Geophysical Journal International

Geophys. J. Int. (2024) **237,** 31–48
Advance Access publication 2024 January 18
GJI General Geophysical Methods

https://doi.org/10.1093/gji/ggae026

Bayesian tomography using polynomial chaos expansion and deep generative networks

Giovanni Angelo Meles [1], Macarena Amaya [1], Shiran Levy [1], Stefano Marelli[2] and Niklas Linde[1]

[1] *Institute of Earth Sciences, Department of Applied and Environmental Geophysics, University of Lausanne,* 1015 *Lausanne, Switzerland.*
E-mail: Giovanni.Meles@unil.ch
[2] *Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Institute of Structural Engineering, 8093 Zurich, Switzerland*

Accepted 2024 January 17. Received 2023 December 19; in original form 2023 October 10

SUMMARY

Implementations of Markov chain Monte Carlo (MCMC) methods need to confront two fundamental challenges: accurate representation of prior information and efficient evaluation of likelihood functions. The definition and sampling of the prior distribution can often be facilitated by standard dimensionality-reduction techniques such as Principal Component Analysis (PCA). Additionally, PCA-based decompositions can enable the implementation of accurate surrogate models, for instance, based on polynomial chaos expansion (PCE). However, intricate geological priors with sharp contrasts may demand advanced dimensionality-reduction techniques, such as deep generative models (DGMs). Although suitable for prior sampling, these DGMs pose challenges for surrogate modelling. In this contribution, we present a MCMC strategy that combines the high reconstruction performance of a DGM in the form of a variational autoencoder with the accuracy of PCA–PCE surrogate modelling. Additionally, we introduce a physics-informed PCA decomposition to improve accuracy and reduce the computational burden associated with surrogate modelling. Our methodology is exemplified in the context of Bayesian ground-penetrating radar traveltime tomography using channelized subsurface structures, providing accurate reconstructions and significant speed-ups, particularly when the computation of the full-physics forward model is costly.

Key words: Ground penetrating radar; Machine Learning; Numerical modelling; Probability distributions; Tomography.

1 INTRODUCTION

Bayesian inversion methods can account for data and modelling uncertainties as well as prior knowledge, thus, representing an attractive approach for tomography and its uncertainty quantification. Nevertheless, the difficulties in specifying appropriate prior distributions and the high computational burden associated with repeated forward model evaluations often hinder proper implementations of Bayesian tomography (Chipman *et al.* 2001). In geophysical settings, prior distributions have traditionally been specified by assuming the subsurface to be represented by a Gaussian random field. More advanced options are made possible by relying on the information content of training images (TI), that is, large gridded 2-D or 3-D unconditional representations of the expected target spatial field that can be either continuous or categorical (Mariethoz & Caers 2014; Laloy *et al.* 2017, 2018). Bayesian inversion needs not only a parametrization of the prior that accurately represents the prior information, but also one that is easy to manipulate and one with a an underlying distribution for which small perturbations lead to comparatively small changes in the data response. In many practical implementations, it is advantageous to rely on another parametrization than the one used for physics-based modelling and visualization, that is, typically a pixel-based parametrization (Asher *et al.* 2015; Alizadeh *et al.* 2020). Physical media are typically associated with points in $\mathbb{R}^N$, where N is the number of elements in the corresponding pixel-based representation. While allowing easy implementation of forward modelling schemes [e.g. Finite Difference (FD) based on partial differential equations], pixel-based N-dimensional parametrizations are often not suitable to effectively parametrize the prior distribution, as N can be very large. When prior knowledge suggests constrained spatial patterns, such as covariance or connected spatial structures, the prior-compatible models populate manifolds embedded in $\mathbb{R}^N$. If this manifold can be locally assimilated to a subset of $\mathbb{R}^M$, with $M \ll N$, the actual prior distribution reduces to a function of M variables only, which effectively implies an M-dimensional forward/inverse problems.

Various approaches can be used to achieve manifold identification through dimensionality reduction. Among these techniques,

principal component analysis (PCA) and related methods are the most common *linear* dimensionality reduction methods (Boutsidis *et al.* 2008; Jolliffe & Cadima 2016). Based on a set of prior model realizations and the eigenvalues/eigenvectors of the corresponding covariance matrix, PCA provides optimal M-dimensional representations in terms of uncorrelated variables that retain as much of the sample variance as possible. PCA has found widespread application in geophysics, using both deterministic and stochastic inversion algorithms, with recent advancements offering the potential to reconstruct even discontinuous structures (Reynolds *et al.* 1996; Giannakis *et al.* 2021; Thibaut *et al.* 2021; Meles *et al.* 2022). In the context of complex geological media with discrete interfaces (Strebelle 2002; Zahner *et al.* 2016; Laloy *et al.* 2018) investigated in this study, PCA-based methods are ineffective in encoding the prior knowledge. Promising alternatives are offered by deep generative models (DGM) that learn the underlying input distribution and generate synthetic samples that closely resemble the statistics in a provided data set based on TIs. For instance, Variational Autoencoders (VAEs) encode data patterns into a compact latent space for both data reconstruction and generation; Generative Adversarial Networks (GANs) use adversarial training to create synthetic output that closely resembles real reference data; Diffusion Models are parametrized Markov chains, trained using variational inference, to generate samples that converge to match the underlying data distribution (Kingma & Welling 2013; Jetchev *et al.* 2016; Goodfellow *et al.* 2020; Ho *et al.* 2020). In the context of Bayesian inversion, DGMs possess a crucial property: they generate realizations that exhibit patterns consistent with the TI when sampling an uncorrelated standard normal or uniform distributed latent space (Laloy *et al.* 2017). The incorporation of such a low-dimensional parametrization representing the prior distribution simplifies the sampling process, thus making DGMs well-suited for MCMC schemes.

Using an effective parametrization for the prior (e.g. via PCA or DGMs) does not alleviate the computational cost associated with repeated forward modelling, which can often limit practical application of MCMC schemes. However, it opens up the possibility of using surrogate modelling. Various classes of surrogate models, such as those based on Gaussian process models or Kriging (Sacks *et al.* 1989; Rasmussen 2003) and polynomial chaos expansions (PCEs; Xiu & Karniadakis 2002; Blatman & Sudret 2011), can be used to effectively solve low-dimensional Bayesian inverse problems (Marzouk *et al.* 2007; Marzouk & Xiu 2009; Higdon *et al.* 2015; Nagel 2019; Wagner *et al.* 2020, 2021a).

Recently, Meles *et al.* (2022) presented a Bayesian framework that uses a shared PCA-based parametrization for characterizing prior information and modelling travel times through PCE. In this paper, we show that the direct application of that methodology produces suboptimal results when the input is parametrized in terms of latent variables associated with a DGM. Indeed, a prerequisite in this framework is that the parametrization chosen to describe and explore the prior allows for accurate surrogate modelling of the physical data used in the inversion. However, the relationship between the parametrization provided by DGMs and the corresponding physical output is typically highly non-linear, and training a surrogate capturing such complex relationship can be extremely challenging or infeasible.

We present a strategy that allows the use of a DGM parametrization to define the prior distribution and explore the posterior distribution while still enabling accurate surrogate modelling. We achieve this goal by using surrogate modelling with either global or local PCA decompositions of the input generated by the underlying DGM during the MCMC process. Through the decoupling of the MCMC parametrizations and the parametrizations, we preserve the advantageous prior representation capabilities of DGMs while simultaneously achieving precise PCA-based surrogate modelling, thereby significantly speeding up forward calculations.

2 METHODOLOGY

2.1 Bayesian inversion

Forward models are mathematical tools that quantitatively evaluate the outcome of physical experiments. We refer to the relationship between input parameters and output values as the 'forward problem':

$$\mathcal{F}(\boldsymbol{u}) = \boldsymbol{y} + \epsilon. \tag{1}$$

Here, $\mathcal{F}$, $\boldsymbol{u}$, $\boldsymbol{y}$ and ϵ stand for the physical law or forward operator, the input parameters, typically representing local media properties, the n-dimensional output of a physical experiment and a noise term, respectively. Given the manuscript's significant emphasis on modelling, we deviate from the standard formalism, and position the forward operator on the left-hand side and related subsequent equations, while placing the observed data and the noise term on the right-hand side. The goal of the 'inverse problem' is to infer properties of $\boldsymbol{u}$ conditioned by the data $\boldsymbol{y}$ while taking into account any available prior information about $\boldsymbol{u}$. For a constant model dimension, a general solution to this problem can be expressed in terms of the unnormalized posterior distribution:

$$P(\boldsymbol{u}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{u})P(\boldsymbol{u}) = L(\boldsymbol{u})P(\boldsymbol{u}). \tag{2}$$

Here, $P(\boldsymbol{u}|\boldsymbol{y})$ is the posterior distribution of the input parameter $\boldsymbol{u}$ given the data $\boldsymbol{y}$, $P(\boldsymbol{y}|\boldsymbol{u})$ (also indicated as $L(\boldsymbol{u})$ and known as 'the likelihood') is the probability of observing the data $\boldsymbol{y}$ given the input parameter $\boldsymbol{u}$, while $P(\boldsymbol{u})$ is the prior distribution in the input parameter domain. In case of Gaussian noise, the likelihood takes the following form:

$$L(\boldsymbol{u}) = \left(\frac{1}{2\pi}\right)^{n/2} |\boldsymbol{C}_d|^{-1/2} \exp\left[-\frac{1}{2}(\mathcal{F}(\boldsymbol{u}) - \boldsymbol{y})^T \boldsymbol{C}_d{}^{-1}(\mathcal{F}(\boldsymbol{u}) - \boldsymbol{y})\right]. \tag{3}$$

where the covariance matrix $\boldsymbol{C}_d$ accounts for data uncertainty. Note that throughout this paper, we adhere to the common formalism used in Geophysics, utilizing the same symbol to denote both individual instances of a random variable and the random variable itself (Tarantola 2005; Aster *et al.* 2018). To draw samples from the unnormalized posterior in eq. (2), common practice is to use a Markov chain Monte Carlo (MCMC) methods to draw samples proportionally from $P(\boldsymbol{u}|\boldsymbol{y})$ (Hastings 1970). However, computing $L(\boldsymbol{u})$ requires the solution of a forward problem, which can be demanding in Bayesian inversions as this evaluation needs to be repeated many times. In the following sections we discuss how this problem can be approached by using a low-dimensional latent representation and surrogate modelling to evaluate $P(\boldsymbol{u})$ and approximate $L(\boldsymbol{u})$, respectively.

2.2 Bayesian inversion in latent spaces

Parametrizations are well suited for surrogate modelling when they encode the prior distribution and effectively simplify the input–output relationship within the problem under investigation. Meles *et al.* (2022) used variables defined in terms of Principal Components to (i) represent the prior distribution related to a random

Gaussian field on a low-dimensional manifold and (ii) learn an accurate surrogate to compute the forward problem. However, it is not generally granted that a parametrization can achieve both (i) and (ii). For representing the prior distribution, we can utilize manifold identification using DGMs. This involves utilizing a DGM to characterize a latent space using a set of coordinates (here indicated as $\boldsymbol{z}$) with a statistical distribution defined prior to the training. The DGM allows mapping between this latent space to the physical space through a decoder/generator, denoted here as $\mathcal{G}_{\rm DGM}$. For a given random realization $\boldsymbol{z}$, the decoder operation $\mathcal{G}_{\rm DGM}(\boldsymbol{z})$ produces an output $\boldsymbol{u}$ in the physical space that adheres to the characteristics of the prior distribution. The use of this new set of coordinates $\boldsymbol{z}$ casts the inverse problem on the latent manifold as:

$$P(\boldsymbol{z}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{z})P(\boldsymbol{z}). \tag{4}$$

While formally identical to eq. (2), eq. (4) involves significant advantages. Not only is $\boldsymbol{z}$ typically low-dimensional (consisting typically of a few tens of variables instead of many thousands of variables) but we can also design the corresponding statistical prior distribution $P(\boldsymbol{z})$ as desired. Here, we impose during training that $P(\boldsymbol{z})$ is a multivariate standard Gaussian distribution.

2.3 Decoupling of inversion and forward modelling domains in MCMC inversions

We now rewrite the forward problem in eq. (1) using the new coordinates:

$$\mathcal{M}_{\rm DGM}(\boldsymbol{z}) = \boldsymbol{y} + \epsilon, \tag{5}$$

where $\mathcal{M}_{\rm DGM} = \mathcal{F} \circ \mathcal{G}_{\rm DGM}$, $\circ$ stands for function composition, and we assume no error induced by the DGM dimensionality reduction, which in turns implies that the corresponding likelihood is the same as that in eq. (3). The complexity and non-linearity of $\mathcal{G}_{\rm DGM}$ imply that the forward operator $\mathcal{M}_{\rm DGM}$ exhibit considerable irregularity, making it difficult to learn a surrogate model due to typical issues of neural networks, such as those associated with limited data availability and overfitting (Bejani & Ghatee 2021). Consequently, we investigate alternative approaches that avoids using the latent parametrization for surrogate modelling while retaining it for the prior representation. Building upon Meles *et al.* (2022), we explore surrogate modelling based on PCA-inputs spanning the *Global* spatial extent of the input. Without any loss of generality, we consider a complete set of Global Principal Components (in the following, GPCs) for realizations of the DGM (implemented via $\mathcal{G}_{\rm GPC}(\boldsymbol{x}^{\rm full}_{\rm GPC}) = \mathcal{G}_{\rm DGM}(\boldsymbol{z}) = \boldsymbol{u}$) and rewrite eq. (1) as:

$$\mathcal{M}_{\rm GPC}\left(\boldsymbol{x}^{\rm full}_{\rm GPC}\right) = \boldsymbol{y} + \epsilon, \tag{6}$$

where $\mathcal{G}_{\rm GPC}$, and $\mathcal{M}_{\rm GPC}$ are the physical distribution and the model associated with the GPCs and therefore $\mathcal{M}_{\rm GPC} = \mathcal{F} \circ \mathcal{G}_{\rm GPC}$. We will show in that the linear relationship $\mathcal{G}_{\rm GPC}(\boldsymbol{x}^{\rm full}_{\rm GPC}) = \boldsymbol{u}$ helps in implementing a surrogate of $\mathcal{M}_{\rm GPC}$, provided that the input and the model can be faithfully represented as operating on an effective M-dimensional truncated subset $\boldsymbol{x}_{\rm GPC}$ of the new coordinates $\boldsymbol{x}^{\rm full}_{\rm GPC}$, that is:

$$G_{\rm GPC}(\boldsymbol{x}_{\rm GPC}) \approx \boldsymbol{u} \Rightarrow \mathcal{M}_{\rm GPC}(\boldsymbol{x}_{\rm GPC}) = \boldsymbol{y} + \hat{\epsilon}, \tag{7}$$

where $\hat{\epsilon}$ is a term including both observational noise and modelling errors related to the projection on the subset represented by $\boldsymbol{x}_{\rm GPC}$. Due to the error introduced by the truncation, the likelihood associated with the GPC parametrization deviates from the one of eq. (3) However, when a substantial number of GPCs are utilized, the extent of such difference tends to be minimal.

