Geophysical Journal International

Geophys. J. Int. (2025) **243,** 1–18 https://doi.org/10.1093/gji/ggaf349
Advance Access publication 2025 September 05
Research Paper

Bayesian full waveform inversion with sequential surrogate model refinement

Giovanni Angelo Meles,[1,2] Stefano Marelli[3] and Niklas Linde[2]

[1] *Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek, TNO, 2597 AK Den Haag, the Netherlands. E-mail: giovanni.meles@tno.nl*
[2] *Institute of Earth Sciences, University of Lausanne, 1015, Lausanne, Switzerland*
[3] *Institute of Structural Engineering, ETH Zurich, 8093 Zürich, Switzerland*

Accepted 2025 September 1. Received 2025 July 11; in original form 2025 May 7

SUMMARY

Bayesian formulations of inverse problems are attractive due to their ability to incorporate prior knowledge, account for various sources of uncertainties and update probabilistic models as new information becomes available. Markov chain Monte Carlo (MCMC) methods sample posterior probability density functions (PDFs) provided accurate representations of prior information and many evaluations of likelihood functions. Dimensionality-reduction techniques such as principal component analysis (PCA) can assist in defining the prior PDF and the input bases can be used to train surrogate models. Surrogate models offer efficient approximations of likelihood functions that can replace traditional and costly forward solvers in MCMC inversions. Many problem classes in geophysics involve intricate input/output relationships that conventional surrogate models, constructed using samples drawn from the prior PDF fail to capture, leading to biased inversion results and poor uncertainty quantification. Incorporating samples from regions of high posterior probability in the training may increase accuracy, but identifying these regions is challenging. In the context of full waveform inversion, we identify and explore high-probability posterior regions using a series of successively trained surrogate models covering progressively expanding wave bandwidths. The initial surrogate model is used to invert low-frequency data only as the input/output relationship of high-frequency data are too complex to be described across the full prior PDF with a single surrogate model. After a first MCMC inversion, we retrain the surrogate model on samples from the resulting posterior PDF and repeat the process. By focusing on progressively narrower input domain regions, it is possible to progressively increase the frequency bandwidth of the data to be modelled while also decreasing model errors. Through this iterative scheme, we eventually obtain a surrogate model that is of high accuracy for model realizations exhibiting significant posterior probabilities across the full bandwidth of interest. This surrogate model is then used to perform an MCMC inversion yielding the final estimation of the posterior PDF. Numerical results from 2-D synthetic cross-hole Ground Penetrating Radar (GPR) examples demonstrate that our method outperforms ray-based approaches, as well as results obtained when only training the surrogate model using samples from the prior PDF. Our methodology reduces the overall computational cost by approximately two orders of magnitude compared to using a classical finite-difference time-domain forward scheme.

Key words: Ground penetrating radar; Bayesian inference; Numerical modelling.

1 INTRODUCTION

Tomography encompasses a diverse array of non-invasive imaging methods that are widely used across numerous application fields, such as exploration for natural resources (Taillandier *et al.* 2009), investigating moons and other celestial bodies (Zhao *et al.* 2008; Khan *et al.* 2016), diagnosing medical conditions (Stotzka *et al.* 2002) and conducting non-destructive testing (Tant *et al.* 2018). Full Waveform Inversion (FWI) stands out given its potential to maximize information retrieval across multiple scales. Advancements in computational capabilities have facilitated FWI developments (Dhabaria & Singh 2024), leading to its growing use for characterizing geological structures, assessing natural resources and mitigating environmental risks (Ernst *et al.* 2007a; Mulder & Plessix 2008; Virieux

& Operto 2009; Fichtner *et al.* 2013). Deterministic FWI can generate images that reveal intricate geological features and resolve material properties at sub-wavelength scale. However, the associated algorithms are prone to a number of limitations, particularly with regard to their limited capabilities for uncertainty quantification and the high risk of converging to a local minimum (Virieux *et al.* 2017; Klotzsche *et al.* 2019). In contrast, Bayesian inversion methods can avoid these issues by treating model parameters as random variables (Liu & Liu 2001; Ulrych *et al.* 2001). Nevertheless, the substantial computational demand stemming from the need to perform many forward waveform simulations can be exceedingly demanding. Various strategies have been employed to enhance convergence and mitigate computational expenses of Bayesian FWI. These strategies include advanced Markov chain Monte Carlo (MCMC) schemes inspired by evolutionary concepts, Hamiltonian Monte Carlo (HMC) and Variational Bayesian inference (VBI) (Hunziker *et al.* 2019; Gebraad *et al.* 2020; Aleardi 2020; Zhang *et al.* 2023).

