} = N_{\text{samp}}$	1.10^6
N_{adapt}	5.10^4
$\text{diag}(C^{(0)})$	$[0.1, 0.05, 0.002, 0.2, 0.5, 0.05]$

Table 2: Values/Range of parameters used in MCMC inference.

3.4 Results

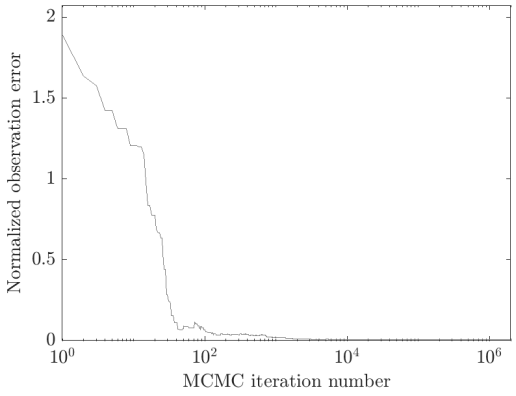
3.4.1 Results of inference in reference basis

The MCMC inference is led with the parameters listed in Table 2. The coordinates $\boldsymbol{\xi}$ follows $\mathcal{N}(0, \Sigma(q))$, the large bounds just ensure to keep a bounded distribution. The logarithm of the noise level α is used rather than its direct value in order to keep a relative symmetry and reversibility during the exploration. The noise indeed follows a Jeffrey's prior for the standard deviation of a Gaussian distribution [2, 14]. However, using this prior while exploring the possible states of the chain introduces a dissymmetry between the lower states of the current one (that are *a priori* more likely) and the higher one (that are *a priori* less likely). Using instead the log-uniform prior allows reestablishing the equilibrium. Denoting $\beta = \log(\alpha)$ and considering $\pi_\alpha(\alpha) \propto \frac{1}{\alpha}$, an inverse transformation gives

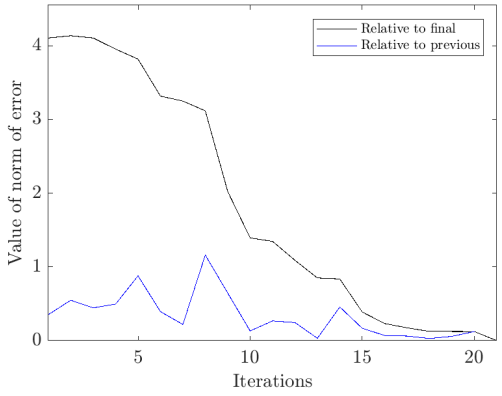
$$\pi_\beta(\beta) = \left|\frac{\partial \alpha}{\partial \beta}\right| \pi_\alpha(\alpha) \propto e^\beta \frac{1}{\alpha} = 1.$$

The number of steps $N_{\text{burn-in}}$ and N_{samp} respectively are the number of steps with adaptation of covariance matrix proposal every N_{adapt} steps and the number of steps used to sample the posterior distribution.

First observations are the convergence of the computed times toward the observation times along iterations, and the convergence of the covariance matrix proposal, plotted respectively in Figures 19a and 19b. The error on observations is rapidly decreasing, and the covariance matrix seems to converge to its final state since the Frobenius norm $\left\|C^{(N_{\text{adapt}})} - C^{(n)}\right\|_{\mathcal{F}}$ is decreasing while the norm of the difference between two successive covariance matrices tends to diminish.



(a) Convergence of times: error $\left\|t^{\text{obs}} - \widetilde{t}(Y^{(n)})\right\|$



(b) Convergence of covariance matrix proposal: Frobenius norm

Figure 19: Convergence study for the MCMC algorithm.

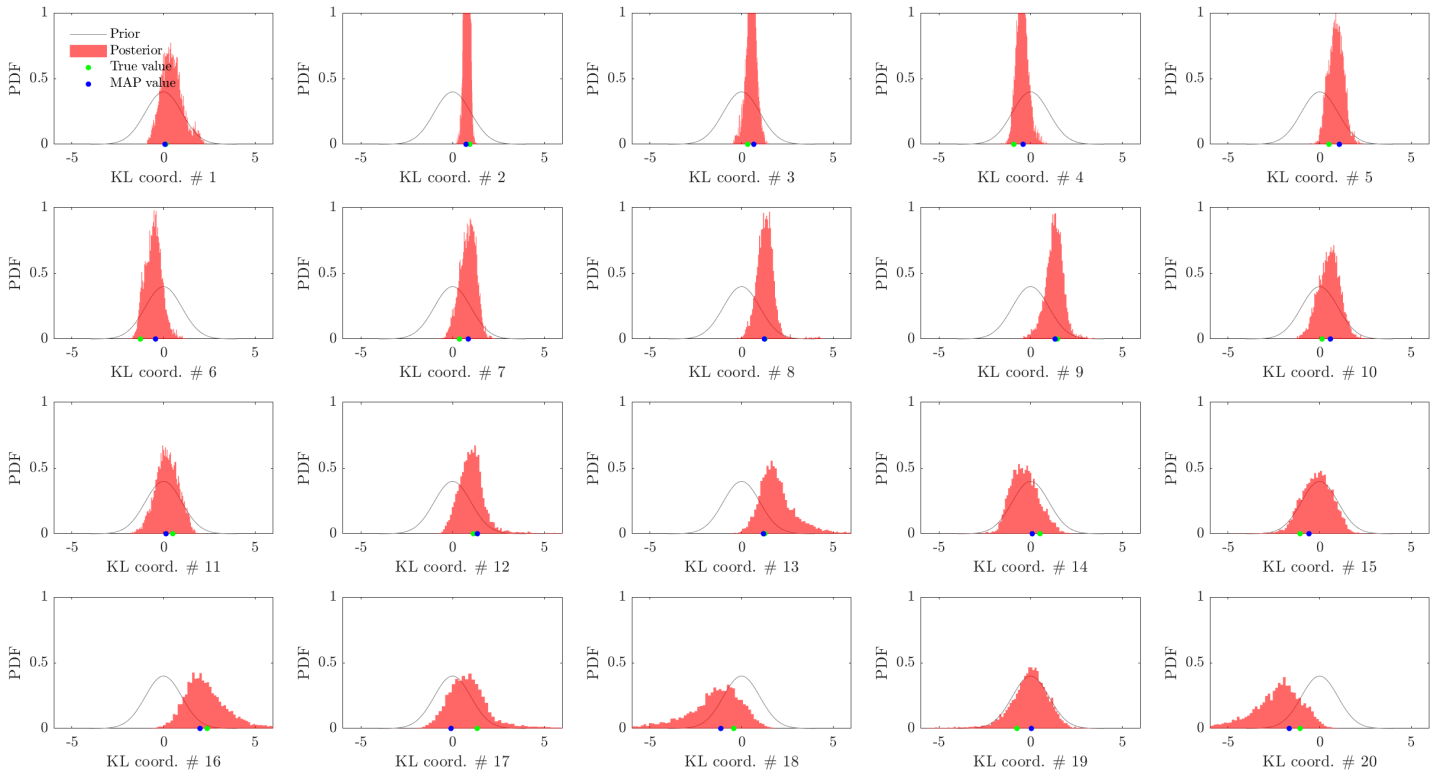


Figure 20: Marginal posterior distributions of the K–L coordinates $\boldsymbol{\xi}$.

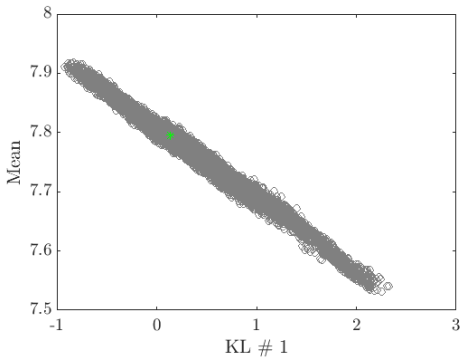
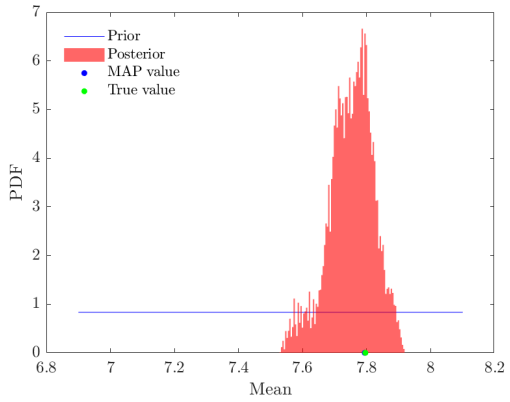


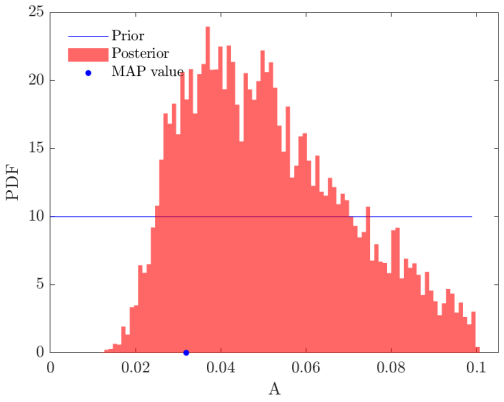
Figure 21: Correlation between the first K–L coordinate ξ_1 and the mean μ.

In Figure 20 are represented the marginal distributions of the K–L coordinates, which correspond to the histogram of values taken in the sampling step. These distributions are compared with the average along q with prior distribution $\mathcal{N}(0,1)$. The first coordinates are

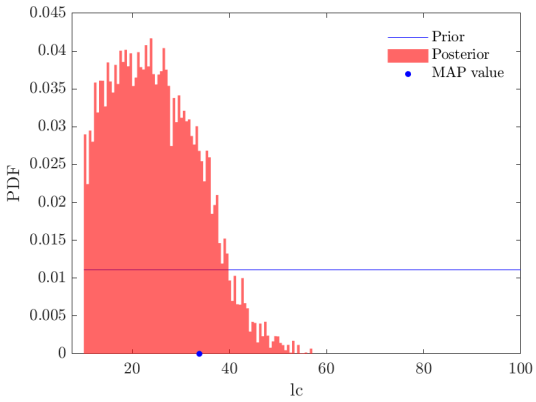
rather well inferred, with narrow distributions containing the K–L projection coordinates, while the distribution of the last ones remain close to their prior. A notable exception to this fact is the first coordinate. It is due to the correlation between the mean μ and ξ_1, which is shown in Figure 21. As both μ and ξ_1 are evolving along iterations, their values can compensate thanks to the very smooth shape of $\overline{u}_1$. This leads to enlarge the posterior distribution support since allowing more exploration.