The formulation given above relies on the *weak* hypothesis that *Global* proximity in the input domain leads to proximity in the output domain $\boldsymbol{u}$ (i.e. the observed quantity under investigation). Critically, when the output functionally depends mainly on a subset L of the entire domain of $\boldsymbol{u}$, proximity in the output domain can be attained by approximating the input within this *Local* region. Based on this *strong* physically-informed assumption, we can achieve this goal by means of a Local PCA decomposition restricted to L:

$$G_{\rm LPC}(\boldsymbol{x}_{\rm LPC}) \restriction_L \approx \boldsymbol{u} \restriction_L \Rightarrow \mathcal{M}_{\rm LPC}(\boldsymbol{x}_{\rm LPC}) = \boldsymbol{y} + \hat{\epsilon}, \tag{8}$$

where *LPC* refers to Local Principal Components (in the following LPCs) and $\restriction_L$ restricts the validity of these relationships to the subset L. However, because the area spanned by LPCs is smaller than that of GPCs, we expect to need fewer *LPC*s than *GPC*s to achieve a satisfactory approximation in the output domain. Note that this change of coordinates is not invertible even if a complete set of LPCs is used. Note that also for the LPC parametrization, the corresponding likelihood differs from that in eq. (3), but that using a proper number of components can render this misfit negligible.

Generally speaking, a surrogate model is a function that seeks to emulate the behaviour of an expensive forward model at negligible computational cost per run. Clearly, the function a forward solver has to model depends on the set of coordinates used to represent the input. For simplicity, we discuss here a surrogate for $\mathcal{M}_{\rm GPC}(\boldsymbol{x}_{\rm GPC})$, namely $\tilde{\mathcal{M}}_{\rm GPC}$ satisfying:

$$\tilde{\mathcal{M}}_{\rm GPC}(\boldsymbol{x}_{\rm GPC}) \approx \mathcal{M}_{\rm GPC}(\boldsymbol{x}_{\rm GPC}). \tag{9}$$

Once available, surrogate models can be used for likelihood evaluation in MCMC inversions, with a modified covariance operator $\boldsymbol{C_D} = \boldsymbol{C}_d + \boldsymbol{C}_{\rm Tapp}$ comprising the covariance matrices $\boldsymbol{C}_d$ and $\boldsymbol{C}_{\rm Tapp}$ accounting for data uncertainty and modelling error, respectively. In such cases, the likelihood not only shows a mild dependence on the parametrization but also on the surrogate model. This relationship is expressed as follows:

$$L(\tilde{M}(\boldsymbol{x}_{\rm GPC})) = \left(\frac{1}{2\pi}\right)^{n/2} |\boldsymbol{C_D}|^{-1/2} \exp\left[-\frac{1}{2}(\tilde{\mathcal{M}}_{\rm GPC}(\boldsymbol{x}_{\rm GPC}) - \boldsymbol{y})^T \boldsymbol{C_D}^{-1}(\tilde{\mathcal{M}}_{\rm GPC}(\boldsymbol{x}_{\rm GPC}) - \boldsymbol{y})\right]. \tag{10}$$

where $|\boldsymbol{C_D}|$ is the determinant of the covariance matrix $\boldsymbol{C_D}$ (Hansen *et al.* 2014). Similar formulas for the likelihood can be derived for any other parametrization of the input space, such as those associated with DGMs of LPCs. The substantial differences in likelihoods associated with the DGM, GPC and LPC parametrizations can potentially give rise to significantly different estimations of the posterior distribution.

Surrogate models all adhere to a fundamental principle: the more complex the input–output relationship, the higher the computational demands needed to build the surrogate model (e.g. in terms of training set). Furthermore, the efficiency of constructing a surrogate is significantly affected by the number of input parameters, and it can become unfeasible when the input dimensionality exceeds several tens of parameters, thus surrogate modelling often relies on some kind of dimensionality reduction. The dimensionality reduction step does not necessarily need to be invertible since what holds significance is the supervised performance, specifically the minimization of modelling error (Lataniotis *et al.* 2020). Surrogate models exhibit their peak potential when dealing with low-dimensional input spaces, provided that such simplicity does not entail a complex input–output relationship.

Based on these general considerations, we can qualitatively anticipate different surrogate performances when operating on different

latent variables (i.e. $\tilde{\mathcal{M}}_{\mathrm{DGM}}(\boldsymbol{z})$), GPCs (i.e. $\tilde{\mathcal{M}}_{\mathrm{GPC}}(\boldsymbol{x}_{\mathrm{GPC}})$) and LPCs (i.e. $\tilde{\mathcal{M}}_{\mathrm{LPC}}(\boldsymbol{x}_{\mathrm{LPC}})$). We can expect that using the latent variables of a DGM involves a lower dimensionality, albeit at the cost of a complex input–output relationship. On the other hand, using GPCs likely results in a considerably simpler input–output relationship but necessitates a larger number of variables, as the entire input domain needs to be approximated. In cases where the underlying forward model permits, the LPCs projection is likely to offer similarly straightforward input–output characteristics as the GPCs approach but with fewer parameters, thus facilitating the training of a surrogate.

To effectively combine MCMC methods with surrogate modelling, it is essential to ensure a high level of accuracy in the output domain. To achieve this goal, we propose decoupling the parametrizations used for inversion and surrogate modelling. The latent representation provided by the DGM is used to evaluate $P(\boldsymbol{z})$ and explore the posterior. Once a prior has been defined *and* a surrogate modelling strategy devised, a Metropolis–Hastings MCMC algorithm can be used to sample the posterior distribution $P(\boldsymbol{z}|\boldsymbol{y})$ (Hastings 1970). A sample of the posterior in the physical space, $P(\mathcal{G}_{\mathrm{DGM}}(\boldsymbol{z})|\boldsymbol{y})$, is then also available via mere application of the $\mathcal{G}_{\mathrm{DGM}}$ to draws from $P(\boldsymbol{z}|\boldsymbol{y})$. For each step in the MCMC process, we consider three strategies with surrogates operating on latent variables, on GPCs and or LPCs associated with the field $\mathcal{G}_{\mathrm{DGM}}(\boldsymbol{z})$. We depict these three strategies for a travel time tomography problem in Fig. 1.

3 APPLICATION TO GPR CROSSHOLE TRAVEL TIME TOMOGRAPHY

In the previous section, we briefly covered the basic principles of Bayesian methods and emphasized the significance of dimensionality reduction and surrogate modelling for their implementation. Now, we integrate these concepts to tackle GPR crosshole traveltime tomography with MCMC. GPR wave-propagation is governed by the distribution of dielectric permittivity (ϵ) and electric conductivity (σ) in the subsurface. The wave propagation velocity mainly depends on permittivity, which is in turn related to porosity and water content through petrophysical relationships (Gloaguen *et al.* 2005). GPR data can be collected in a variety of configurations, with crosshole designs being particularly well suited for groundwater investigations (LaBrecque *et al.* 2002; Annan 2005).

3.1 Experimental setup

We target lossless media represented by binary images with two facies of homogeneous GPR velocity (6×10^7 and 8×10^7 m s^{-1}) resembling river channels (Strebelle 2002; Zahner *et al.* 2016; Laloy *et al.* 2018). For the representation of the prior and posterior exploration, we consider coordinates induced by a VAE (in the following the subscript $_{\mathrm{VAE}}$ refers to this specific DGM parametrization) as recent research suggests that their lower degree of non-linearity in the corresponding networks compared with GANs makes them more amenable for modelling and inversion (Lopez-Alvis *et al.* 2021; Levy *et al.* 2023). The details of the VAE utilized in this study can be found in Laloy *et al.* (2017) and Lopez-Alvis *et al.* (2021) with the dimension of the latent space being 20. As for the output, we consider arrival-times associated with the crosshole configuration displayed in Figs 2(a)–(e) with 12 evenly spaced sources and receivers located in two vertically oriented boreholes. The distance between the boreholes is 4.6 m, while the spacing between sources/receivers is 0.9 m, leading to 144 traveltimes. We use both an eikonal and a 2-D finite difference time domain (FDTD) solver to simulate noise-free propagation of GPR waves (Irving & Knight 2006; Hansen *et al.* 2013). For the FDTD code each source is characterized by a 100 MHz Blackman–Harris function, while perfectly matched layers surrounding the propagation domain are used to prevent spurious reflections from contaminating the data, while appropriate space-time grids are used to avoid instability and dispersion artefacts. Traveltimes are picked automatically based on a threshold associated with the relative maximum amplitude of each source–receiver pair.

3.2 PCE modelling within MCMC

For surrogate-based modelling, we rely on PCE modelling due to its efficiency, flexibility and ease of deployment (Xiu & Karniadakis 2002; Blatman & Sudret 2011; Métivier *et al.* 2020; Lüthen *et al.* 2021). We here summarize the most relevant aspects of PCE modelling with a surrogate $\mathcal{M}_{\mathrm{GPC}}(\boldsymbol{x}_{\mathrm{GPC}})$, but similar considerations apply to $\mathcal{M}_{\mathrm{DGM}}$, $\mathcal{F}$ or $\mathcal{M}_{\mathrm{LPC}}$, albeit with relevant caveats or advantages that will be discussed below. PCE approximate functions in terms of linear combinations of orthonormal multivariate polynomials $\Psi_{\boldsymbol{\alpha}}$:

$$\tilde{\mathcal{M}}_{\mathrm{GPC}}(\boldsymbol{x}_{\mathrm{GPC}}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}_{\mathrm{GPC}}), \tag{11}$$

where M is the dimension of $\boldsymbol{x}_{\mathrm{GPC}}$ and $\mathcal{A}$ is a subset of $\mathbb{N}^M$ implementing a truncation scheme to be set based on accuracy requirements and available computational resources (Xiu & Karniadakis 2002). The training of the coefficients $a_{\boldsymbol{\alpha}}$ is computationally unfeasible when the input domain is high-dimensional (the case for a surrogate $\tilde{\mathcal{F}}(\boldsymbol{u})$ of $\mathcal{F}(\boldsymbol{u})$). Moreover, when the imposed truncation pattern cannot fully account for the degree of non-linearity of the underlying model [the case for a surrogate $\tilde{\mathcal{M}}_{\mathrm{VAE}}(\boldsymbol{z})$ of $\mathcal{M}_{\mathrm{VAE}}(\boldsymbol{z})$], the still unbiased PCE predictions are inevitably affected even if the input domain is low-dimensional. On the other hand, using tailored PCA decomposition as required by $\tilde{\mathcal{M}}_{\mathrm{LPC}}(\boldsymbol{x}_{\mathrm{LPC}})$ of $\mathcal{M}_{\mathrm{LPC}}(\boldsymbol{x}_{\mathrm{LPC}})$ could decrease the computational burden and increase accuracy. In any case, the surrogate forward modelling predictor can be evaluated at a negligible cost by direct computation of eq. (11) and its accuracy estimated using a validation set or cross-validation techniques (Blatman & Sudret 2011; Marelli *et al.* 2021).

We here test the three strategies discussed in Section 2.3 using PCE for surrogate modelling in conjunction with the VAE parametrization for prior-sampling. For each strategy we build a corresponding PCE to model traveltime arrivals using the Matlab Package UQlab (Marelli & Sudret 2014; Marelli *et al.* 2021). To offer a fair comparison, we use the same training and validation data sets for all proposed schemes.

3.2.1 *VAE-PCE strategy*

In the first strategy, referred to as VAE-PCE, the input for the PCE modelling are 20-D $\boldsymbol{z}$ vectors mapping the latent space into the physical one, that is: $\mathcal{G}_{\mathrm{VAE}}(\boldsymbol{z}) = \boldsymbol{u}$ (Lopez-Alvis *et al.* 2021). This choice amounts to applying the strategy by Meles *et al.* (2022), as the same parametrization is used for both prior characterization and surrogate modelling.