An increasingly common set of tools to reduce computational costs in uncertainty quantification is surrogate modelling, such as Kriging (Sacks *et al.* 1989), polynomial chaos expansions (PCEs) (Xiu & Karniadakis 2002; Lüthen *et al.* 2022; Meles *et al.* 2022) or deep neural networks (DNN) methods (Adler & Öktem 2024; Wang *et al.* 2018; Wu *et al.* 2018; Yang & Ma 2019; Sun *et al.* 2023). Algorithms using DNN architectures that directly map waveform data to physical property distributions have also been proposed. However, training a well-generalized and robust DNN for this task remains computationally expensive (Liu *et al.* 2021). Developing accurate surrogate models for physical systems often requires reducing the dimensionality or simplifying the representation of input variables to effectively approximate complex system behaviours. Successful surrogate modelling will enhance computational feasibility while preserving a high level of accuracy in input representation and output prediction. Various approaches exists to achieve effective dimensionality reduction, with principal component analysis (PCA) being the most common (Reynolds *et al.* 1996; Boutsidis *et al.* 2008; Jolliffe & Cadima 2016; Thibaut *et al.* 2021; Giannakis *et al.* 2021; Meles *et al.* 2022).

Recently, Meles *et al.* (2022) performed Bayesian inversion with a PCA-based parametrization of prior information to predict travel times with PCE. Extending this strategy to full-waveform inversion is non-trivial due to more intricate input-output relationships compared to traveltime modelling (Meles *et al.* 2024) and the limited expressiveness provided by PCE. This limitation arises not from theoretical arguments but from the truncation schemes that are needed to limit the number of polynomial coefficients to be evaluated.

To address the Bayesian FWI problem, we introduce a sequential formulation (Peherstorfer *et al.* 2018) involving surrogate models that progressively cover expanding data bandwidths, trained on progressively smaller manifolds. This approach leverages heuristic arguments about input–output relationships in wave phenomena, focusing initially on constructing effective surrogate models for low-frequency data components over the whole prior probability density function (PDF). By iteratively refining model training within narrower input regions (i.e. drawn from intermediate posteriors obtained by inverting lower frequency data), our method aims to achieve accurate modelling across the full frequency bandwidth of interest over the support of the posterior PDF to enable accurate and efficient inversion of all data.

2 METHODOLOGY

2.1 Bayesian inversion via surrogate modelling

Numerical forward models are used to predict the outcomes of physical experiments for given values of input parameters. They describe the relationship between input parameters and the output as:

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

Here, $\mathcal{F}$ represents the physical law or forward operator, typically involving partial differential equations (PDEs) acting on locally defined parameters. The variable $\boldsymbol{u}$ denotes the input parameters, $\boldsymbol{y}$ the data and ϵ is a noise term. The inverse problem seeks to infer properties of $\boldsymbol{u}$ given the data $\boldsymbol{y}$ with a prescribed noise model of the data acquisition process while considering any available prior information about $\boldsymbol{u}$.

For a constant model input dimension, a general formulation of this problem can be expressed through the unnormalized posterior PDF:

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

In this context, $P(\boldsymbol{u}|\boldsymbol{y})$ represents the posterior PDF of the input $\boldsymbol{u}$ given the n-dimensional data $\boldsymbol{y}$. The probability $P(\boldsymbol{y}|\boldsymbol{u})$, also denoted as $L(\boldsymbol{u})$ and referred to as 'the likelihood,' is the probability of observing the data $\boldsymbol{y}$ given the input parameter values $\boldsymbol{u}$. Finally, $P(\boldsymbol{u})$ is the prior PDF of the input parameters. In this paper, we follow standard formalism in geophysics and use the same symbol (e.g. $\boldsymbol{u}$) to denote both individual instances of a random variable and the random variable itself (Tarantola 2005; Aster *et al.* 2018).

For zero-mean Gaussian observational noise and a perfect forward operator, the likelihood takes the following form:

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

where the data covariance matrix $\boldsymbol{C}_d$ quantifies data uncertainty. To characterize the unnormalized posterior PDF in eq. (2), it is common to use MCMC methods to draw samples proportionally from $P(\boldsymbol{u}|\boldsymbol{y})$ (Gelman *et al.* 1995).

Computing the likelihood $L(\boldsymbol{u})$ typically involves using physics-based solvers, which can be computationally demanding for Bayesian inversion that rely on very large numbers of likelihood computations. In recent years, surrogate modelling has gained popularity as an alternative to physics-based solvers, even if its use in geophysics has been limited until recently (Linde *et al.* 2017). Surrogate modelling offers an approximate and computationally inexpensive representation of the forward model. The coordinate system used to express the forward model in eq. (1) may not be suitable for implementing a surrogate model, particularly if the dimensionality of the input $\boldsymbol{u}$ is high. In many cases, the forward model may nonetheless be effectively represented in a lower dimensional manifold, enabling a proper surrogate formulation.