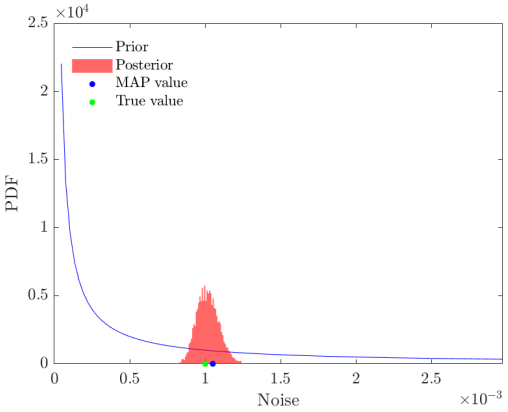
(a) Marginal posterior distribution of the mean μ



(b) Marginal posterior distribution of the amplitude A



(c) Marginal posterior distribution of the length of correlation ℓ



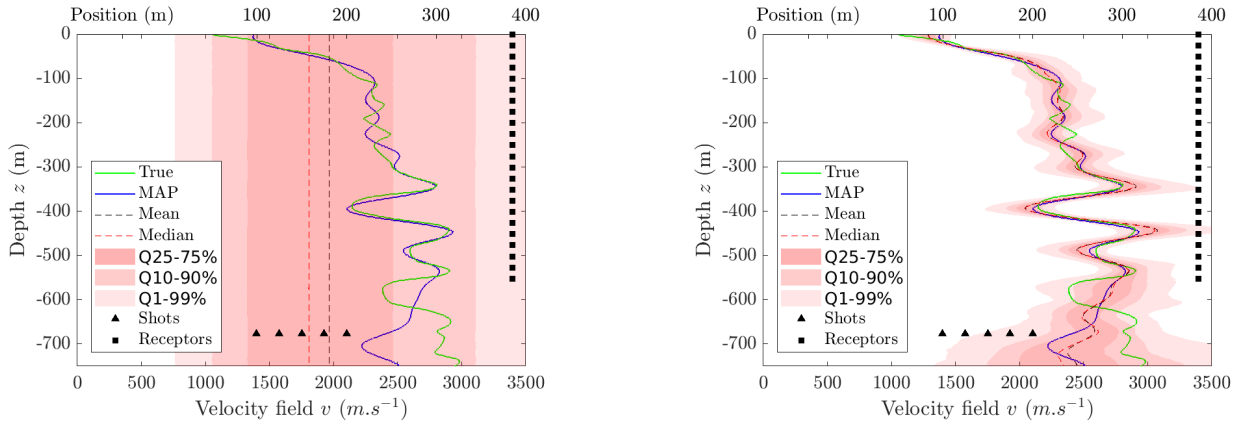
(d) Marginal posterior distribution of the noise level α

Figure 22: Marginal posteriors distribution of the parameters.

The marginal posterior distributions of the other parameters are plotted in Figure 22. The selection of the subspace for the length of correlation during the MCMC algorithm is shown in Figure 22c: even if the search space is very large (from 10 to 100), the maximum sampled length scale is around 60. This ensures the possibility to use large search spaces when not

having a strong *a priori* information without committing errors. It is also worth to note that the marginal distribution of the noise level α is similar to a Gaussian centered on the true value of noise. This validates the use of the log-uniform prior.

The distribution of Y is then used to compute a distribution of velocity fields by extracting 1% of the sampling chain elements and calculating the corresponding field. This posterior distribution is compared to the prior distribution in Figure 23. The MCMC algorithm is reducing the uncertainty of the field while keeping the true velocity field in the probable space. The area below $650m$ keeps a large uncertainty, because of the fact that it is a shaded area. The small reduction of uncertainty in this region is due to the inferred shape of the field: since it is relatively smooth, the values and variations at the limit of the inferred domain gives information on the behavior in the shaded area.



(a) Prior distribution of the field v

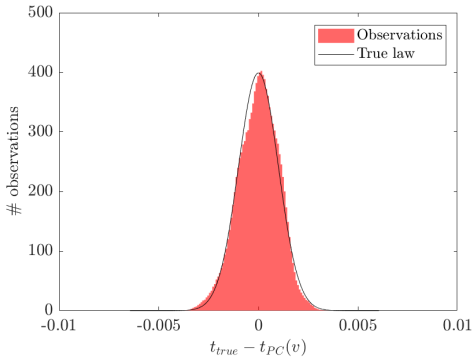
(b) Posterior distribution of the field v

Figure 23: Distribution of the velocity field v.

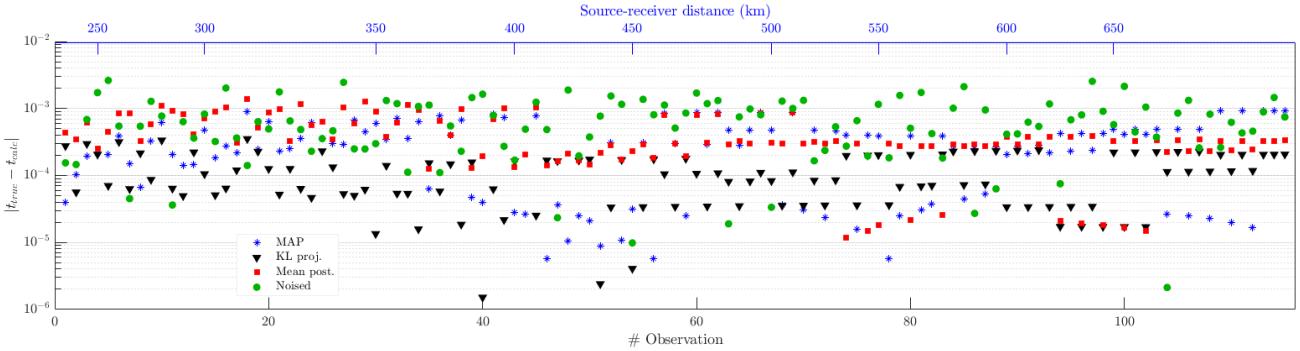
The last analysis focuses on times of arrival, which are in reality the only available data if the true field is unknown. First, in Figure 24a is plotted the distribution of the errors $t_i^{\text{obs}} - \widetilde{t}_i(Y^{(n)})$ compared to a centered Gaussian distribution with a standard deviation equal to 0.001, *i.e.* the theoretical noise distribution. Both are similar, which indicates a good inference of the noise as Figure 22d. This point is discussed more precisely in Section 3.4.3. The error on exact times of arrival (computed with the eikonal solver) $|t_i^{\text{true}} - t_i^{\text{calc}}(v)|$ is compared for three different fields to the error $|t_i^{\text{true}} - t_i^{\text{obs}}|$ in Figure 24b. The three fields are

- the K–L projection of the true field on the reference basis,
- the *maximum* a posteriori (MAP) field,
- the mean field considering the posterior distribution plotted in Figure 23b.

The two inferred fields leads to higher errors than the K–L projection, however, these errors are in general lower than the noise added in the observations. This allows to conclude that it is not possible to be more accurate with the data used to lead the inference.



(a) Histogram of errors $t_i^{\text{obs}} - \widetilde{t}_i(Y^{(n)})$



(b) Comparison of errors $|t_i^{\text{true}} - t_i^{\text{calc}}(v)|$. Blue: MAP; black: K–L projection; red: mean posterior; green: noised observations.

Figure 24: Results concerning times of arrival.

3.4.2 Comparison with parametrized basis

The objective of this section is to compare the results obtained with the reference basis with results obtained with different parametrized basis. Three cases are studied:

- $\mathcal{B}(\ell = 10)$: the assumed length of correlation is too small to explain accurately the velocity field shape;
- $\mathcal{B}(\ell = 80)$: the assumed length of correlation is too large to explain accurately the velocity field shape;
- $\mathcal{B}(\ell = \ell_{\text{MAP}})$: the assumed length of correlation is close to the MAP value, supposed to explain accurately the velocity field shape;

The first step is to assess that the optimal length of correlation ℓ_{opt} is close to the inferred ℓ_{MAP} with $r = 20$. The optimal length of correlation is defined as the ℓ that minimizes the projection error:

$$\ell_{\text{opt}} = \text{argmin}_{\ell \in [\ell_-, \ell_+]} \left\| v_{\text{true}} - v_{\text{KL}}(\ell) \right\|_{L^2(\Omega)}.$$

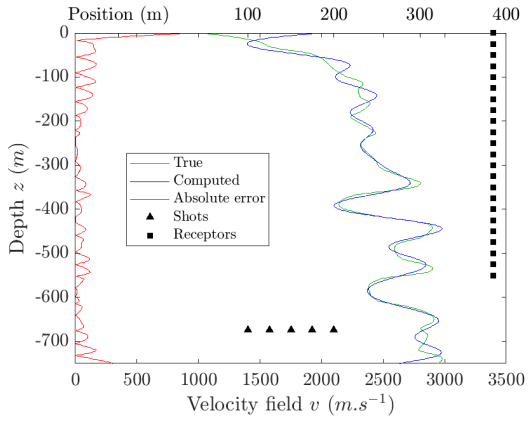
Solving this optimization problem, ℓ_{opt} is found to be equal to 49, which is not really in adequation with the distribution and the MAP value found by the MCMC inference (see Figure 22c). The difference between both is explained by the fact that the data do not give information about all the spatial domain Ω. The inferred length of correlation is therefore the length of correlation that allows to have the most accurate representation of the inferred part of the field Ω^{obs}:

$$\ell_{\text{opt}} = \text{argmin}_{\ell \in [\ell_-, \ell_+]} \left\| v_{\text{true}} - v_{\text{KL}}(\ell) \right\|_{L^2(\Omega^{\text{obs}})}.$$

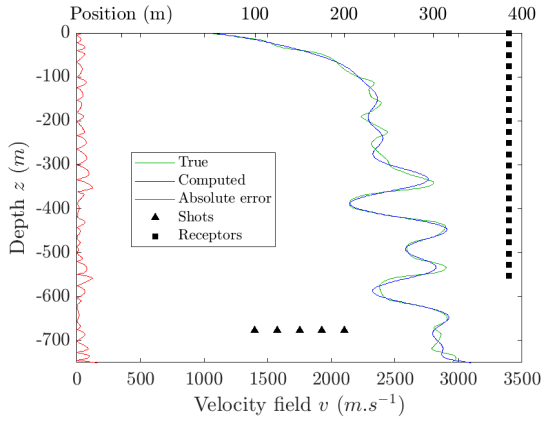
Considering that $\Omega^{\text{obs}} = [1, 550]$, *i.e.* the domain on which information is available and accurate stops at the depth of the last receptor, the optimal value of the length of correlation becomes 36, which is consistent with the inferred value.