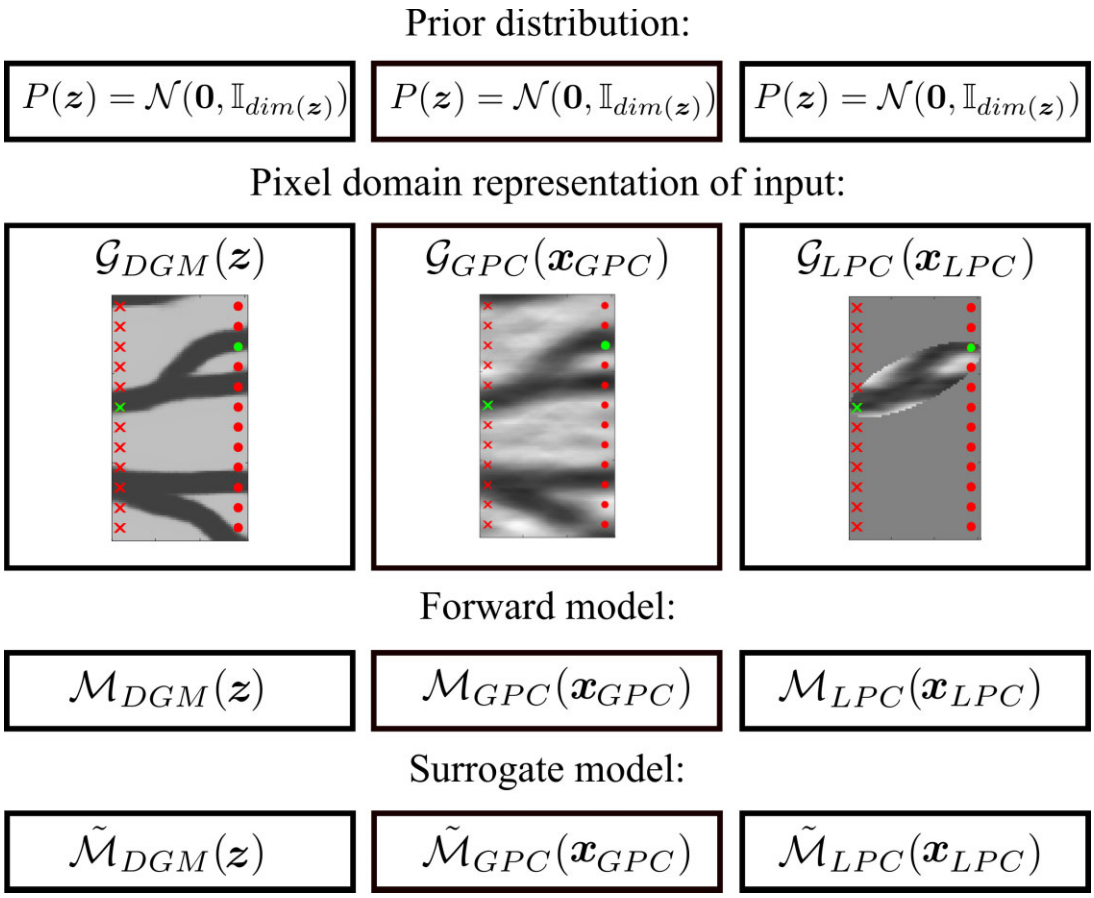

Figure 1. Graphical schematic of the three strategies discussed in this manuscript. The prior input distribution is always parametrized by the latent space of the DGM, with $P(\boldsymbol{z}) = \mathcal{N}(0, \mathbb{I}_{dim(\boldsymbol{z})})$, while the input to the reference/surrogate model depends on the chosen strategy. When the DGM parametrization is chosen, draws of the latent variables are considered (left column). With the PCA strategies, Globally or Locally defined G/LPCs are used for the Global and Local approach, respectively (central and right column). The Global approach uses a set $\boldsymbol{x}_{\text{GPC}}$ of PCs defined across the entire domain, illustrated by the corresponding velocity field that extends throughout the entire medium. Conversely, the Local approach utilizes a set $\boldsymbol{x}_{\text{LPC}}$ of PCs that cover only specific portions of the domain chosen based on physical considerations. This is exemplified by the corresponding velocity field, which is confined to a limited portion of the medium. In this illustration, as we consider travel time modelling, the Local domain is defined in terms of fat-ray sensitivity kernels (for more details, please refer to Section 3.2).

3.2.2 Global-PCA-PCE strategy

The second strategy, in the following Global-PCA-PCE, uses inputs of the PCE modelling defined in terms of projections on prior-informed PCA components spanning the entire domain. More specifically, in the Global-PCA-PCE approach we randomly create a total of 1000 slowness realizations $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$ from the prior and compute the corresponding principal components (see Fig. 3). The input for PCE in the Global-PCA-PCE approach are then the projections of $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$ on a to-be-chosen number of such GPCs. Following Meles *et al.* (2022), all PCA processes are defined in terms of slowness.

The effective dimensionality of the input with respect to $\mathcal{M}_{\text{GPC}}$, that is, the number of GPCs representing the input, is not a-priori defined. Following a similar approach to Meles *et al.* (2022), the effective dimensionality is here assessed by comparison with the reference solution in the output domain with respect to the noise level. In Figs 4(a) and (e), two velocity distributions are shown next to the approximate representations (Figs 4b–d and f–h) obtained by projecting them on 30, 60 and 90 GPCs, respectively. As expected, the reconstruction quality improves as more principal components are included.

To quantify the faithfulness of the various reduced parametrizations in terms of the output, we consider 100 realizations of the generative model, and compute the resulting histograms of the traveltime residuals using the reference forward solver. The root-mean-square error (in the following, rmse) of the misfit between the data associated with the original distribution and its projections on 30, 60 and 90 principal components, shown in Figs 4(i)–(k), are 1.60, 0.85 and 0.55 ns, respectively, that are to be compared to the expected level of GPR data noise of 1 ns for 100 MHz data (Arcone *et al.* 1998). The number of principal components (i.e. 90 PCs) required to approximate the forward solver below the expected noise level is larger than for the example considered by Meles *et al.* (2022) (i.e. 50 PCs). Building a PCE on such a large basis is challenging in terms of computational requirements and efficiency, and could lead to poor accuracy if a small training set is used. To address this, one approach is to either reduce the number of components, which introduces larger modelling errors, or explore alternative parametrizations that offer improved computational efficiency and accuracy. In this study, the Global-PCA-PCE approach utilizes 60 GPCs, while an alternative strategy is considered below that is based on physical considerations.

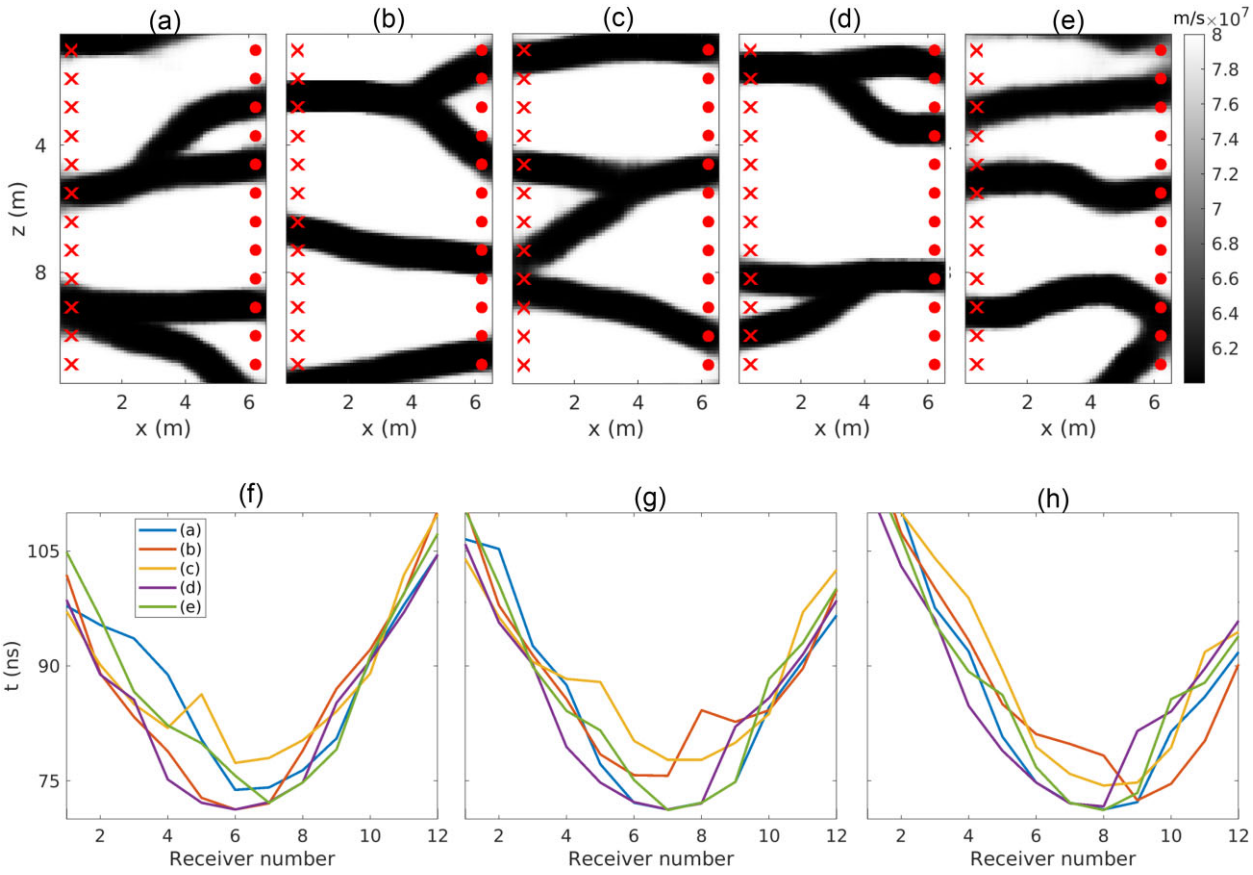

Figure 2. (a)–(e) Representative velocity fields generated by the decoder of the used VAE; crosses and circles stand for sources and receivers, respectively. (f)–(h) Corresponding exemplary traveltime gathers.

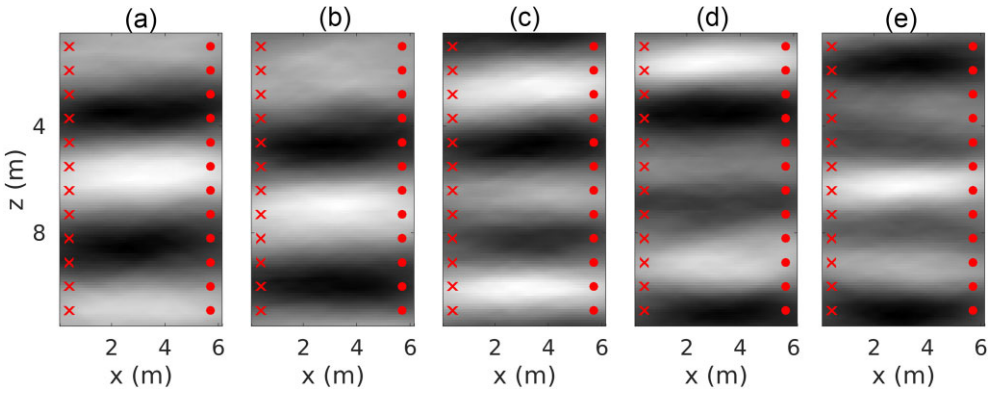

Figure 3. (a)–(j) The first five GPCs in the input domain corresponding to entire slowness fields. Crosses and circles stand for sources and receivers, respectively.

3.2.3 Local-PCA-PCE strategy

As mentioned in Section 2.3, an improved parametrization for surrogate modelling can sometimes be found by considering the forward problem's specific characteristics. The GPCs in the Global-PCA-PCE approach refer to the input field in its entirety. However, the actual first-arrival time for a given source–receiver combination depends only on a subdomain of the entire slowness distribution. This leads us to suggest a Local approach, in the following referred to as Local-PCA-PCE. Instead of using principal components describing the entire slowness field, we aim to use output-specific LPCs that characterize only the sub-domains impacting the physics of the problem (Jensen *et al.* 2000; Husen & Kissling 2001). We then expect that fewer LPCs are needed than GPCs to achieve the desired input–output accuracy. In practice, our construction of LPCs involves utilizing fat-ray sensitivity kernels, which capture the sensitivity of arrival-times to small perturbations in the subsurface model, thus providing valuable insights into the regions [corresponding to the L subset in eq. (8)] that have the most significant impact on the observed data (Husen & Kissling 2001). For a given source–receiver pair, the corresponding sensitivity kernel depends on the slowness field itself and its patterns can vary significantly (see Figs 5a–j). The sought Local decomposition needs to properly represent any possible slowness field within the prior, thus it reasonable to define it based on a representative sample of input realizations. To reduce