To achieve this, one needs to express the forward problem in terms of a new set of variables. Among various approaches, PCA is particularly effective for this task, as it optimally captures input variability through a linear representation, enabling accurate low-dimensional approximations. The input $\boldsymbol{u}$ in eq. (1) can be uniquely represented as $\boldsymbol{x}_{\text{full}}$ by projecting it onto a complete, possibly high-dimensional, set of principal components, which preserves the input–output relationship without introducing any approximation.

$$\mathcal{F}(\boldsymbol{u}(\boldsymbol{x}_{\text{full}})) = \mathcal{M}(\boldsymbol{x}_{\text{full}}) = \boldsymbol{y} + \epsilon, \tag{4}$$

where $\mathcal{M} = \mathcal{F} \circ \boldsymbol{u}$ with '$\circ$' indicating function composition. It is then possible to truncate the PCA expansion and retain only the first M principal components. In the following, we denote by $\boldsymbol{x}$ the M-dimensional approximation of $\boldsymbol{x}_{\text{full}}$. The forward model can be accurately approximated using a reduced M-dimensional representation $\boldsymbol{x}$, whenever the discrepancy between the model output computed on $\boldsymbol{x}_{\text{full}}$ and on its truncated counterpart $\boldsymbol{x}$ is negligible. When $\mathcal{M}(\boldsymbol{x}_{\text{full}}) \approx \mathcal{M}(\boldsymbol{x})$, we can then write:

$$\mathcal{M}(\boldsymbol{x}) = \boldsymbol{y} + \hat{\epsilon}, \tag{5}$$

where $\hat{\epsilon}$ encompasses both observational noise and modelling errors due to the truncation with the truncation level M dictating the accuracy in representing the input domain. The required truncation needed to achieve the desired accuracy in modelling depend on the forward problem. This is because nonlinear modelling non-trivially propagates the inaccuracies from the input to the output, making it impossible to determine M *a priori*. Thus, different forward models may require different levels M of PCA truncation to ensure satisfactory results. Operating on an effective set of coordinate $\boldsymbol{x}$ casts the inverse problem on the M-dimensional manifold as:

$$P(\boldsymbol{x}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x}) = L(\boldsymbol{x})P(\boldsymbol{x}). \tag{6}$$

Once the forward problem and Bayes' theorem are expressed in terms of a low-dimensional input, surrogate modelling can be used for MCMC (Chib & Greenberg 1995) scheme, with the reference solver $\mathcal{M}(\boldsymbol{x}_{\text{full}})$ being replaced by a surrogate $\tilde{\mathcal{M}}(\boldsymbol{x})$:

$$\tilde{\mathcal{M}}(\boldsymbol{x}) \approx \mathcal{M}(\boldsymbol{x}_{\text{full}}). \tag{7}$$

For likelihood evaluation in MCMC algorithms, surrogate models should be employed with a modified covariance operator $\boldsymbol{C}_D = \boldsymbol{C}_d + \boldsymbol{C}_{\text{Tapp}}$ with $\boldsymbol{C}_{\text{Tapp}}$ accounting for the impact of truncation and surrogate modelling errors:

$$L(\tilde{\mathcal{M}}(\boldsymbol{x})) = \left(\frac{1}{2\pi}\right)^{n/2} |\boldsymbol{C_D}|^{-1/2} \exp\left[-\frac{1}{2}(\tilde{\mathcal{M}}(\boldsymbol{x}) - \boldsymbol{y})^T \boldsymbol{C_D}^{-1}(\tilde{\mathcal{M}}(\boldsymbol{x}) - \boldsymbol{y})\right], \tag{8}$$

where $|\boldsymbol{C_D}|$ is the determinant of the covariance matrix $\boldsymbol{C_D}$ (Hansen *et al.* 2014). In this expression, $\boldsymbol{x} \in \mathbb{R}^M$ is an M-dimensional input, and the observed data $\boldsymbol{y} \in \mathbb{R}^n$ is an n-dimensional output.

Since the computational cost of evaluating a surrogate model typically scales with the output dimension, reducing the number of output variables can significantly accelerate execution. When the output exhibits strong correlations or redundancy, techniques such as PCA can be leveraged to construct a more compact yet informative representation, improving efficiency without compromising predictive capability (Meles *et al.* 2022). In the context of waveform data, PCA can, in principle, be applied to full-bandwidth time-domain signals. In the following sections, we discuss the challenges of directly applying surrogate modelling to time-domain data and propose a potential workaround that combines monochromatic filtering with PCA decomposition of the output to enable an effective implementation.