Before inferring distributions with the parametrized MCMC algorithm, the best projection fields are plotted in Figure 25. The three results are rather good approximation, however, as explained in 3.2, the two extremal choice include defaults. The smallest length of correlation cannot infer the smooth behavior on the top of the domain. As for the largest length of correlation, its projection coordinates are really unlikely (up to 350 for the last modes), this make the inference really difficult without *a priori*. The reference basis does not bring out these hurdles.

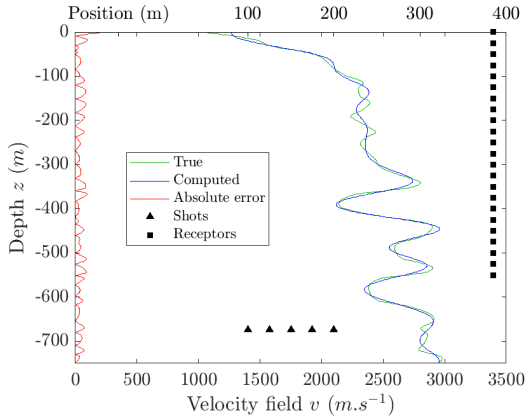
The three field posterior distributions are represented in Figure 26. For the largest length of correlation (Figure 26b), it is clear that the optimal coordinates are not achieved, even with an enlargement of the search space. Moreover, the general problem is the same for the three tests: fixing the shape parameters reduces the search space a lot and prevents from exploring enough possibility to keep the true values probable. Therefore, even after inferring a first time the field with reference basis, re-inferring using the MAP parametrized basis is not a way to improve modeling. The use of a restrained reference basis using the mean over the posterior distribution is investigated too. The best projection is compared to the projection on the initial reference basis. The K–L coordinates are likely in both cases. This shows that a more accurate basis does not lead to a more accurate projection: a second use of the MCMC algorithm is not pertinent to improve the model.



(a) $\mathcal{B}(\ell = 10)$

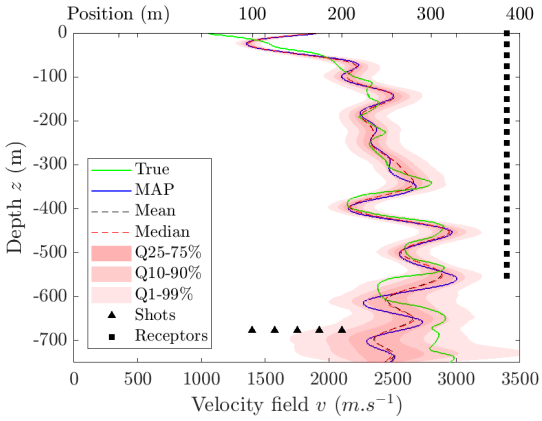


(b) $\mathcal{B}(\ell = 80)$

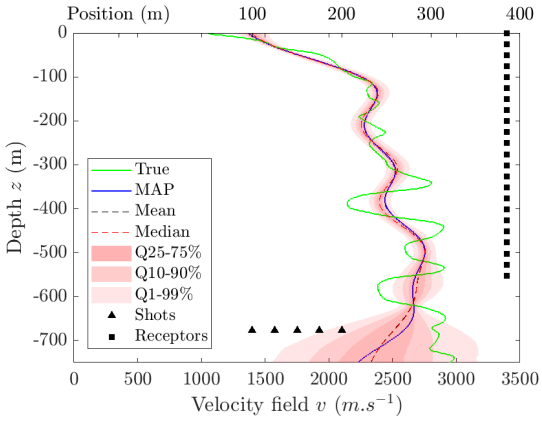


(c) $\mathcal{B}(\ell_{\text{MAP}} = 34)$

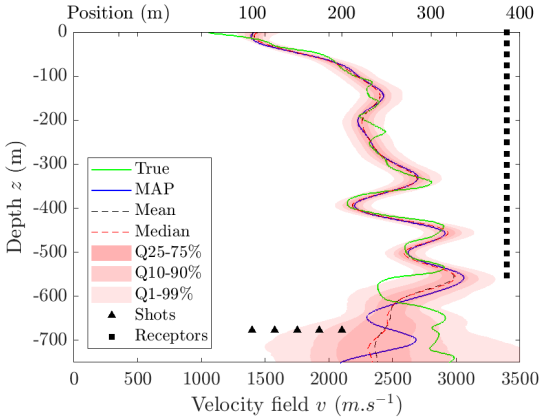
Figure 25: K–L projections of the velocity field v_{true} on parametrized bases.



(a) $\mathcal{B}(\ell = 10)$



(b) $\mathcal{B}(\ell = 80)$



(c) $\mathcal{B}(\ell_{\text{MAP}} = 34)$

Figure 26: Posterior distributions of the velocity field v for parametrized bases.

3.4.3 Error sources quantification

The inference of observations noise, whose results are presented in Figures 22d,24a, seems to be very accurate. The model however does not take into account the other error sources. In this section, a quantification of each source of error is performed. The model includes five main type of errors:

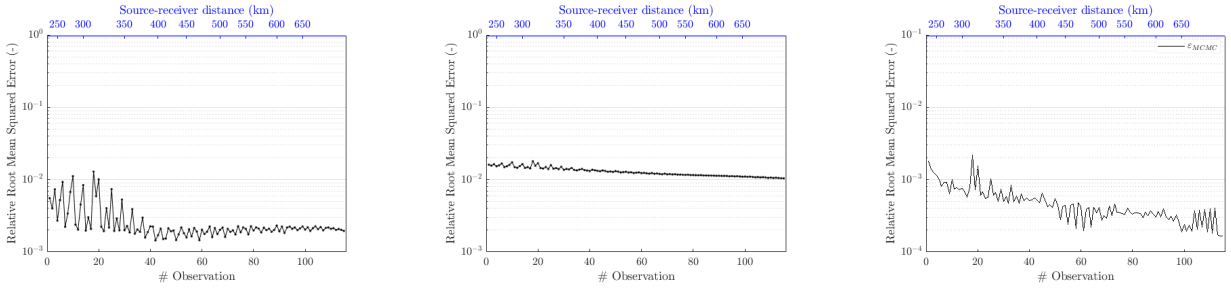
1. the observations noise ε_{obs}, which is known in this application: its law is a Gaussian centered law of standard deviation $\alpha = 0.001$s,
2. the error of model (e.g. the use of the eikonal solver) ε_{eik}, which is set to 0 in this application, since the eikonal model is used to obtain the observations $(t_i^{\text{true}})_{1 \leqslant i \leqslant N_{\text{obs}}}$.
3. the K–L truncation error ε_{KL} that is due to the representation error of the field,
4. the error induced by PC eikonal surrogate ε_{PC}, that accounts for the fact that only the PC surrogate values are known during the MCMC algorithm,
5. the uncertainty error $\varepsilon_{\text{MCMC}}$, which results from the MCMC inference.

The true times of arrival can therefore be decomposed as follows

$$t^{\text{true}} = t^{\text{obs}} + \varepsilon_{\text{obs}} = t(v_{\text{true}}) + \varepsilon_{\text{obs}} + \underbrace{\varepsilon_{\text{eik}}}_{=0} = \widetilde{t}(v_{\text{MC, KL}}) + \varepsilon_{\text{obs}} + \underbrace{\varepsilon_{\text{KL}} + \varepsilon_{\text{PC}} + \varepsilon_{\text{MCMC}}}_{:=\varepsilon_M},$$

where the different errors can compensate each other. The modeling error ε_M accounts for the final modeling error. It is quantified thanks to a set of fields $\mathcal{V}$ obtained by generation of Gaussian processes with different lengths of correlation $\ell \in [10, 150]$.