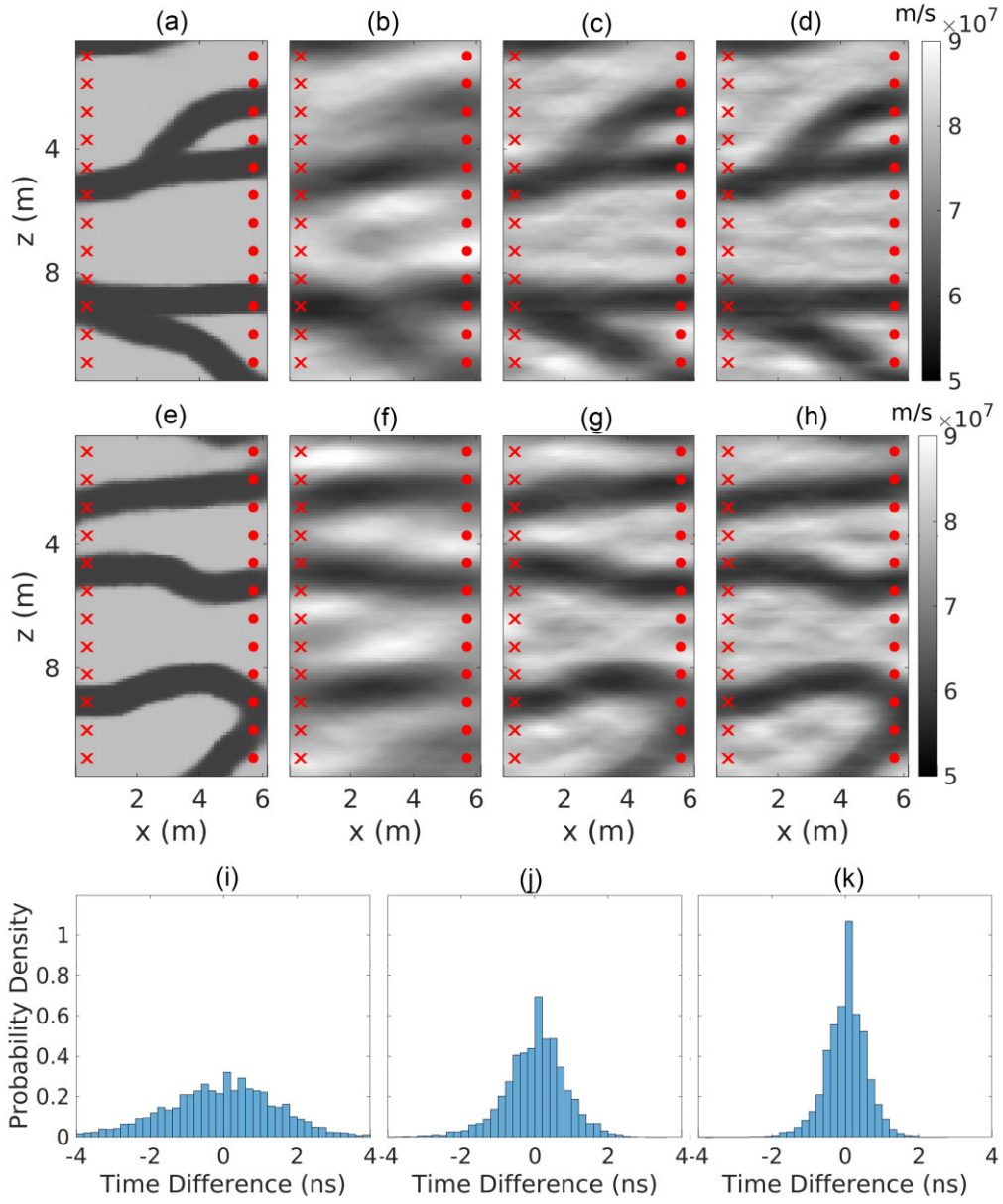

Figure 4. (a) and (e) Random velocity distributions and the corresponding representations using the first (b) and (f) 30, (c) and (g) 60, and (d) and (h) 90 GPCs defined with the Global approach; (i–k) corresponding histograms of traveltime residuals based on simulations on the true field and partial reconstructions for 100 velocity distributions. Crosses and circles stand for sources and receivers, respectively.

the overall PCE computational cost it is also convenient to limit the number of used output-driven decompositions. To achieve these goals, we assume that the prior model is stationary with respect to translation. Instead of focusing on each specific source–receiver pair (a total of 144), we can then consider each source–receiver angle (a total of 23). We then use a total of 1000 slowness realizations $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$ from the prior and build the corresponding fat-ray sensitivity kernels using the reference eikonal solver for each of the 23 possible angles (Hansen *et al.* 2013). For any given angle, we consider a cumulative kernel consisting of the sum of the absolute values of each kernel (green areas in Figs 5k–t). Such a cumulative kernel cover an area larger than each individual kernel but is still considerably smaller than the entire slowness field. For any possible additional input model, the corresponding sensitivity kernel is then very likely to be *geometrically* included in the area covered by the cumulative kernel (see Figs 5k–t). Based on this insight, we define principal components spanning only the area covered by such cumulative kernels or relevant parts thereof (e.g. a threshold can be taken into consideration to reduce the size of these subdomains). For the practical definition of the components, the cumulative kernels are either set to 0 or 1 depending on whether the corresponding value is below or larger than the threshold, respectively. We then multiply point by point the slowness distributions with the cumulative kernels, and consider the principal components of such products.

The first five LPCs are shown for given source–receiver pairs (Figs 6a–e and f–j). Note that the pattern variability is confined within the cumulative kernels, while in the complementary areas the values are 0. Note also that compared to the five principal components in Fig. 3, higher resolution details can be identified. Given the

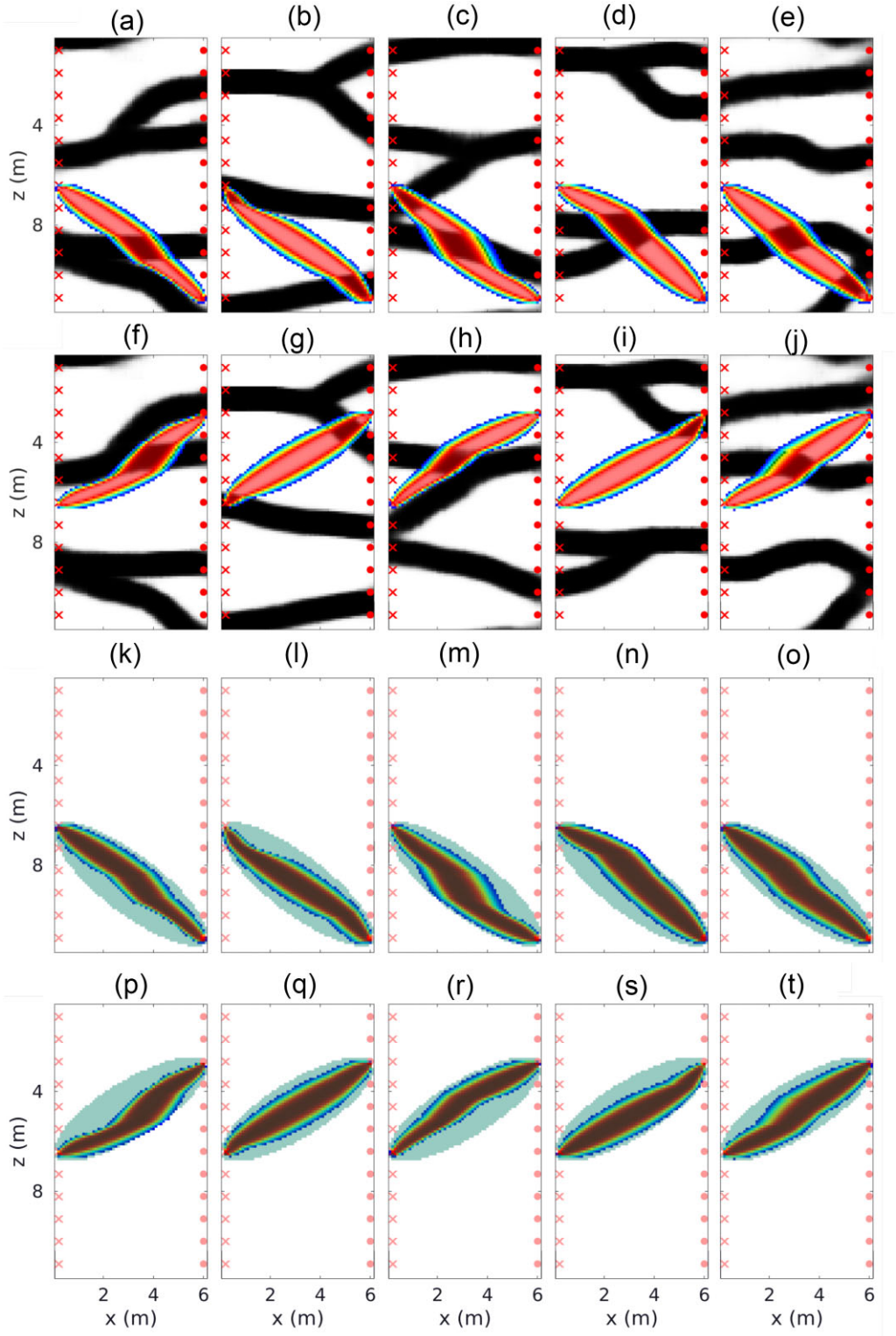

Figure 5. (a)–(j) For the velocity fields in Fig. 2, the sensitivity-based kernels for two arbitrarily selected source–receiver pairs are shown superimposed on the corresponding velocity distributions. (k)–(t) The same sensitivity kernels as in (a)–(j), but superimposed on the cumulative kernels (green shaded areas) used to define the support of the sensitivity-based LPCs used in the Local-PCA-PCE approach.

same number of principal components, we can then expect the input to be better presented in the physically relevant subdomain when the Local-PCA-PCE rather then the Global-PCA-PCE approach is followed. For all source–receiver pairs corresponding to the same altitude angle, the same kernels and principal components are used, provided they are shifted to cover the appropriate geometry.

In Figs 7(a)–(g) the two slowness distributions from Fig. 4 are shown next to the representations obtained by projecting them on

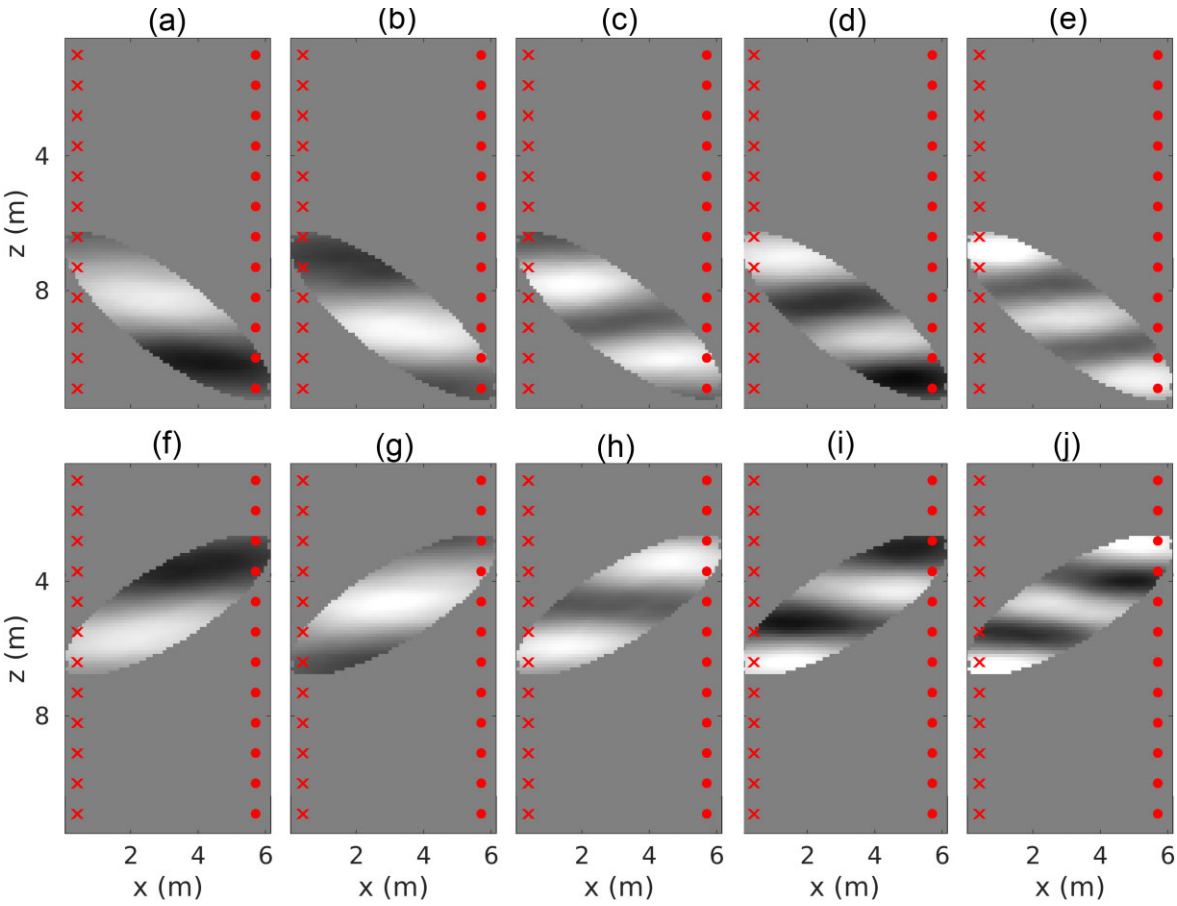

Figure 6. The first five LPCs in the input domain corresponding to the cumulative kernels associated with the source–receiver pairs considered in Figs 5(k)–(t).

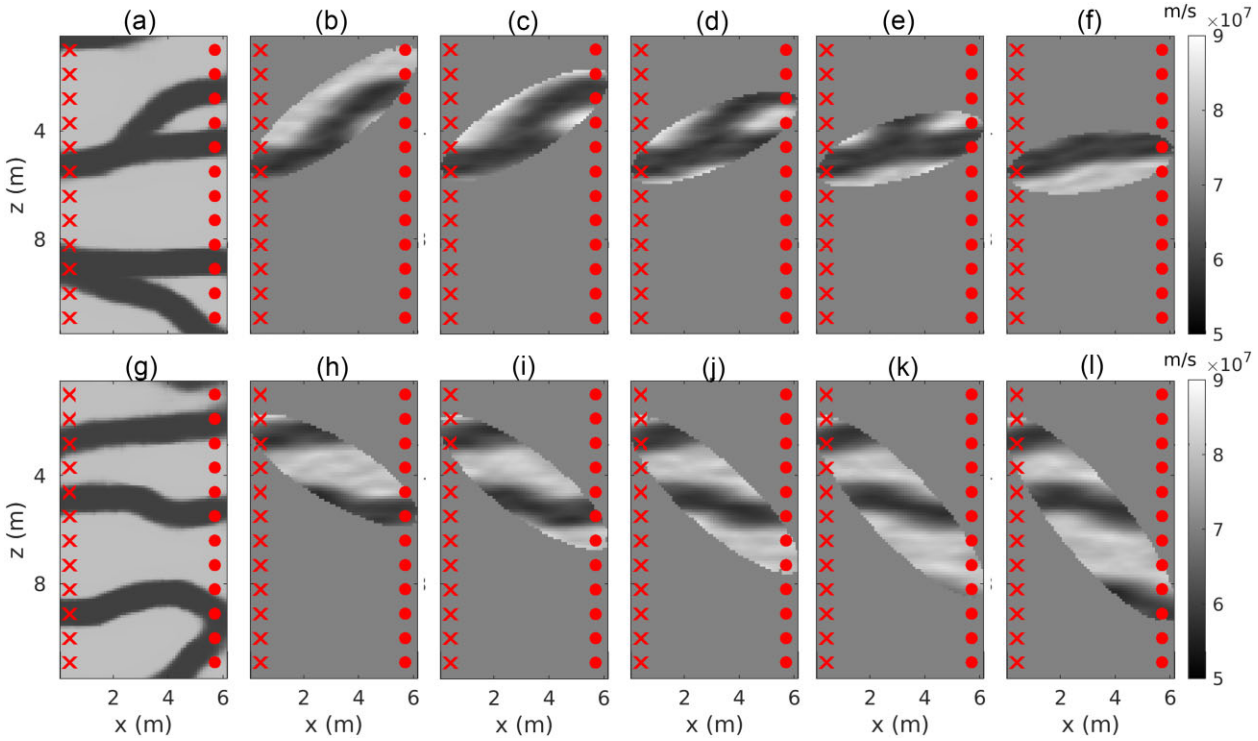

Figure 7. (a) and (g) Velocity fields with (b)–(f) and (h)–(l) the corresponding representations used for surrogate modelling based on the first 30 LPCs used in the Local-PCA-PCE approach. Different kernels are used for each source–receiver angle.

30 LPCs. In the areas complementary to the sensitivity kernels, the speed is set to 0.07 m ns^{-1}. Input reconstructions are remarkably improved with respect to when using the same number of GPCs [compare Figs 7(a)–(g) and (h)–(l) to Figs 4(b) and (f)] of the entire slowness field. More importantly, the modelling errors provided by using just 30 sensitivity-based LPCs is lower than what was previously provided by 90 standard components (i.e. rms ≈0.45 ns). By incorporating these tailored LPCs, we can attain enhanced output fidelity when utilizing truncated representations of the input. This enhanced fidelity proves particularly advantageous for the implementation of PCE, allowing for more precise and efficient modelling. Consequently, this approach holds substantial promise in achieving superior accuracy and computational efficiency in PCE-based analyses.

In summary, we have introduced three different parametrizations to be used as input for PCE. We consider coordinates inherited by the VAE, and principal components derived by considering either entire slowness fields or sensitivity-based subdomains. We refer to these three parametrizations in what follows as VAE-PCE, Global-PCA-PCE and Local-PCA-PCE, respectively.

3.3 PCE performance

We here analyse the PCE performance of the different input parametrizations for surrogate modelling introduced in Section 3.2. In agreement with Meles *et al.* (2022), we consider for each surrogate a total of 900 training sets consisting of noise-free travel times calculated by the eikonal solver and a polynomial degree p of five.

3.3.1 VAE-PCE performance

When the VAE parametrization is used as input, the PCE performance is rather poor, with a rmse of 2.01 ns, which is well beyond the expected noise level in the data and is consequently considered unsatisfactory (Fig. 8a). The poor performance is due to the highly non-linear map $\mathcal{M}_{\text{VAE}}$. In such a scenario, PCE does not provide accurate results even if the physical input of the validation set is exactly defined by $\mathcal{G}_{\text{VAE}}(\boldsymbol{z}) = \boldsymbol{u}$. In this scheme, note that the evaluation of $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$ is not required to compute the corresponding PCE. Given the low accuracy of the PCE performance for this parametrization, we can expect the corresponding likelihood to differ significantly from that involving exact modelling.

3.3.2 Global-PCA-PCE performance

Despite the comparatively poor reconstruction of the input (e.g. $\mathcal{G}_{\text{G/LPC}}(\boldsymbol{x}) \approx \boldsymbol{u}$) provided by the PCA approaches, the corresponding parametrizations perform well when used as input to build PCE surrogates, with the Global-PCA-PCE approach being outperformed by the Local-PCA-PCE scheme in terms of accuracy [rmse of 1.31 and 0.68 ns, respectively, see Figs 8(b) and (c) for the corresponding histograms]. In both of these cases, the PCE operates on more variables (i.e. 60 and 30 for the Global-PCA-PCE and Local-PCA-PCE parametrizations, respectively, versus 20 for the VAE-PCE scheme). Moreover, the input for the Global-PCA-PCE and Local-PCA-PCE schemes are projections of images, which require the evaluation of $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$, on G/LPCs. As such, evaluation of the corresponding PCEs is computationally more expensive for the PCA-based approaches than for the VAE-PCE case. Given the good accuracy of the PCE performance for this parametrization, we can expect the corresponding likelihood to mildly differ from that involving no truncation in the GPC projection and exact modelling.

3.3.3 Local-PCA-PCE performance

In the Local-PCA-PCE approach, for each of the 23 angles considered, training involves randomly chosen source–receiver pair data associated with identical angles, while the final rmse is computed on the standard 144 traveltime gathers. For a more comprehensive evaluation of the Local-PCA-PCE scheme, we incorporate the results of FDTD data processing. In this analysis, training and validation rely on the FDTD data, while the PCE implementation remains consistent with that of the eikonal data. Results are similar and well below the noise, with an rmse of 0.65 ns (the corresponding histogram is displayed in Fig. 8d). All PCE results are unbiased and the model errors have Gaussian-like distributions. Given the excellent accuracy of the PCE performance for this parametrization, we can expect the corresponding likelihood to only slightly differ from that involving no truncation in the LPC projection and exact modelling.

Depending on the input parametrization, PCEs approximate eikonal and FDTD solvers to different degrees. Fig. 9 represent the corresponding covariance matrices accounting for the modelling error of each surrogate model.

We now discuss the computational burden of each strategy when running on a workstation equipped with 16GB DDR4 of RAM and powered by a 3.5 GHz Quad-Core processor running Matlab and Python on Linux. We emphasize that our goal with the present manuscript is to propose novel methods to conjugate VAE and PCE rather than offer optimized implementations.

There are up to three relevant computational steps in the execution of a forward run for the VAE-PCE, Global-PCA-PCE, Local-PCA-PCE, eikonal and FDTD simulations, namely the evaluation of the physical input, $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$, its decomposition on either GPCs or LPCs and the actual evaluation of the forward or surrogate model. Not all methods require each of these steps. The VAE-PCE model, for example, is not a function of $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$ but depends on $\boldsymbol{z}$ only. Evaluation of the VAE-PCE model is extremely fast both when involving one or 35 (as in the MCMC inversion discussed below that is based on 35 chains) simultaneous model runs, taking on average ≈0.06s and ≈0.08s, respectively. Evaluation of the Python-based decoder $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$ required for all forward models except the VAE-PCE, is the bottleneck of the Matlab algorithms used herein, requiring ≈1.35s and ≈1.43s, respectively, when operating on one or 35 input when considering the eikonal solver. Nevertheless, this cost could be reduced by either implementing the decoder and the PCE within the same Python environment or by optimizing the calls to the Matlab/Python scripts in our code for greater efficiency. When in its native environment, evaluation of $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$ is actually very fast, taking only ≈0.005s and ≈0.08s when operating on one or 35 inputs, respectively. Still, note that this cost is overall negligible even in our non-optimized setting when considering expensive physics-based forward solvers such as FDTD. Only the Global-PCA-PCE and Local-PCA-PCE strategies require PCA decompositions. The Global-PCA-PCE approach is faster, requiring only up to ≈0.002s and ≈0.05s when applied to one and 35 input elements, respectively, while the Local-PCA-PCE method is slower, taking up to ≈0.06s and ≈0.23s in the same situation. For the Global-PCA-PCE method the cost of a single forward run is ≈0.52s, which is significantly more than for VAE-PCE. The difference between these two PCE strategies can be attributed to the significantly larger number of input parameters of the Global-PCA-PCE with respect

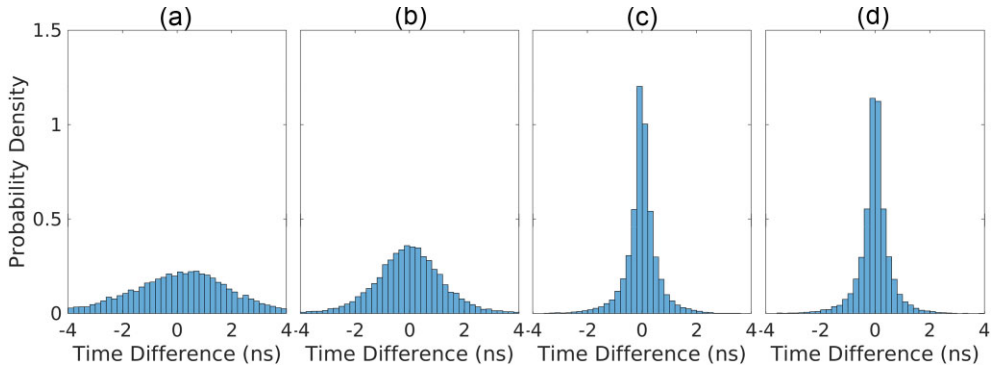

Figure 8. (a)–c) Histograms of the model error with respect to the PCE prediction when using the VAE (20 input parameters), Global (60 input parameters) and Local (30 input parameters per angle) parametrizations of the input in the PCE-based surrogate modelling, respectively, using the eikonal solver to compute the training set. (d) Histogram of the model error using Local parameters and FDTD reference data.

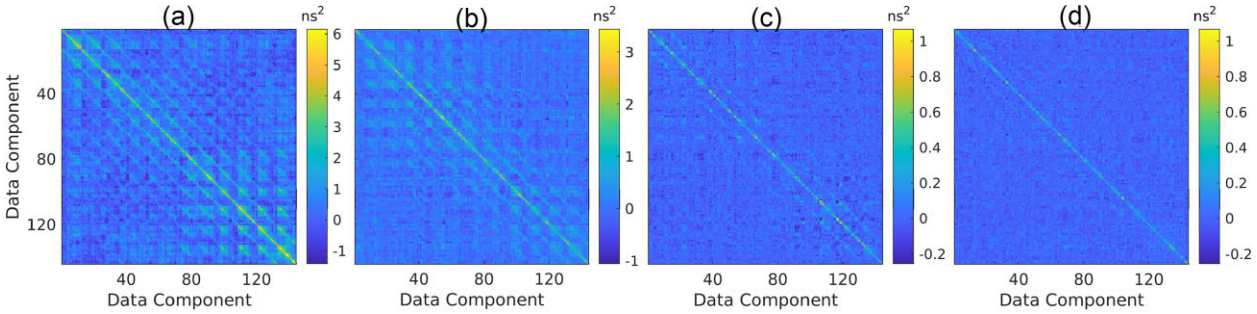

Figure 9. Model error covariance matrices associated with PCE-based surrogate modelling based on: (a) VAE-PCE, (b) Global-PCA-PCE and (c) Local-PCA-PCE schemes trained with 900 eikonal data sets. (d) Model error covariance matrix for the Local-PCA-PCE scheme trained with FDTD data.

to the VAE-PCE scheme (i.e. 60 versus 20). Note that the PCE model evaluations are vectorized and, therefore, the cost is almost the same when applied to 35 input ($\approx 0.57s$). Moreover, the computational cost of the Global-PCA-PCE method could be reduced by applying a PCA decomposition of the output, akin to what is proposed in Meles *et al.* (2022). Despite involving fewer variables than the Global-PCA-PCE approach, the Local-PCA-PCE method is slightly more computationally demanding with a cost of $\approx 0.64s$ and $\approx 0.65s$, respectively, when operating on one or 35 input, respectively. The increase in cost compared to the Global-PCA-PCE method depends on the fact that for each source–receiver angle (θ), the Local-PCA-PCE utilizes its own polynomial basis, denoted as Ψ^θ in the following.

In comparison with the PCE methods, the cost of the reference eikonal solver is basically a linear function of the number of input distributions it operates on. A single run requires $\approx 0.05s$, while for 35 velocity distributions the cost increases up to $\approx 1.67s$. As such, its cost is either significantly smaller or slightly larger than what is required by the Global/Local-PCA-PCE approaches. Finally, the cost required by the reference FDTD code is $\approx 120s$ and $\approx 4200s$ if operating on either one or 35 velocity distributions, which is orders of magnitude longer than for the eikonal or PCE models. These results are summarized in Table 1, where we estimate the performance of an ideally-optimized Local-PCA-PCE method benefiting from (a) evaluating $\mathcal{G}_{\text{VAE}}(\mathbf{z})$ in its native environment and (b) using a single family of polynomials Ψ^θ for all angles. In numerical results not presented herein, we find that choosing *any one* of the Ψ^θ families for all models provides nearly identical fidelity to what is achieved by using specifically tailored polynomials for each angle at the considerably smaller cost of $\approx 0.06s$ and $0.16s$ when applied to either one or 35 input, respectively. While such a result cannot be generalized, it is always possible to test the corresponding PCEs accuracy with a representative evaluation set. The option of relying on a single family of polynomials for the Local-PCA-PCE method is certainly to be taken into account when optimising the approach.

3.4 Inversion results

We now explore the performance of the different input parametrizations used for PCE-based surrogate modelling, namely VAE-PCE, Global-PCA-PCE and Local-PCA-PCE, when performing probabilistic MCMC inversion. The inversions were carried out using the UQLAB Matlab-based framework (Marelli & Sudret 2014; Wagner *et al.* 2021b). As reference model, consider the field shown in Fig. 10, which is used to generate a total of 144 traveltimes using the reference eikonal and FDTD solvers. Note that this field is not used to train the PCEs. Uncorrelated Gaussian noise characterized by $\sigma^2 = 1$ ns^2 was added to the data used for inversion.

We use a Metropolis–Hastings algorithm, and run 35 non-interacting Markov chains in parallel for 4×10^5 iterations per chain. During burn-in determined according to the Geweke method, the scaling factor of the proposal distribution is tuned such that an acceptance rate of about 30 per cent is achieved for each experiment (Geweke 1992; Brunetti *et al.* 2019). Finally, outlier chains with respect to the interquartile range statistics discussed in Vrugt *et al.* (2009) are considered aberrant trajectories and are ignored in the analysis of the results.