Regarding ε_{KL}, it is evaluated by computing with the eikonal solver the times of arrival for each velocity field $v_{\ell,i} \in \mathcal{V}$ and comparing it with the times obtained for their corresponding truncated field on the reference basis $v_{\text{KL},\ell,i}$. The RRMSE is plotted in Figure 27a.
The error induced by the use of PC surrogates is analyzed by comparing for each truncated velocity field $v_{\text{KL},\ell,i}$ the times of arrival computed with the eikonal solver $t(v)$ and the ones computed with the surrogates $\widetilde{t}(v)$. Using the truncated velocity fields is a way to discard the truncation error. The RRMSE is plotted in Figure 27b.
Finally, the error due to the MCMC inference $\varepsilon_{\text{MCMC}}$ is studied by comparing the times of arrival in the posterior distribution to the time of arrival of the fields we seek to infer. In order to get rid of the other error sources, the fields to infer are the truncated fields, and the observations are obtained by computing the times of arrival with the PC surrogate and without adding noise. The RRMSE is plotted in Figure 27c. Comparing the different obtained RRMSEs to the one due to the noise in observation as done in Figure 28, the dominant source of error seems to be the PC surrogate. However, results approximate very well the value of α even if it infers in reality the global modeling error too. In order to understand this, the comparison



(a) RRMSE $t(v_{\ell,i})$ versus $t(v_{\mathrm{KL},\ell,i})$: $\varepsilon_{\mathrm{KL}}$ (b) RRMSE $t(v_{\mathrm{KL},\ell,i})$ versus $\widetilde{t}(v_{\mathrm{KL},\ell,i})$: $\varepsilon_{\mathrm{PC}}$ (c) RRMSE $\widetilde{t}(v_{\mathrm{KL},\ell,i})$ versus $\widetilde{t}(Y)$: $\varepsilon_{\mathrm{MCMC}}$

Figure 27: RRMSE for the different error sources.

of times of arrival is also plotted for the application case in Figure 29. In this application, the dominant source of modeling error is also the PC surrogate, as shown in Figure 29a, but its order of magnitude is equivalent to the observation noise one. This is due to the fact that the LHS validation set in the general case samples uniformly the search space, while the PC training points, obtained by the building of a SG, brings less information about the area at the domain limit. The search space explored during the MCMC algorithm is not uniform on all the possible space, and remains in region where the PC approximation is accurate. Gathering all the sources of errors, the final modeling error ε_M has yet the same order of magnitude as the observation noise, as plotted in Figure 29b. The noise is represented by a centered Gaussian variable of standard deviation α, so the associated mean absolute error is computed as $\alpha\sqrt{\frac{2}{\pi}}$. Reducing the noise of observation would lead to have another dominant error that will be the one inferred with α. In real applications, the objective is to have a noise of observation that is greater than the other uncertainties due to the modeling. According to Figure 28, it could therefore be useful to improve the PC surrogate.

4 Discussions and perspectives

4.1 Discussions

4.1.1 Increase of the parameters number

Since the objective is to infer the velocity field with the less *a priori* as possible, more general shapes of fields are investigated.

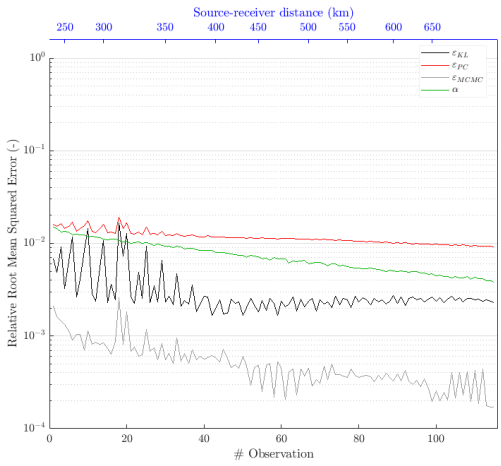
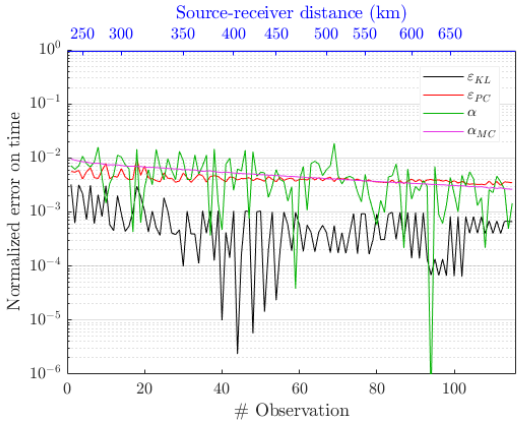
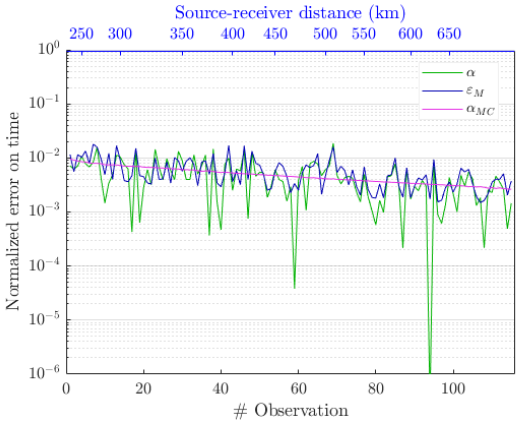


Figure 28: Comparison of errors in the general case.



(a) Comparison of error on times for the different error sources inferring v_{true}



(b) Comparison of global error on times inferring v_{true}

Figure 29: Comparison of errors in the application case.

Increasing the hyperparameters number: use of the Matérn's kernel. The Matérn's kernel is a general type of covariance kernel that includes the exponential one and admits the squared exponential one as its limit [9, 22]. It writes

$$\forall(\boldsymbol{x},\boldsymbol{x}')\in\Omega^2,\qquad k(\boldsymbol{x},\boldsymbol{x}',q)=A\frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{\ell}\right)^{\nu}K_\nu\left(\sqrt{2\nu}\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{\ell}\right),$$

where $\nu>0$ (and generally <100) is a supplementary parameter to infer, Γ is the Gamma function and K_ν is the modified Bessel function of the second kind. Using this kernel increases thus the dimension of the set q by 1. The interest of the change of measure method is the fact

that this increase in dimension does not pass on the computation of the eikonal surrogate and therefore does not make the cost of the inference grow a lot. As shown in Figure 30, the squared exponential reference kernel used in the study of Section 3.4 can be very well approximated by a parametrized Matérn's kernel with length of correlation $\ell = 92.4$ and supplementary parameter $\nu = 0.89$. This is a supplementary argument to explore the Matérn's kernel as a possible covariance kernel for velocity fields. The influence of both parameters ℓ and ν is

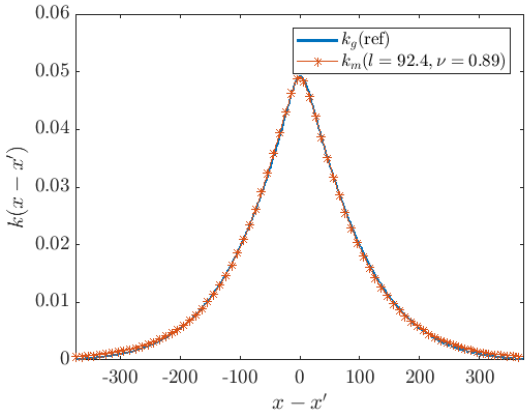
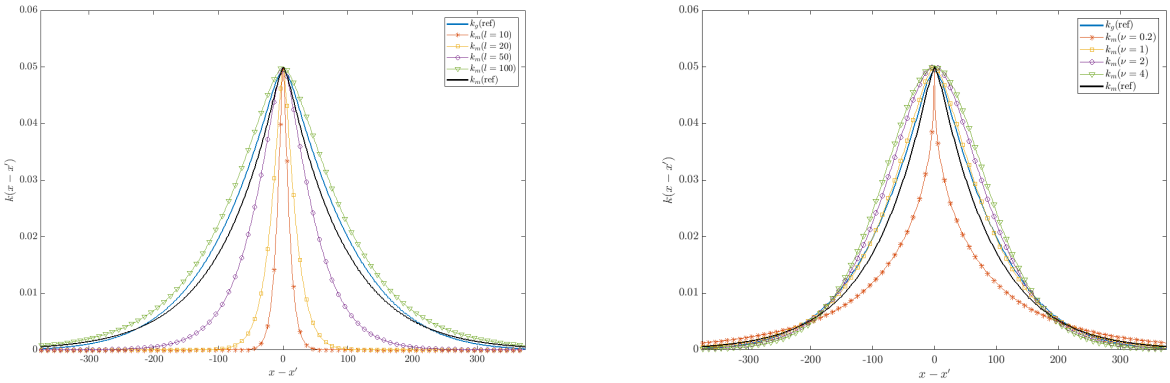


Figure 30: Comparison between the reference squared exponential kernel $k_g(\text{ref})$ and the parametrized Matérn's one $k_m(\ell = 92.4, \nu = 0.89)$.

highlighted in Figure 31. While the length of correlation controls the scope of the kernel, the supplementary parameter rather controls the shape of it: the larger the parameter, the more the kernel ressembles a Gaussian curve.

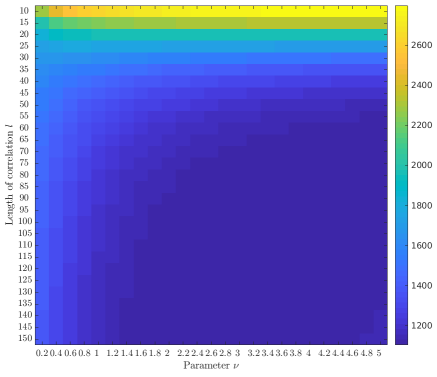


(a) Influence of ℓ on the kernel shape. ν is set to 1.

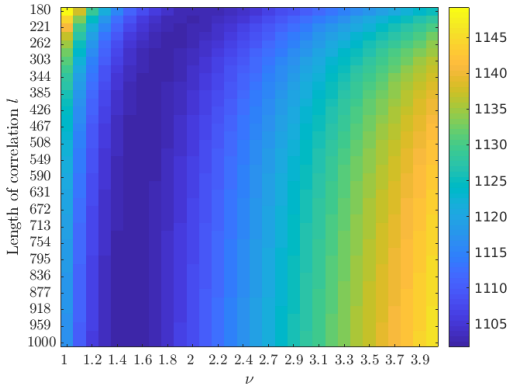
(b) Influence of ν on the kernel shape. ℓ is set to 90

Figure 31: Parameters influence on the Matérn's kernel shape.

Projections of the velocity field on different parametrized basis are computed and compared to the projection on the reference squared exponential basis in Figure 32. The optimal value for ν seems to be between 1.5 and 2 whatever the length of correlation. As for this latter, the error decreases with ℓ. However, it converges exponentially so there is no interest to use too large lengths of correlation. The comparison with the projection error for the squared



(a) Projection error for reference subspace.