We first present the results for training data generated by an eikonal solver. We compare VAE-PCE, Global-PCA-PCE and

Table 1. Summary of the computational cost of the various steps for a single realization/batch of 35 input of the proposed algorithms. In addition to the strategies used in the MCMC examples discussed in the manuscript and summarized in the first four columns (i.e. the VAE-PCE, Global- and Local-PCA-PCE and eikonal schemes), we also consider the cost of the FDTD approach using the reference code (fifth column) and an optimized Local-PCA-PCE approach ideally benefiting from executing the VAE decoder in the same environment as the PCE model and based on a single polynomial family for all angles (sixth column).

	VAE-PCE	Global-PCA-PCE	Local-PCA-PCE	Eikonal	FDTD	Optimized Local-PCA-PCE
$\mathcal{G}_{\text{VAE}}(z)$	0	$\approx$1.35/1.43s	$\approx$1.35/1.43s	$\approx$1.35/1.43s	$\approx$1.35/1.43s	$\approx$0.005/0.08s
PCA	0	$\approx$0.002/0.05s	$\approx$0.06/0.23s	0	0	$\approx$0.06/0.23s
Forward	$\approx$0.06/0.08s	$\approx$0.52/0.57s	$\approx$0.64/0.65s	$\approx$0.05/1.67s	$\approx$120/4200s	$\approx$0.06/0.16s

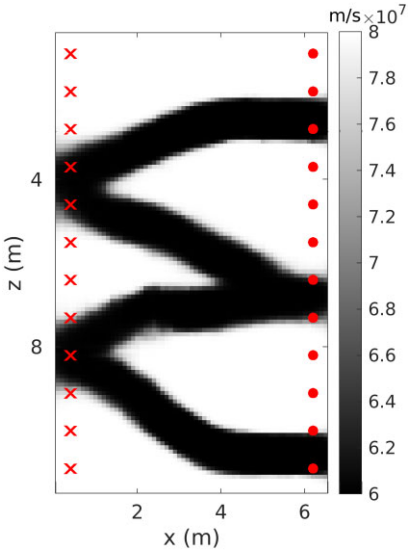

Figure 10. The reference velocity distribution used in the numerical inversion experiments.

Local-PCA-PCE inversion results to those achieved by using the eikonal solver, which represent the reference solution since the full physics solver is used in the entire MCMC process. This test provides an ideal benchmark as we compare the proposed PCE strategies with results derived from exact modelling. Note that such a comparison is not feasible for the FDTD scenario due to the computational burden associated with the corresponding MCMC process. Inversion results in terms of mean and standard deviations incorporating the model error [i.e. using the PCE-derived $\boldsymbol{C}_{\text{Tapp}}$ in eq. (10)] are shown in Figs 11(a)–(e). The mean of the posterior obtained using the VAE-PCE poorly resembles the reference velocity field, with relevant topological differences between the two (compare Fig. 10 to Fig. 11a). Note that the data misfit is particularly large (i.e. 3.49 ns). This poor performance is also obtained when considering other test models (see the Appendix). The mean of the posterior provided by the Global-PCA-PCE approach shares many features with the reference distribution, but differences are apparent in the lower and upper channelized structures. The similarity between the posterior mean and the true distribution increases significantly when the Local-PCA-PCE is used (compare Fig. 10 to Fig. 11c). These results also show close proximity with the posterior mean solution obtained by the eikonal solver (see Fig. 11d), that is, without any surrogate modelling. Also, when the FDTD Local-PCA-PCE is used, that is when training and inversion is performed based on FDTD data, an almost identical solution to what is achieved using the Local-PCA-PCE is obtained (see Fig. 11e). For this alternative data set, the use of the FDTD solver in the inversion algorithm would be extremely expensive and is not considered feasible to run for comparison purposes. The quality of the solution offered by the surrogate on FDTD data can be heuristically appreciated by noting its similarity to results obtained by the Local-PCA-PCE based on eikonal data, which in turn produces results close to those of the eikonal solver on eikonal data. Although not strictly consequential, it is to be expected that the results offered by the surrogate-based on FD data would also be similar to those that would have been obtained if using the FDTD solver on FDTD data.

High standard deviations of the posterior distribution are distributed in wide domains of the image when VAE-PCE is used (see Fig. 11f). In contrast, when Global-PCA-PCE (Fig. 11g), Local-PCA-PCE (Fig. 11h) and eikonal (Fig. 11i) solvers are used, high standard deviation values are found mainly along channel boundaries in agreement with other studies (Zahner *et al.* 2016). Convergence is assessed using the potential-scale reduction factor $\hat{R}$ that compares the variance of the individual Markov chains with the variance of all the Markov chains merged together (Gelman & Rubin 1992) calculated from the second half of the chains. Convergence is usually assumed if $R < 1.2$ for all parameters. In our experiments, full convergence for all of the 20 parameters is achieved when the VAE-PCE, the Global-PCA-PCE and the Local-PCA-PCE approaches are used. Six parameters do not converge when the eikonal solver is used, but the values of R are nevertheless close to 1.2 (see Fig. 12). Further quantitative assessments can be achieved by comparing the reference input and the corresponding inversion solutions in terms of input domain rmse, structural similarity (in the following, SSIM) and rmse in the output domain (Gneiting & Raftery 2007; Levy *et al.* 2022). Among these metrics, SSIM specifically evaluates the structural similarity between images, emphasizing their underlying patterns and details. It assigns a value between -1 and 1, with -1 indicating a notable dissimilarity and 1 denoting perfect similarity. Again, we see that the VAE-PCE performs poorly, with a low SSIM value. Better results are provided by the Global-PCA-PCE strategy. The Local-PCA-PCE scheme results are the closest to the reference solutions achieved using the eikonal solver. It is found that the FDTD Local-PCA-PCE performs similarly to the Local-PCA-PCE strategy. The results for all considered metrics are summarized in Table 2.

We also consider histograms of SSIM values in the corresponding posterior distributions (Fig. 13). The VAE-PCE posterior has low similarity with the reference model, with the maximum SSIM value being below 0.5. Closer proximity is found among samples obtained using the Global-PCA-PCE approach, a trend that is further improved when considering the Local-PCA-PCE scheme that shows some overlap with the results of the reference eikonal inversion. Note again that the statistics of the FDTD Local-PCA-PCE algorithm is again similar to the Local-PCA-PCE scheme. These findings can be further appreciated by considering random posterior realizations for each of the inversion strategies (see Fig. 14).

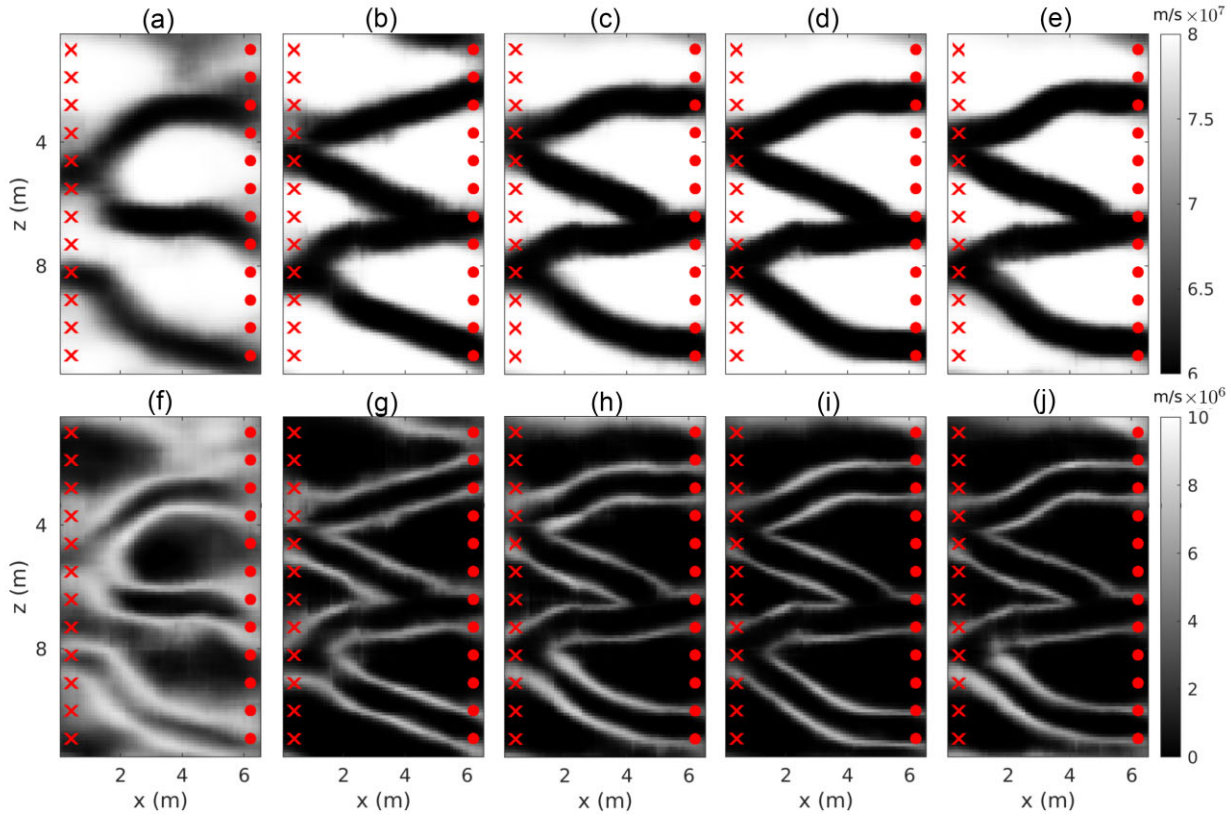

Figure 11. Posterior mean and standard deviation respectively for the (a, f) VAE-PCE inversion scheme, (b, g) Global-PCA-PCE, (c, h) Local-PCA-PCE, (d, i) eikonal and (e, j) FDTD Local-PCA-PCE inversion strategies.

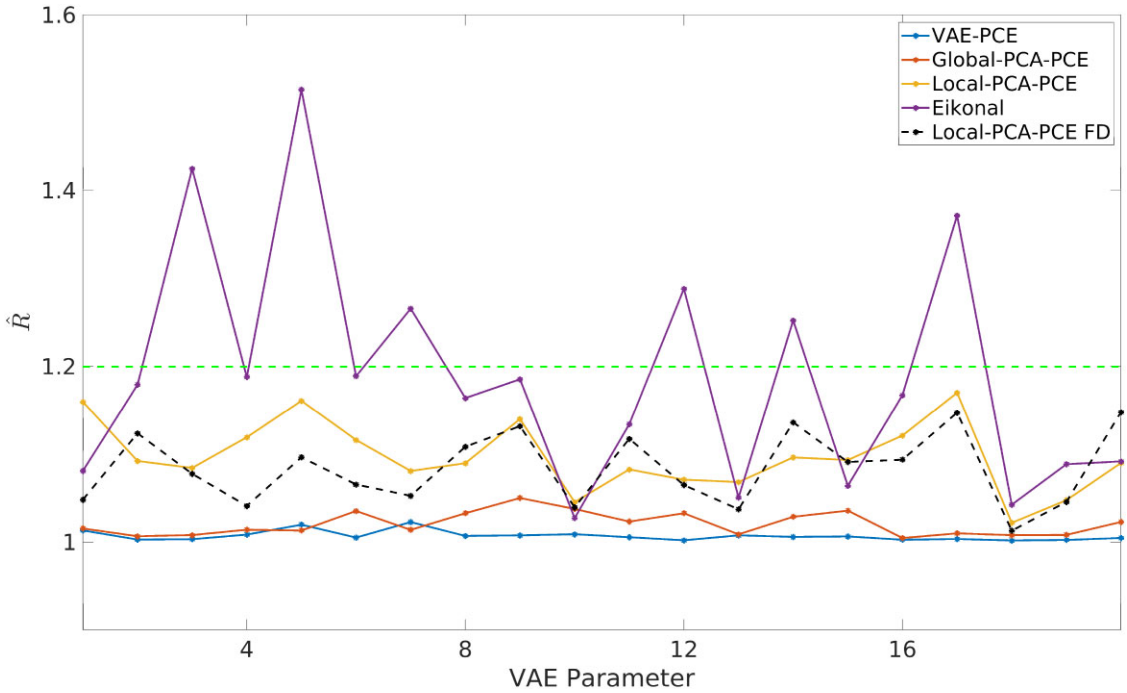

Figure 12. Gelman–Rubin statistics for the various inversion strategies using 35 chains after 4×10^5 iterations per chain.

4 DISCUSSION

Deep generative models offer a flexible framework for encoding complex spatial priors, enabling the description of intricate input distributions. However, proper parametrization of prior distributions alone does not ensure efficient estimation of posterior distributions when using MCMC methods. In many practical situations, the use of surrogate models becomes beneficial or even essential for evaluating the likelihood functions effectively. Surrogate modelling with

Table 2. Assessment of the inversion results in the input and output domains for the various surrogate modelling strategies.