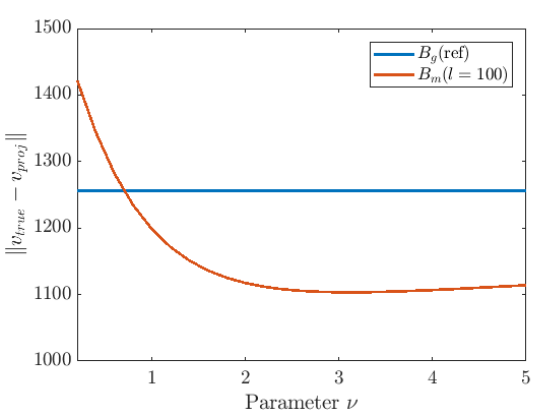


(b) Projection error for larger ℓ and reduced space for ν.

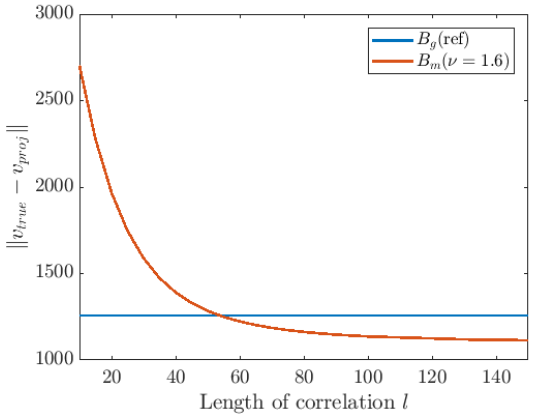
Figure 32: Search for best Matérn's kernel parameters in the sense that they minimize the projection error.

exponential reference basis in Figure 33a, 33b assesses the pertinence of the Matérn's kernel use. The reference Matérn's kernel is built considering $\ell \in [10, 150]$ as the Gaussian's one, and $\nu \in [0.2, 5]$. The shape of it is shown in Figure 31 (black curve). The resulting projection is computed and given in Figure 34a. The error is of the same order of magnitude as for the reference squared exponential basis. However, this reference kernel is actually very similar to a parametrized one as shown in Figure 34b. This reduces a lot the interest of using a reference basis, since it does not bring any supplementary information than a parametrized one.

As a conclusion, the parametrized Matérn's projection may best approximate the field than a squared exponential one. Using the change of measure in its case results in a similar parametrized shape. Concerning the utility of this parametrized shape, it is necessary to compare it to the squared exponential reference one on a set of velocity fields. The best basis to use is the one that gives in average the best projections (best according to the L^2-norm), and that limits the maximal projection error (best according to the L^∞-norm). However, the use of the reference Matérn's kernel, even if it is similar to a parametrized one, presents an advantage over it. It allows exploration on hyperparameters space, which results in limiting

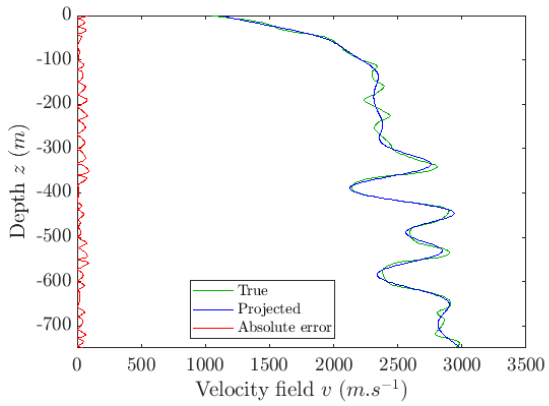


(a) Projection error for $\ell = 100$.

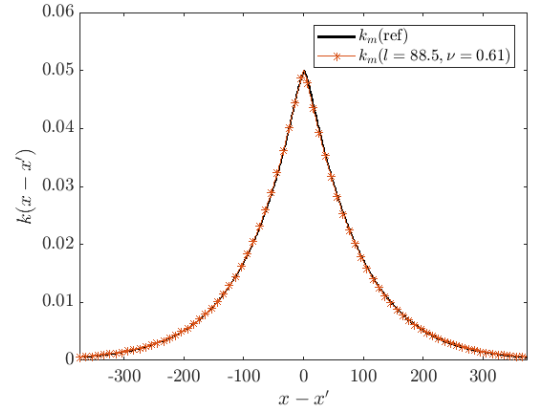


(b) Projection error for $\nu = 1.6$.

Figure 33: Projection errors on Matérn's parametrized bases. Comparison with the squared exponential reference basis.
Norm used is $\|v_{\text{true}} - v_{\text{KL}}(r = 20, \mathcal{B})\|_{L^2}$. Compared only for the projection of the continuous 1-D field introduced in Section 3.1.



(a) Comparison of the true velocity field and its projection on $\mathcal{B}_m$(ref).

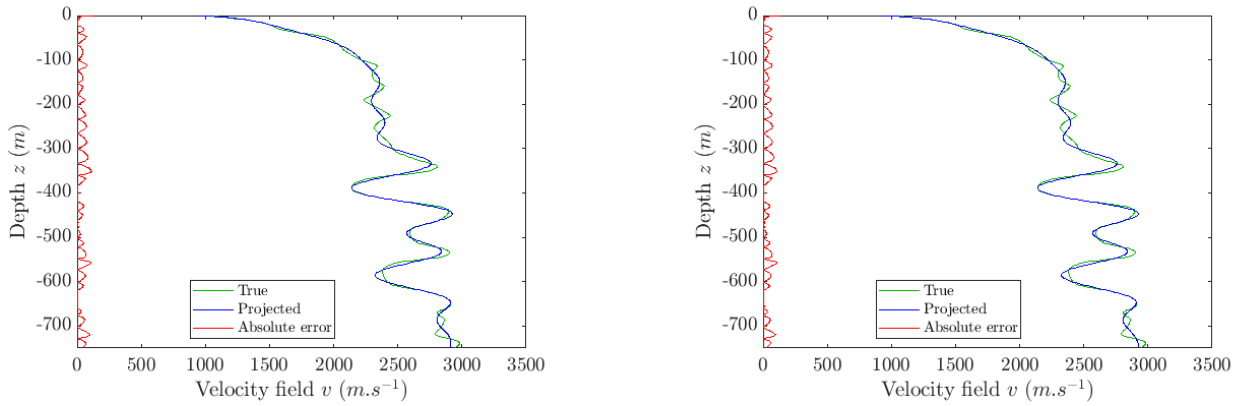


(b) Comparison between reference Matérn's kernel and parametrized one.

Figure 34: Results for the reference Matérn's kernel.

the overfitting, as seen in Section 3.4.2.

Introducing a spatial dependency for the mean. Another way to precise the model is to not consider anymore a mean μ that is constant along the depth z. This constant mean is



(a) Projection on parametrized basis. $\ell = 29$

(b) Projection on reference basis.

Figure 35: Preliminary results using non constant mean $\mu = A + bZ$.

replaced by the following formulation, where $(a, b) \in \mathbb{R}_+ \times \mathbb{R}$:

$$v(z) = az^b \times \exp\left(\sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(z)\xi_i\right).$$

This gives the following form for the logarithmic quantity u

$$u(z) = \underbrace{\log(a)}_{:=A} + b\underbrace{\log(z)}_{:=Z} + \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(z)\xi_i.$$

This model generalizes the previous one, since taking $b = 0$ is sufficient to retrieve $\mu = A$. Contrary to the use of supplementary parameter in the covariance function as in the building of Matérn's kernel, this adds a dimension to the inference of the PC surrogate.

Finding the optimal parameters A and b is easy when the field is known since it consists in a linear regression. The optimal length of correlation for the parametrized basis is found to be 29. Both the projection on the best parametrized basis and on the reference basis are plotted in Figure 35. In order to keep the same number of variables, the number of K–L terms is set to $r = 19$ instead of 20. This allows as the Matérn's kernel to better approximate the top of the domain. The trade-off between this gain and the loss of computational time due to the increasing of the dimension has to be determined.

4.1.2 Robustness of the method to the data

The application in Section 3 is an ideal case where a lot of accurate data are known on the inferred domain. It could be interesting to test the robustness of the method to different noises

and to a lack of data.

Robustness relatively to the noise. Changing the magnitude of the noise level α, and even changing the observed times of arrival without changing the noise law is a way to study the results dependency to this parameter. The results dependency to the noise will be quantified through the measure of adequation between the posterior distributions for the velocity fields and the times of arrival. The Kolmogorov–Smirnov test [19] could be used in order to verify if the results are robust (and therefore independent) to a change of noise with keeping the same law for it. To quantify the influence of a change of this law, the Kullback–Leibler distance [1] between posterior distributions could be considered as a first step.

Robustness relatively to the data. The results of Section 3.4 will be put in perspective with the same inference led with only 50% of the receptors stations. The times of arrival of the posterior possible fields will be compared to the true data and the distribution of them according to the added noise. As for the noise, the Kullback–Leibler distance will be a first step to quantify the resulting change in times of arrival distributions.

4.2 Perspectives

This six-months internship is a preliminary work for a PhD thesis. In the following, some current hurdles and points of improvement are highlighted.

Convergence of the MCMC algorithm. The update of the covariance allows the MCMC algorithm to converge toward a local minimum. Indeed, without update, it is possible to find a state which has a good posterior and never moves to a very near state that is better.
In order to ensure convergence toward the global minimum, another strategy must be put in place, like the running of several chains in parallel, or the introduction of a criterion to scale the proposition matrix balancing the exploration/exploitation criterion.
It can be difficult, when the true values are not known, to qualify and quantify the superiority of a result versus another. The scatter plot of the values through the iterations is an indicator of the exploration of the space: if the values are converging too fast, maybe there was no convergence toward a local minimum.

Two-dimensional case. The approach developed in this report is applicable on two dimensional fields. The difficulty in that case is the needed number of K–L terms to have an accurate

approximation of the field. This brings out the interest to use supplementary reduction methods, that implies for instance to better model the error sources and to use adaptive models, as presented in the following paragraphs.

Error models. As discussed in Section 3.4.3, the quantification of the different sources of errors is a crucial point to assess the pertinence of a solution and to know how to improve the method. The main error sources are

- the observation noise ε_{obs},
- the error of model (e.g. the use of the eikonal solver) ε_{eik}, which is generally considered as negligible toward the others,
- the truncation K–L error ε_{KL},
- the surrogate error ε_{PC} due to the PC approximation,
- the uncertainty error $\varepsilon_{\text{MCMC}}$, which results from the MCMC inference.

The true times of arrival are decomposed as follows

$$t^{\text{true}} = t^{\text{obs}} + \varepsilon_{\text{obs}} = t(v_{\text{true}}) + \varepsilon_{\text{obs}} + \varepsilon_{\text{eik}} = \widetilde{t}(v_{\text{MC, KL}}) + \varepsilon_{\text{obs}} + \varepsilon_{\text{eik}} + \varepsilon_{\text{KL}} + \varepsilon_{\text{PC}} + \varepsilon_{\text{MCMC}}.$$

The PC error is itself a sum of two different errors: the truncation error $\varepsilon_{\text{PC, lev}}$ due to the choice of the level for the polynomial projection basis, and the error on coefficients $\varepsilon_{\text{PC, coef}}$, which comes from the sampling method and the corresponding introduced bias in the coefficients computation. A trade-off between the computational cost and the accuracy of the model remains to be theoretically defined. Another way to limit the error is to consider the variability of all the modes in the K–L truncation, even if only the r-th first are inferred. That is to say, the velocity field v^r_{KL}, instead of being the K–L projection truncated to r terms, is a mean of the behaviour on the highest modes:

$$v^r_{\text{KL}} = \int_{\Xi=\{\mathcal{X}=\{\xi_k,\ k>r\}\}} \exp\left(\mu + \sum_{i=1}^{\infty} \lambda_i^{1/2} u_i \xi_i\right) \pi_{\boldsymbol{\xi}}(X) dX.$$

Adaptive models. Many parameters are set by default in the current analysis. In particular, for the eikonal model, a step of discretization is chosen, and for the different surrogates, a set of training points and a maximum level for polynomial basis is set. These choices are made according to a trade-off between the computational cost and the accuracy. In reality, a surrogate model does not need to be accurate uniformly on its defined space, but only on a subspace of interest. In order to take this into account, it is interesting to investigate adaptive models and multi-fidelity ones.

The adjustment of a model is for instance the addition of training points on the subspace of interest. As seen in Section 3.4.3, the eikonal surrogate is less accurate on the bounds of the search space. In reality, during the inference, these bounds are not achieved and the approximation is better than the average. If, as in the application, the inferred error is of the same order of magnitude than the surrogate's error, one could add training points to improve the model. These training points would be chosen according to the posterior distribution after the inference. This would allow to improve the subspace of interest by infill sampling. This approach is to put in perspective with the bias it brings out: being more accurate in a subspace could lead the optimization to stay in that subspace without exploring anymore. The use of more types of data in an adaptive framework is considered too. First the general shape of the field is inferred thanks to times of arrival, then polarization is used to discriminate the posterior fields, and finally a local optimization is performed thanks to a full-wave form model. This method is close to a multi-fidelity one.

Multi-fidelity methods consists in defining several models that have different accuracies and computational costs. For instance, an eikonal solver with a fine-mesh discretization t_H, accurate but expensive to compute would be the high-fidelity model, while a coarse-mesh solver t_L would be the low-fidelity model. Combining both knowledge with the computation of some precise points and much more inaccurate ones allow building and enriching a model at low cost [18].

Source location. The interest of a better knowledge of the velocity field is above all the possibility to improve source location and characterization. Two points are raised:

- what is the impact of the field inference on the source location compared to a known determined field ?
- how does the field uncertainties propagate to the location uncertainty ?

5 Conclusion

The built surrogate models and the implementation of the algorithm solving the corresponding inverse problem meet the different expectations. In the studied case, the velocity field is indeed well-inferred, and the associated uncertainties are in adequation with the geophysical observations. Moreover, the use of several surrogate models allows a drastic improvement of the computation cost, while limiting the induced error. The numerical implementation with `Matlab` will be converted in `Python` for the following studies, in order to be more adaptable and user-friendly. The factorization of the code already eases the test of several configurations.

From a geophysical point of view, this approach opens new possibilities to qualify and quantify uncertainties linked to seismic tomography, and therefore to improve the risk assessments. As for the mathematical framework, the use of surrogate models and first and foremost the change of measure is innovative in this domain of application.

The major issue of this model is the increase in dimension of the field, that will be possible by using reduction methods and adaptive approaches. The use of this method for the location of the event will be considered too.

References

[1] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological)*, 28(1):131–142, 1966.

[2] N. Chopin, C. Robert, and J. Rousseau. Harold Jeffreys' Theory of Probability revisited. *Statistical Science*, 107(3):141–172, 2009. 53 pages.

[3] P. R. Conrad and Y. M. Marzouk. Adaptive smolyak pseudospectral approximations. *SIAM Journal on Scientific Computing*, 35(6):A2643–A2670, 2013.

[4] P. G. Constantine, M. S. Eldred, and E. T. Phipps. Sparse pseudospectral approximation method. *Computer Methods in Applied Mechanics and Engineering*, 229-232:1–12, 2012.

[5] A. Doucet, A. Smith, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Information Science and Statistics. Springer New York, 2013.

[6] L. Fejér. On the infinite sequences arising in the theories of harmonic analysis, of interpolation, and of mechanical quadratures. *Bulletin of the American Mathematical Society*, 39(8):521 – 534, 1933.

[7] Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO). Disaster warning and promoting human welfare – the civil and scientific uses of ctbto data public information, 2022. Vienna International Centre.

[8] A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient metropolis jumping rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, pages 599–608. Oxford University Press, Oxford, 1996.

[9] M. G. Genton. Classes of kernels for machine learning: A statistics perspective. *J. Mach. Learn. Res.*, 2:299–312, mar 2002.

[10] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Element Method: Response Statistics*, pages 101–119. Springer New York, New York, NY, 1991.

[11] H. Haario, Saksman E., and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223 – 242, 2001.

[12] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 04 1970.

[13] R. L. Iman, J. C. Helton, and J. E. Campbell. An approach to sensitivity analysis of computer models: Part i—introduction, input variable selection and preliminary variable assessment. *Journal of Quality Technology*, 13(3):174–183, 1981.

[14] R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370, 1996.

[15] 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.

[16] M. O'Brien and C. Regone. Model created by Mike O'Brien and Carl Regone, 1994, and is provided courtesy of Amoco and BP.

[17] S. A. Orszag. Comparison of pseudospectral and spectral approximation. *Studies in Applied Mathematics*, 51(3):253–259, 1972.

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

[19] N. Smirnov. Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, 19(2):279 – 281, 1948.

[20] P. Sochala, A. Gesret, and O. Le Maitre. Polynomial surrogates for bayesian traveltime tomography. *GEM - International Journal on Geomathematics*, 12(1), December 2021.

[21] I. Sraj, O. P. Le Maître, O. M. Knio, and I. Hoteit. Coordinate transformation and polynomial chaos for the bayesian inference of a gaussian process with parametrized prior covariance function. *Computer Methods in Applied Mechanics and Engineering*, 298:205–228, 2016.

[22] D. Stoyan. Matérn, b.: Spatial variation. 2nd ed., springer-verlag, berlin, heidelberg, new york, london, paris, tokyo 1986, 151 s., dm 33,—. *Biometrical Journal*, 30(5):594–594, 1988.

[23] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.

[24] Dapeng Zhao, Yingbiao Xu, Douglas A. Wiens, LeRoy Dorman, John Hildebrand, and Spahr Webb. Depth extent of the lau back-arc spreading center and its relation to subduction processes. *Science*, 278(5336):254–257, 1997.

Mathematical symbols

A Amplitude of the Gaussian process, $\in \mathbb{R}_+^\star$. 18

Ω Spatial domain, $\subset \mathbb{R}^2$. 13

Σ Covariance matrix of the K–L coordinates $\boldsymbol{\xi}$, $\Sigma : \mathbb{H} \to \mathbb{R}^{r \times r}$. 18

α Noise level, corresponds to the standard deviation of the observation noise, $\in \mathbb{R}_+^\star$. 20

$\boldsymbol{\eta}$ Set of K–L coordinates $(\eta_i)_{1 \leqslant i \leqslant r}$ associated to a decomposition on a parametrized basis, $\boldsymbol{\eta} \sim \mathcal{N}(0, 1)$. 17

$\boldsymbol{\xi}$ Set of K–L coordinates $(\xi_i)_{1 \leqslant i \leqslant r}$ associated to a decomposition on a reference basis, $\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma(q))$. 18

$\boldsymbol{x}$ Point of the spatial domain, $\in \Omega$. 13

ℓ Length of correlation, $\in \mathbb{R}_+$. 18

$\left(\overline{u}_i, \overline{\lambda}_i\right)$ Eigenelements associated to a reference covariance kernel $\overline{k}(\cdot, \cdot)$. 18

$(u_i(q), \lambda_i(q))_{i \in \mathbb{R}_+^\star}$ Eigenelements associated to a parametrized covariance kernel $k(\cdot, \cdot, q)$. 17

$\mathbb{H}$ Hyperparameters space. 16

μ Mean of the Gaussian process, $\in \mathbb{R}$. 16

$\overline{k}$ Reference covariance kernel of the Gaussian process, $\overline{k} : \Omega \times \Omega \to \mathbb{R}$. 18