Model	Rmse mean velocity	SSIM mean velocity	Rmse mean output
VAE-PCE	8.01×10^6 m s^{-1}	0.30	3.49 ns
Global-PCA-PCE	5.38×10^6 m s^{-1}	0.54	1.49 ns
Local-PCA-PCE	2.67×10^6 m s^{-1}	0.73	1.15 ns
Eikonal	1.57×10^6 m s^{-1}	0.87	1.01 ns
FD Local-PCA-PCE	3.06×10^6 m s^{-1}	0.71	1.15 ns

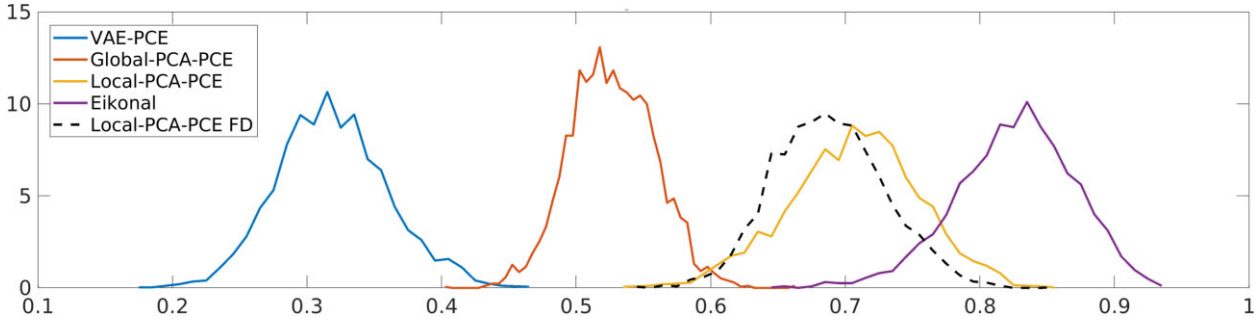

Figure 13. Histograms of SSIM values across the posterior distributions for the various inversion strategies.

PCE has become widespread in many disciplines. The massive decrease of the computational costs associated with PCE is achieved by approximating demanding computational forward models with simple and easy-to-evaluate functions. A key requirement is that the number of input variables describing the computational model is relatively small (i.e. up to a few tens) and that the target model can be approximated by truncated series of low-degree multivariate polynomials. The number of coefficients defining the PCE model grows polynomially in both the size of the input and the maximum degree of the expansion. When the reference full-physics model response is highly non-linear in its input parameters, the problem is typically non-tractable (Torre *et al.* 2019). Since the high-fidelity mapping of complex prior distributions provided by DGMs is based on highly non-linear relationships between latent variables and physical domains, replicating the performance of such networks and/or composite functions thereof (e.g. $\mathcal{M}_{\text{VAE}} = \mathcal{F} \circ G_{\text{VAE}}$ in eq. 5) using PCE is problematic. To circumvent this challenge, we have explored two PCA-based decompositions that facilitated the straightforward implementation of PCE. One decomposition was designed to encompass the entire input domain, while the other specifically focused on subdomains of particular physical relevance. Although this latter concept is investigated here in the context of traveltime tomography, the integration of problem-related PCs operating synergistically with DGM-defined latent parametrizations has the potential for broader applications.

Whatever the choice of input coordinates, for example, based on PCA or local properties of the input, the determining criterion for evaluating the quality of the corresponding PCE should always be performance on a representative validation set. In case of PCA, the lower bound of prediction misfit rmse can be *a priori* estimated by comparing the accuracy of the reference model acting on the full and the compressed domains, that is, $\mathcal{M}_{\text{G/LPC}}(\boldsymbol{x}_{\text{full}})$ and $\mathcal{M}_{\text{G/LPC}}(\boldsymbol{x})$. In our case, such lower bounds for the Global-PCA-PCE approach operating with 60 components is 0.85 ns. Using the Local-PCA-PCE scheme with 30 components only, the rmse drops to 0.55 ns. However, the accuracy of the corresponding Global/Local-PCA-PCE is worse, with rmse of 1.31 and 0.67 ns, respectively, mainly due to the small size of the training set. Note that while the lower bound of the PCA decomposition decreases when more PCs are taken into account, the corresponding accuracy of PCE is primarily limited by the size of the training set. Increasing the number of components can actually worsen PCE performance if the training set is insufficient to determine the polynomial coefficients, which, as mentioned, grow significantly with input size. In our case, using 90 components would imply an rmse of 1.39 ns, which is worse than what was obtained by the 60 components PCE.

Both the VAE-PCE and Global-PCA-PCE consist of 144 (i.e. the total number of source–receiver combinations) different PCE models operating, for each traveltime simulation, on an identical input, that is, the latent variables of the DGM or the 60 PCs characterizing the entire physical domain. On the other hand, the Local-PCA-PCE scheme consists of 23 (i.e. the total number of source–receiver altitude angles) different PCE models operating, for each traveltime simulation, on 144 different input, that is, the 30 LPCs characterizing each local subdomain. Since each of the 23 models operates on specific LPCs, the corresponding families of orthonormal polynomials Ψ^{θ} are different. This is in contrast with the Global-PCA-PCE method, for which each model operates via a single family of polynomials. Thus, the Local-PCA-PCE scheme is computationally more demanding than the Global-PCA-PCE (see Table 1). However, the use of a single family of polynomials can also be considered for the Local-PCA-PCE method, resulting in shorter run time. When considering computational performance, an optimal implementation of $\mathcal{G}_{\text{VAE}}(\boldsymbol{z})$ should also be sought.

In this study, to determine the minimum number of G/LPCs for constructing an accurate PCE, we assess the lower bound of output prediction misfit rmse as a function of the number of G/LPCs used. We project the input onto subsets of G/LPCs, typically ranging in the tens. This process generates non-binary images, which are then utilized to compute the output using the reference forward modelling. Alternatively, we could consider rebinarizing the reconstructed images as done in Thibaut *et al.* (2021). This approach would bring the projected input back into the prior, but this property is not necessarily relevant for the determination of PCE accuracy. Irrespective of the chosen reconstruction algorithm, the Local approach maintains a significant advantage over the Global method. When using

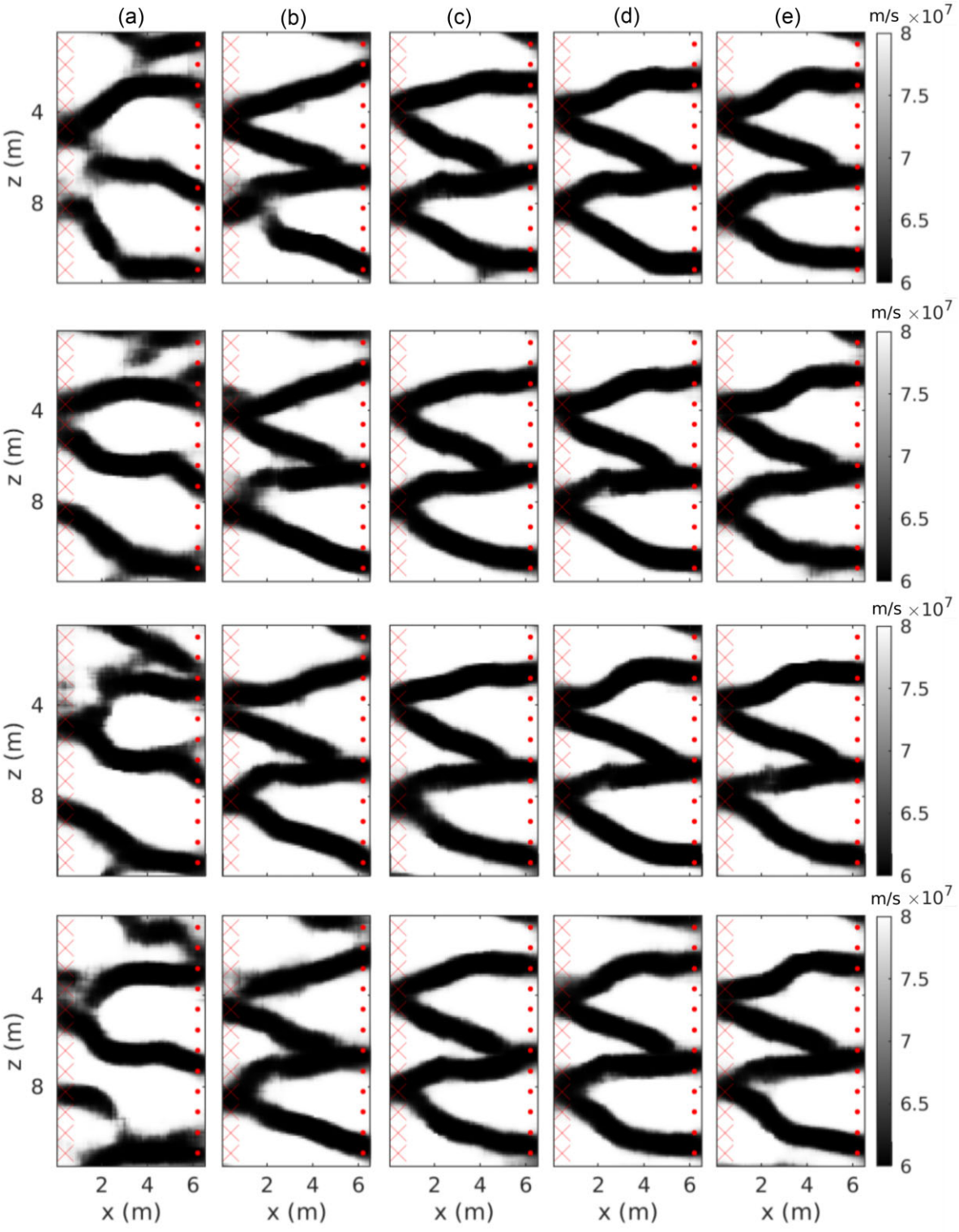

Figure 14. Random samples from the posterior obtained by using the (a) VAE-PCE, (b) Global-PCA-PCE, (c) Local-PCA-PCE, (d) eikonal and (e) FD Local-PCA-PCE strategies. The results in (c)–(e) are visually similar.

an equal number of components, LPCs, in fact, consistently yield superior approximations of the relevant input compared to GPCs.

We have seen that once a DGM-based latent parametrization has been found to reduce the effective dimensionality of the input domain and, based on PCA decompositions, high fidelity PCEs have been trained, MCMC inversion can be efficient. Relying on PCE rather than advanced deep learning methods for surrogate modelling can be advantageous in terms of ease of implementation, as potentially complex training of a neural network is not needed. Many effective sampling methods, such as Adaptive Metropolis, Hamiltonian Monte Carlo, or the Affine Invariant Ensemble Algorithm (Duane *et al.* 1987; Haario *et al.* 2001; Goodman & Weare

2010) could be considered in our workflow instead of the current use of a standard Metropolis–Hastings sampling algorithm.

Adaptation of the Global-PCA-PCE strategy presented here could easily be used for other imaging problems such as active or passive seismic tomography at different scales (Bodin & Sambridge 2009; Galetti *et al.* 2017). On the other hand, implementation of the Local-PCA-PCE schemes would depend on the properties of the corresponding sensitivity kernels, which would require more careful evaluation and problem-specific design.

5 CONCLUSIONS

Low-dimensional latent variables associated with deep generative models, such as VAE, conform well to complex priors, and provide an attractive basis to explore posterior model distributions using MCMC strategies. MCMC methods can also benefit greatly from surrogate modelling based on PCE, provided the forward model can be approximated by low-degree multivariate polynomials. However, this type of latent variable models tend to have a highly non-linear relation to data, and are, thus, poorly approximated by low-degree PCEs. As such, performing PCE-accelerated MCMC inversion based on a latent VAE-based parametrization for both inversion and as input to surrogate modelling leads to large posterior uncertainty due to the need to account for important modelling errors in the likelihood function. In the context of GPR traveltime tomography, for instance, PCE-based surrogate modelling using VAE latent variables as input results in modelling errors that are well beyond the expected noise level in the data. By separating the parametrization used for inversion and the one used as input for surrogate modelling, we can circumvent this problem and perform MCMC in a latent space defined by deep generative models while surrogate modelling is approximated by PCEs operating on globally or locally defined principal components. We find that these two approaches largely outperform surrogate modelling based on VAE input parametrizations. For the channelized structures considered, modelling errors are comparable to the typical observational errors when PCE are based on globally defined principal components and significantly lower when locally defined principal components are considered. Generally speaking, using PCE significantly reduces the computational burden of MCMC, but it can be successfully used to perform non-linear MCMC inversion only if the corresponding modelling error is comparatively low. In this manuscript we have shown how PCE based on VAE parametrizations performs poorly in MCMC inversion, whereas PCE based on globally and locally defined principal components produce results comparable or close to those obtained using full-physics forward solvers. The methods presented herein are extendable to other problems involving wave-based physics of similar complexity.

ACKNOWLEDGMENTS

We thank Associate Editor Prof Andrew Valentine (Durham University), Prof Thomas Mejer Hansen (Aarhus University), Dr Robin Thibaut (Ghent University) and the anonymous reviewers whose comments helped us improve the paper. Niklas Linde, Macarena Amaya and Shiran Levy acknowledge support by the Swiss National Science Foundation (project number: 184574).

DATA AVAILABILITY

The data underlying this paper are available upon request to the corresponding author.

REFERENCES

Alizadeh, R., Allen, J.K. & Mistree, F., 2020. Managing computational complexity using surrogate models: a critical review, *Res. Eng. Des.,* **31,** 275–298.

Annan, A.P., 2005. GPR methods for hydrogeological studies, in *Hydrogeophysics,* Part of the Water Science and Technology Library book series (WSTL, Vol. 50), pp. 185–213, eds Rubin, Y. & Hubbard, S.S., Springer.

Arcone, S.A., Lawson, D.E., Delaney, A.J., Strasser, J.C. & Strasser, J.D., 1998. Ground-penetrating radar reflection profiling of groundwater and bedrock in an area of discontinuous permafrost, *Geophysics,* **63**(5), 1573–1584.

Asher, M.J., Croke, B.F., Jakeman, A.J. & Peeters, L.J., 2015. A review of surrogate models and their application to groundwater modelling, *Water Resour. Res.,* **51**(8), 5957–5973.

Aster, R.C., Borchers, B. & Thurber, C.H., 2018. *Parameter Estimation and Inverse Problems,* Elsevier.

Bejani, M.M. & Ghatee, M., 2021. A systematic review on overfitting control in shallow and deep neural networks, *Artif. Intell. Rev.,* **54,** 6391–6438.

Blatman, G. & Sudret, B., 2011. Adaptive sparse polynomial chaos expansion based on least angle regression, *J. Comput. Phys.,* **230**(6), 2345–2367.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178**(3), 1411–1436.