$\pi_{\cdot}$ Prior distribution of variable $\cdot$. 18

$\widetilde{\cdot}$ Surrogate of the quantity $\cdot$. 23

k Covariance kernel of the Gaussian process depending on hyperparameters q, $k : \Omega \times \Omega \times \mathbb{R}_+^{|q|} \to \mathbb{R}$. 16

q Set of hyperparameters for the covariance kernel k. In general, $q = \{\ell, A\}$. 16

r Number of K–L terms, $\in \mathbb{N}^\star$. 17

t Time (in s), $\in \mathbb{R}_+$. 13

u Logarithm of the velocity field, $u = \log(v)$. 16

v Velocity field (in $\mathrm{m.s}^{-1}$), $v : \Omega \to \mathbb{R}_+$. 13

z Depth (in m), $\in \mathbb{R}_+$. 51

Mathematical acronyms

i.i.d. *independent and identically distributed*. 17

K–L *Karhunen–Loève*. 16–18, 20, 24, 29, 30, 35, 37, 38, 40–46, 52–54, 63, 67

LHS *Latin hypercube sampling*. 25, 47

MAP *maximum* a posteriori. 42–44

MCMC *Markov Chain Monte–Carlo*. 14, 21, 24, 26, 27, 35, 38, 39, 41, 42, 44, 46, 47, 53, 54, 65, 66

MH *Metropolis–Hastings*. 21, 23

OLS *ordinary least-square*. 25

PC *Polynomial Chaos*. 14, 24–26, 35–37, 46, 47, 52, 54, 62, 63

PSP *pseudo spectral projection*. 24, 25, 62

RRMSE *relative root mean squared error*. 35–38, 46, 47, 65, 66

SA *sensitivity analysis*. 26, 64–66

SG *sparse grid*. 24, 38, 47, 62, 63

SVD *singular value decomposition*. 22, 25

Glossary

eikonal Describing the properties of a class of partial differential equation encountered in descriptions of wave propagation. 13, 15, 26, 29, 37, 38, 42, 46, 49, 54, 55, 63, 65

P waves For pressure or primary waves. One of the two main types of seismic body waves. 27

Acronyms

CEA *Commissariat à l'Énergie Atomique et aux énergies alternatives*. 10, 11, 14

CTBT *Comprehensive Test Ban Treaty*. 12

CTBTO *Comprehensive Nuclear-Test-Ban Treaty Organization*. 11

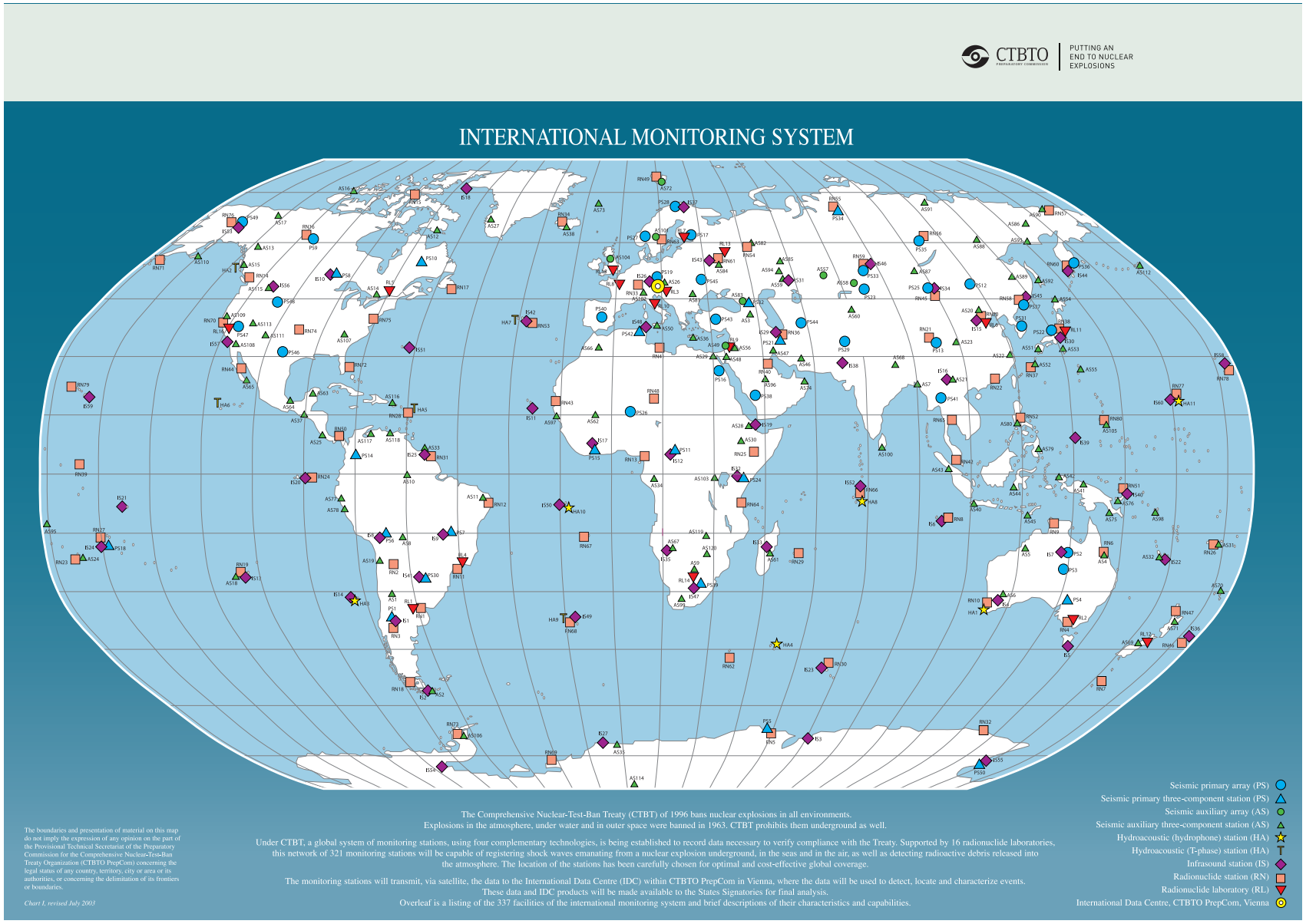
DAM *Direction des Applications Militaires*. 10, 11

IAEA *International Atomic Energy Agency*. 11

IMS *International Monitoring System*. 12, 13

NTP *Non-Proliferation Treaty*. 12

Appendix A International Monitoring System map



Appendix B Smolyak pseudo-spectral projection

In this section, the computation of PC coefficients thanks to a SG and a PSP is explained. This method is based on [3,4]. The objective is to calculate, for any multi-indice κ in the considered multi-indices set $\mathcal{K}$, the corresponding coefficient c_κ, knowing that the PC surrogate model of the function f writes

$$\widetilde{f}(X) = \sum_{\kappa \in \mathcal{K}} c_\kappa \phi_\kappa(X),$$

where $X \sim \mathcal{U}([-1,1]^d)$ and $\{\phi_\kappa\}_{\kappa \in \mathcal{K}}$ is a basis of Legendre's polynomials. By projecting the surrogate on the basis, it holds

$$\forall \kappa \in \mathcal{K}, \qquad c_\kappa = \frac{\left\langle \widetilde{f}(X), \phi_\kappa(X) \right\rangle}{\left\| \phi_\kappa(X) \right\|^2},$$

where $\left\| \phi_\kappa(X) \right\|^2 = 1$ considering an orthonormal basis. The scalar product to compute is a d-dimensional integral, which is exponentially expensive to compute when increasing dimension. In order to decrease this computational cost, a Smolyak tensorization is done. This consists in using a sparse sampling grid and calculating only the coefficients for $k \in \mathcal{K} \subset \mathbb{N}^d$ such that no aliasing is introduced. This sampling grid is deterministic and every level of discretization is nested within each other.

Here, a Fejér grid of type II is used to determine the set of points S_r^l of level l in dimension r [6]. In one dimension, the set is obtained by the univariate quadrature rule and is denoted Q_l. The cardinality of this set according to the level $l \geqslant 0$ is equal to $|Q_l| = n(1,l) = 2^{l+1} - 1$. The maximal degree of the associated monomials is $d_{\max} = 2l+1$. The levels being nested with each other, for any $l > 1$, $n(1,l) - n(1,l-1) = 2^l$. The increase of dimension with the Smolyak tensorization writes

$$S_r^l = \sum_{i=1}^{l} (Q_i - Q_{i-1}) \otimes S_{r-1}^{l-i},$$

setting $Q_{-1} = S_r^{-1} = \emptyset$ by convention. In terms of number of points, it holds

$$n(r,l) = \sum_{i=0}^{l} (n(1,i) - n(1,i-1)) \times n(r-1, l-i) = \sum_{i=0}^{l} 2^i \times n(r-1, l-i).$$

Analyzing this double recurrence in l and r leads to find

$$n(r,l) = 2^{l+1} \sum_{k=0}^{r-1} \begin{pmatrix} l+k \\ l \end{pmatrix} (-1)^{r-1-k} + (-1)^r.$$

For level l equal to 3, as used in the application to compute the PC surrogate of the eikonal solver in Section 3.3, the number of points exponentially increases and becomes too expensive around dimension $r = 30$ as shown in Figure 36. In order to increase the K–L truncation dimension, other sampling methods need to be discussed. For instance, one could use the fact that highest modes are energetically less important and therefore that a lowest density sampling is sufficient in the corresponding dimensions.

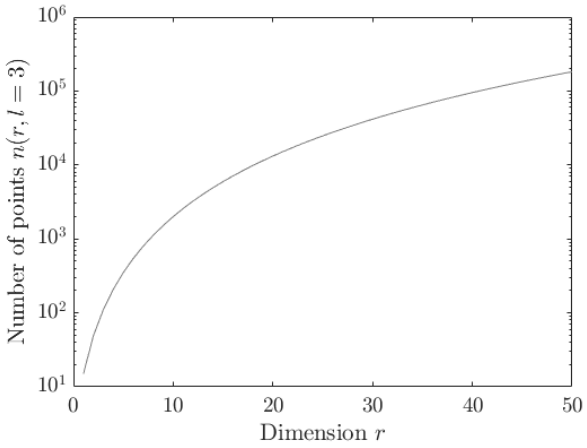


Figure 36: Size of the Smolyak's SG using Fejér's type II quadrature rule according to the dimension for level $l = 3$.

Appendix C Spatial sensitivity analysis

In order to quantify the possibility to infer the velocity for any depth z, a SA is performed. The goal is to see in which measure a perturbation in the velocity field leads to a change in the observations.

C.1. Perturbation of the field

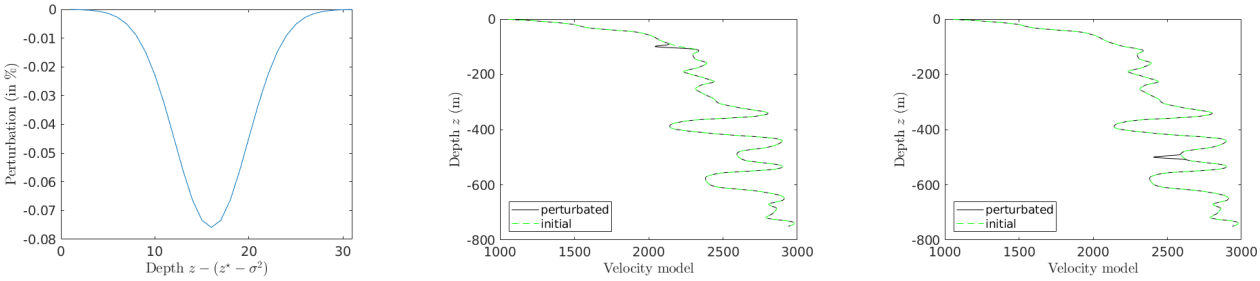
From the true velocity field, a set of perturbed velocity fields is obtained, considering the addition of a random Gaussian noise that is moved on the depth axis. If the perturbation is centered on z_c, the Gaussian noise at this point $\zeta(z_c)$ follows $\mathcal{N}(0, p^2)$ where p is the maximum percentage of the perturbation relatively to the true field. In order to keep a physical field, the perturbation is smoothed thanks to a squared exponential kernel, which has an amplitude σ^2:

$$\zeta(z) := \zeta(z_c) \times \exp\left(-\frac{(z - z_c)^2}{2\sigma^2}\right).$$

This noise is then introduced in the velocity field:

$$v_{\text{perturbed}}(z) := (1 + \zeta(z))v(z).$$

In this analysis, p is chosen equal to 5% and σ^2 to 15. An example of generated noise and its application on the field is shown in Figure 37.



(a) Gaussian noise (b) Perturbed field, $z_c = 100$ (c) Perturbed field, $z_c = 500$

Figure 37: Example of noise generated for the perturbation of the field and associated perturbed fields.

C.2. Results of sensitivity analysis

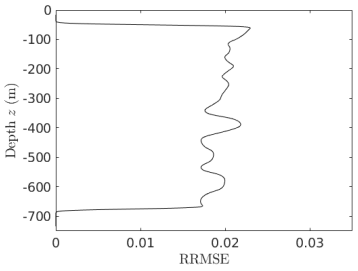
Computation of observations. Ten random Gaussian noises are generated $(\zeta_1, \cdots, \zeta_{10})$. For each of them, the resulting perturbed field is build for any $z_c \in [\sigma^2 + 1, z_{\max} - \sigma^2]$. Then,

the eikonal solver is used to obtain the observations associated to each field: for each noise ζ, for each center of perturbation z_c, for each observation i, $t_{z_c,\zeta,i}$ is obtained.

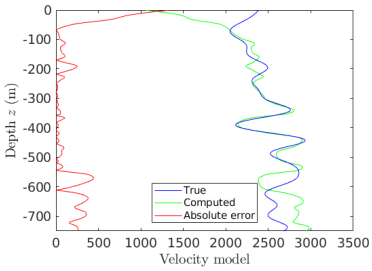
Definition of "influence". A spatial zone is considered as influent if a perturbation on this area leads to a significant difference in the observations. This difference is computed as the RRMSE:

- For each z_c, for each observation i, a RRMSE is computed thanks to the 10 random perturbations:
$$\mathrm{RRMSE}(z_c, i) = \sqrt{\frac{\sum_{1 \leqslant k \leqslant 10} (t_i^{\mathrm{obs}} - t_{z_c,\zeta,i})^2}{K \left(t_i^{\mathrm{obs}}\right)^2}}.$$
- These RRMSEs are then gathered in a global RRMSE, considering all the observations at the same time:
$$\mathrm{RRMSE}_{\mathrm{glob}}(z_c) = \sqrt{\sum_{1 \leqslant i \leqslant N} \mathrm{RRMSE}(z_c, i)^2}.$$

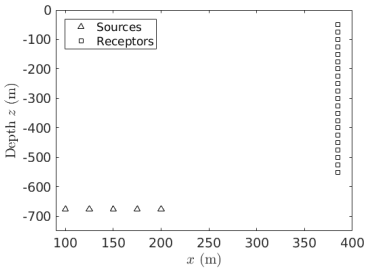
Results. This indicator is plotted in Figure 38a. It is put in perspective with the results obtained by the MCMC inference in Figure 38b. The shaded areas for $z < 50$ and $z > 680$ seem to be explained by this SA: perturbations in these regions do not have any impact on the observations, making the inference impossible. This could be explained by the lack of sources and receptors in these zones (see Figure 38c), this assumption being tested later in Section C.3..



(a) Observations RRMSE according to z_c



(b) Velocity fields. Green: true; blue: obtained with the MCMC algorithm.



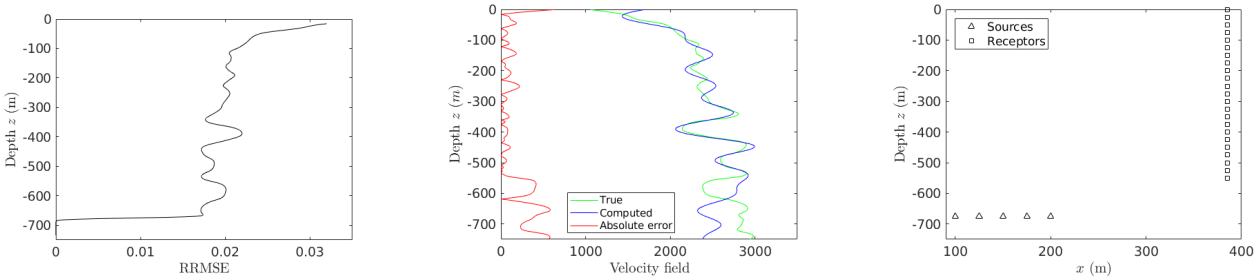
(c) Position of the sources and receptors

Figure 38: SA results: top and bottom area cannot be inferred with the observations.

C.3. Addition of receptors

In order to confirm this analysis, the same procedure is repeated, considering two added receptors at the top of the zone of interest, as represented in Figure 39c. Using the previous MCMC

algorithm of Section 2.4 with these two more receptors, we obtain the results in Figure 39b. It becomes clear that adding receptors in the region reinforces the influence of it and allows to infer the velocity field in it.



(a) Observations RRMSE according to z_c

(b) Velocity fields. Green: true; blue: obtained with the MCMC algorithm.

(c) Position of the sources and receptors

Figure 39: SA results with two supplementary receptors: top area is now inferred too, in contrary to bottom one.

Finally, in order to distinguish two different border regions, the following results will use the two added receptors. It allows to separate the case of the top region, where we have information, from the case of the bottom region, where there is still no receptor nor source. It is worth to say that in a real case, knowing the influence of each area on the observations is a good way to quantify the possibility to infer this zone, and thus to reduce the dimensionality of the problem.

Appendix D Link between projection and amplitude

The amplitude of the kernel A does not influence directly on the projection. Indeed, if the length of correlation ℓ is fixed, it holds

$$k(\boldsymbol{x}, \boldsymbol{x}', A = a, \ell) = a \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{2\ell^2}\right) = ak(\boldsymbol{x}, \boldsymbol{x}', A = 1, \ell),$$

and then, for the eigenelements,

$$\begin{aligned}
&\int_\Omega k(\boldsymbol{x}, \boldsymbol{x}', A = 1, \ell) u_i(\boldsymbol{x}', A = 1, \ell) d\boldsymbol{x}' = \lambda_i(A = 1, \ell) u_i(\boldsymbol{x}, A = 1, \ell), \\
=&\int_\Omega \frac{1}{a} k(\boldsymbol{x}, \boldsymbol{x}', A = a, \ell) u_i(\boldsymbol{x}', A = 1, \ell) d\boldsymbol{x}' = \lambda_i(A = 1, \ell) u_i(\boldsymbol{x}, A = 1, \ell), \\
\iff &\int_\Omega k(\boldsymbol{x}, \boldsymbol{x}', A = a, \ell) u_i(\boldsymbol{x}', A = 1, \ell) d\boldsymbol{x}' = a\lambda_i(A = 1, \ell) u_i(\boldsymbol{x}, A = 1, \ell),
\end{aligned}$$

i.e. $(\lambda(A = a, \ell), u(\boldsymbol{x}, A = a, \ell)) = (a\lambda(A = 1, \ell), u(\boldsymbol{x}, A = 1, \ell))$. This scaling directly reverberates on the K–L coordinates computation:

$$\eta_i(A = a, \ell) = \frac{u_{\text{true}}^T u_i(A = a, \ell)}{\lambda_i^{1/2}(A = a, \ell)} = \frac{u_{\text{true}}^T u_i(A = 1, \ell)}{\sqrt{a\lambda_i(A = 1, \ell)}} = \frac{\eta_i(A = 1, \ell)}{\sqrt{a}}.$$

This leads to the exact same projection:

$$\begin{aligned}
u_{\text{KL}}(A = 1, \ell) &= \mu + \sum_{i=1}^{r} \lambda_i^{1/2}(A = 1, \ell) u_i(A = 1, \ell) \eta_i(A = 1, \ell), \\
&= \mu + \sum_{i=1}^{r} \sqrt{a\lambda_i(A = 1, \ell)} u_i(A = 1, \ell) \frac{\eta_i(A = 1, \ell)}{\sqrt{a}}, \\
&= \mu + \sum_{i=1}^{r} \lambda_i^{1/2}(A = a, \ell) u_i(A = 1, \ell) \eta_i(A = a, \ell) = u_{\text{KL}}(A = a, \ell).
\end{aligned}$$

To conclude, the amplitude just scales the projection coordinates.