Boutsidis, C., Mahoney, M.W. & Drineas, P., 2008. Unsupervised feature selection for principal components analysis, in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 61–69.

Brunetti, C., Bianchi, M., Pirot, G. & Linde, N., 2019. Hydrogeological model selection among complex spatial priors, *Water Resour. Res.,* **55**(8), 6729–6753.

Chipman, H., George, E.I., McCulloch, R.E., Clyde, M., Foster, D.P. & Stine, R.A., 2001. The practical implementation of Bayesian model selection, *Lecture Notes-Monogr. Ser.,* **38,** 65–134.

Duane, S., Kennedy, A.D., Pendleton, B.J. & Roweth, D., 1987. Hybrid Monte Carlo, *Phys. Lett. B,* **195**(2), 216–222.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D. & Nicolson, H., 2017. Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry, *Geophys. J. Int.,* **208**(1), 36–58.

Gelman, A. & Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences, *Stat. Sci.,* **7**(4), 457–472.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments, *Bayesian Stat.,* **4,** 641–649.

Giannakis, I., Giannopoulos, A., Warren, C. & Sofroniou, A., 2021. Fractal-constrained crosshole/borehole-to-surface full-waveform inversion for hydrogeological applications using ground-penetrating radar, *IEEE Trans. Geosci. Remote Sens.,* **60,** doi:10.1109/TGRS.2021.3054173.

Gloaguen, E., Marcotte, D., Chouteau, M. & Perroud, H., 2005. Borehole radar velocity inversion using cokriging and cosimulation, *J. Appl. Geophys.,* **57**(4), 242–259.

Gneiting, T. & Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.,* **102**(477), 359–378.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y., 2020. Generative adversarial networks, *Commun. ACM,* **63**(11), 139–144.

Goodman, J. & Weare, J., 2010. Ensemble samplers with affine invariance, *Commun. Appl. Math. Comput. Sci.,* **5**(1), 65–80.

Haario, H., Saksman, E. & Tamminen, J., 2001. An adaptive Metropolis algorithm, *Bernoulli,* **7**(2), 223–242.

Hansen, T.M., Cordua, K.S., Jacobsen, B.H. & Mosegaard, K., 2014. Accounting for imperfect forward modeling in geophysical inverse problems—exemplified for crosshole tomography, *Geophysics,* **79**(3), H1–H21.

Hansen, T.M., Cordua, K.S., Looms, M.C. & Mosegaard, K., 2013. Sippi: A matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 2—application to crosshole GPR tomography, *Comput. Geosci.,* **52,** 481–492.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika,* **57**(1), 97–109.
Higdon, D., McDonnell, J.D., Schunck, N., Sarich, J. & Wild, S.M., 2015. A Bayesian approach for parameter estimation and prediction using a computationally intensive model, *J. Phys. G: Nucl. Part. Phys.,* **42**(3), doi:10.1088/0954–3899/42/3/034009.
Ho, J., Jain, A. & Abbeel, P., 2020. Denoising diffusion probabilistic models, *Adv. Neural Inform. Process. Syst.,* **33,** 6840–6851.
Husen, S. & Kissling, E., 2001. Local earthquake tomography between rays and waves: fat ray tomography, *Phys. Earth planet. Inter.,* **123**(2–4), 127–147.
Irving, J. & Knight, R., 2006. Numerical modeling of ground-penetrating radar in 2-D using MATLAB, *Comput. Geosci.,* **32**(9), 1247–1258.
Jensen, J.M., Jacobsen, B.H. & Christensen-Dalsgaard, J., 2000. Sensitivity kernels for time-distance inversion, *Solar Phys.,* **192**(1), 231–239.
Jetchev, N., Bergmann, U. & Vollgraf, R., 2016. *Texture synthesis with spatial generative adversarial networks,* preprint (arXiv:1611.08207).
Jolliffe, I.T. & Cadima, J., 2016. Principal component analysis: a review and recent developments, *Phil. Trans. R. Soc., A,* **374**(2065), doi:10.1098/rsta.2015.0202.
Kingma, D.P. & Welling, M., 2013. Auto-encoding variational bayes, preprint (arXiv:1312.6114).
LaBrecque, D., Alumbaugh, D.L., Yang, X., Paprocki, L. & Brainard, J., 2002. Three-dimensional monitoring of vadose zone infiltration using electrical resistivity tomography and cross-borehole ground-penetrating radar, in *Methods in Geochemistry and Geophysics,* Vol. **35,** pp. 259–272, Elsevier.
Laloy, E., Hérault, R., Jacques, D. & Linde, N., 2018. Training-image based geostatistical inversion using a spatial generative adversarial neural network, *Water Resour. Res.,* **54**(1), 381–406.
Laloy, E., Hérault, R., Lee, J., Jacques, D. & Linde, N., 2017. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network, *Adv. Water Resour.,* **110,** 387–405.
Lataniotis, C., Marelli, S. & Sudret, B., 2020. Extending classical surrogate modeling to high dimensions through supervised dimensionality reduction: a data-driven approach, *Int. J. Uncertain. Quantif.,* **10**(1), 55–82.
Levy, S., Hunziker, J., Laloy, E., Irving, J. & Linde, N., 2022. Using deep generative neural networks to account for model errors in Markov chain monte carlo inversion, *Geophys. J. Int.,* **228**(2), 1098–1118.
Levy, S., Laloy, E. & Linde, N., 2023. Variational Bayesian inference with complex geostatistical priors using inverse autoregressive flows, *Comput. Geosci.,* **171**(C), doi:10.1016/j.cageo.2022.105263.
Lopez-Alvis, J., Laloy, E., Nguyen, F. & Hermans, T., 2021. Deep generative models in inversion: the impact of the generator's nonlinearity and development of a new approach based on a variational autoencoder, *Comput. Geosci.,* **152,** doi:10.1016/j.cageo.2021.104762.
Lüthen, N., Marelli, S. & Sudret, B., 2021. Sparse polynomial chaos expansions: literature survey and benchmark, *SIAM/ASA J. Uncertain. Quant.,* **9**(2), 593–649.
Marelli, S., Lüthen, N. & Sudret, B., 2021. UQLab user manual – polynomial chaos expansions, Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, Report # UQLab-V1.4-104.
Marelli, S. & Sudret, B., 2014. UQLab: a framework for uncertainty quantification in Matlab, in *Proceedings of the 2nd International Conference on Vulnerability and Risk Analysis and Management (ICVRAM 2014),* Liverpool, United Kingdom, 13–16 July 2014, pp. 2554–2563.
Mariethoz, G. & Caers, J., 2014. *Multiple-Point Geostatistics: Stochastic Modeling With Training Images,* John Wiley & Sons.
Marzouk, Y. & Xiu, D., 2009. A stochastic collocation approach to Bayesian inference in inverse problems, *Commun. Comput. Phys.,* **6**(4), 826–847.
Marzouk, Y.M., Najm, H.N. & Rahn, L.A., 2007. Stochastic spectral methods for efficient Bayesian solution of inverse problems, *J. Comput. Phys.,* **224**(2), 560–586.
Meles, G.A., Linde, N. & Marelli, S., 2022. Bayesian tomography with prior-knowledge-based parametrization and surrogate modeling, *Geophys. J. Int.,* **231**(1), 673–691.
Métivier, D., Vuffray, M. & Misra, S., 2020. Efficient polynomial chaos expansion for uncertainty quantification in power systems, *Electr. Power Syst. Res.,* **189,** doi:10.1016/j.epsr.2020.106791.
Nagel, J.B., 2019. *Bayesian Techniques for Inverse Uncertainty Quantification,* pp. 504, IBK Bericht.
Rasmussen, C.E., 2003. Gaussian processes in machine learning, in *Summer School on Machine Learning,* pp. 63–71, Springer.
Reynolds, A.C., He, N., Chu, L. & Oliver, D.S., 1996. Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data, *SPE J.,* **1**(04), 413–426.
Sacks, J., Schiller, S.B. & Welch, W.J., 1989. Designs for computer experiments, *Technometrics,* **31**(1), 41–47.
Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics, *Math. Geol.,* **34**(1), 1–21.
Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* SIAM.
Thibaut, R., Laloy, E. & Hermans, T., 2021. A new framework for experimental design using bayesian evidential learning: the case of wellhead protection area, *J. Hydrol.,* **603,** doi:10.1016/j.jhydrol.2021.126903.
Torre, E., Marelli, S., Embrechts, P. & Sudret, B., 2019. Data-driven polynomial chaos expansion for machine learning regression, *J. Comput. Phys.,* **388,** 601–623.
Vrugt, J.A., Ter Braak, C., Diks, C., Robinson, B.A., Hyman, J.M. & Higdon, D., 2009. Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlin. Sci. Numer. Simulat.,* **10**(3), 273–290.
Wagner, P.-R., Fahrni, R., Klippel, M., Frangi, A. & Sudret, B., 2020. Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels, *Eng. Struct.,* **205,** doi:10.1016/j.engstruct.2019.110063.
Wagner, P.-R., Marelli, S. & Sudret, B., 2021a. Bayesian model inversion using stochastic spectral embedding, *J. Comput. Phys.,* **436,** doi:10.1016/j.jcp.2021.110141.
Wagner, P.-R., Nagel, J., Marelli, S. & Sudret, B., 2021b. UQLab user manual – Bayesian inversion for model calibration and validation, Technical report, Report UQLab-V1.4-113, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland.
Xiu, D. & Karniadakis, G.E., 2002. The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.,* **24**(2), 619–644.
Zahner, T., Lochbühler, T., Mariethoz, G. & Linde, N., 2016. Image synthesis with graph cuts: a fast model proposal mechanism in probabilistic inversion, *Geophys. J. Int.,* **204**(2), 1179–1190.

APPENDIX: OVERVIEW OF THE LIMITATIONS OF THE VAE–PCE APPROACH

For the configuration considered in this manuscript, PCEs based on VAE parameters provide poor accuracy in predicting traveltimes. When calculated on a representative validation set, an aggregate rmse of 2.01 ns is observed for the misfit between reference and predicted data. For the velocity distribution in Fig. 10, the traveltime prediction is particularly poor, with an rmse of 3.1 ns. In Fig. A1, we consider six additional reference velocity fields and the corresponding posterior mean images for MCMC inversion when using VAE–PCE surrogate modelling. In some cases the posterior mean resembles the reference velocity field well [compare Fig. A1(a) to A1(g), or Fig. A1(f) to A1(l)]. However, large differences can arise between the reference and the VAE-PCE posterior mean [e.g. compare Fig. A1(b) to A1(h), or Fig. A1(e) to A1(k)]. Even if the corresponding modelling error is accounted for in the inversion implying that the posterior mean models should be unbiased, we find that the modelling error has severe impacts by increasing the posterior model uncertainty.

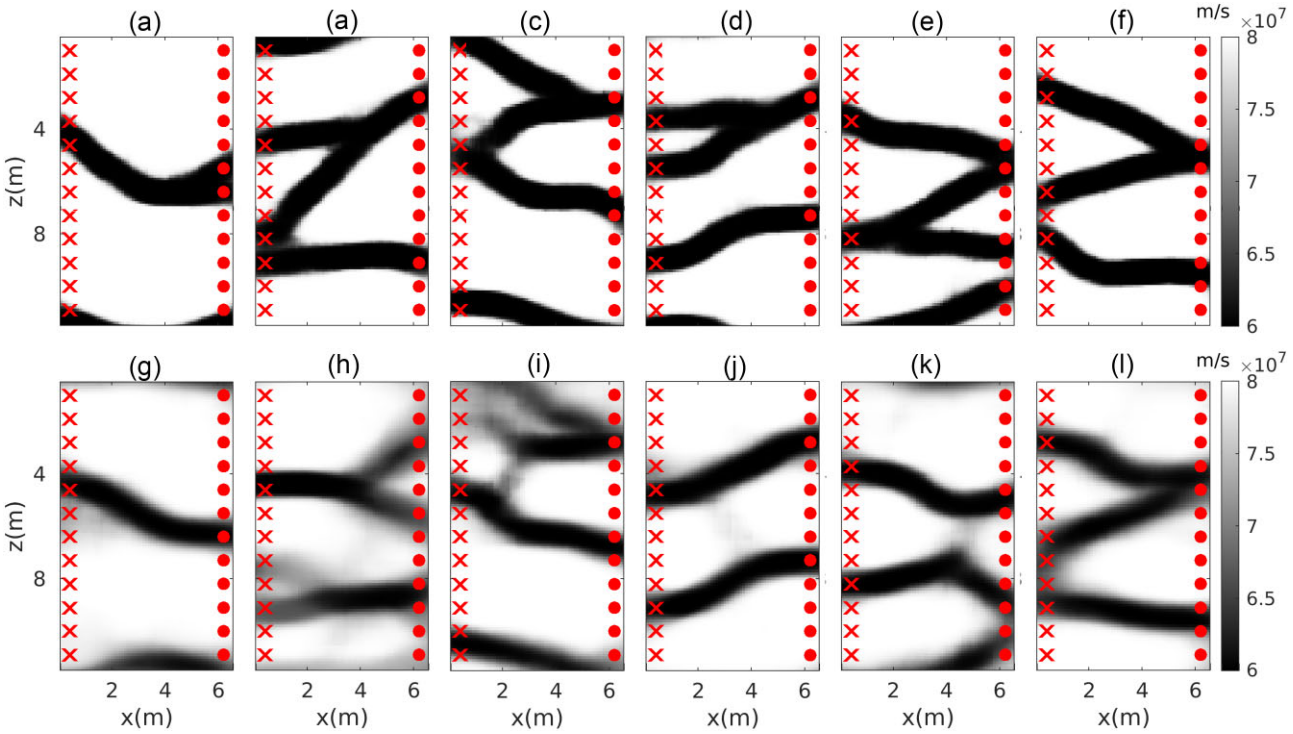

Figure A1. Reference velocity fields (a)–(f) and (g)–(l) the corresponding posterior mean images for MCMC inversion based on VAE-PCE surrogate modelling.