2.2 Surrogate modelling of wave-based phenomena

Common surrogate modelling approaches such as Kriging (Sacks *et al.* 1989) or polynomial chaos expansions (Xiu & Karniadakis 2002), are typically designed to approximate scalar-output models. The simplest extension to models with multivariate outputs is based on separately modelling each output variable, sometimes coupled with dimensionality reduction techniques such as PCA (Blatman & Sudret 2011b). Some specific classes of surrogates, such as Kriging, also admit the explicit inclusion of covariance information, albeit sometimes at the cost of an overall reduced accuracy (Kleijnen & Mehdad 2014).

Surrogate modelling of systems with time-dependent outputs of the form $y(\boldsymbol{x}, t) = \mathcal{M}(\boldsymbol{x}, t)$, however, is generally a much more complex task. This is due to both the output vector dimension, often discretized in the order of $O(10^{2-5})$ time-steps, but mainly because of the increase in prevalence of higher order effects at later times, due to interferences, reflections, etc. We make an important distinction between two main families of dynamical systems, based entirely on the characteristics of the input excitation. We define the first class as *systems with fundamentally simple inputs*, the input of which are either not time-dependent (e.g. Green's functions), or exhibit variability that can be accurately represented with a relatively small number of scalar parameters (e.g. monochromatic waves, fixed wavelets with variable amplitude and dominant frequency, etc.). This class of systems can generally be treated with classical surrogate modelling approaches, combined with advanced dimensionality reduction techniques (Mai & Sudret 2017; Yaghoubi *et al.* 2017; Meles *et al.* 2022).

The second class is that of *systems with fundamentally complex inputs*, characterized by input excitations that are poorly compressible, that is, their variability requires a large to infinite number of parameters to be properly represented. Typical input excitations of this kind are turbulent winds and earthquakes. Approximating this class of systems is much more challenging, and it is the territory of state-space models, such as autoregressive models with exogenous inputs (NARX; Billings 2013), and their more recent incarnations (Schär *et al.* 2024).

In the case of waveform tomography, the system belongs to the first class of models. Despite the very high output dimensionality (each realization of the input model $\boldsymbol{x}$ can result in large set of output traces, typically in the order of $O(10^2)$ traces, each counting $O(10^3)$ time-steps), the set of input wavelets (one per source position) is considered constant. Therefore, even if the overall problem dimensionality can be large, due to the choice of subsurface parametrization, the time dependence of the input excitations is very low dimensional. This peculiarity of the tomographic inverse problem allows us to focus our efforts onto devising effective output dimensionality reduction strategies, while still deploying powerful strong learners from the UQ literature, rather than developing our own problem-specific surrogate model to solve both.

Surrogates typically train on samples drawn from the prior PDF. This generally works well if the prior PDF is not too wide, if there are not too many input and output dimensions and if input–output relationships are not highly nonlinear (Meles *et al.* 2022). However, the strategy by Meles *et al.* (2022) is likely to fall short when applied to nonlinear waveform modelling across prior PDFs with wide support. An alternative training strategy is to focus on the much smaller hypervolume of the parameter space exhibiting significant posterior probability densities (Cui *et al.* 2015). These regions can be located and sampled through an initial inversion using a reasonably accurate surrogate model trained on samples from the prior PDF that in turn is used for a new MCMC inversion. Such sequential approach involving surrogates of different fidelities can often produce more accurate inference than using one fidelity-level only while still maintaining important speed-ups compared with using the original computationally costly forward solver only (Peherstorfer *et al.* 2018; Amaya *et al.* 2024). For nonlinear problems, constructing such an initial surrogate model of sufficient accuracy can be challenging if considering all data. An alternative is to first restrict the inversion to the regions of the input and output space where the initially surrogate model is sufficiently accurate.

For the considered Bayesian FWI problem, we leverage knowledge about wave phenomena to guide the initial selection of data dimensions and input parameters to be used. Our strategy considers the well-established fact that low-frequency data primarily reflects low-wavenumber structures in the velocity distribution while higher wavenumber features, corresponding to smaller and more abrupt spatial variations, have a significant impact on higher frequency waves (Williamson 1990; Bunks *et al.* 1995; Chew & Lin 1995; Pratt *et al.* 1998; Zhou & Greenhalgh 2003; Meles *et al.* 2012). The complexity of wave propagation increases with frequency due to the shorter wavelengths involved, thereby, making them more sensitive to small-scale heterogeneities. This sensitivity leads to intricate interference patterns, multipath effects and strong attenuation and dispersion, contrasting with the more uniform and predictable propagation of low-frequency waves with longer wavelengths (Aki & Richards 2002). Consequently, we expect that low-frequency data can be modelled far more reliably with surrogate modelling than high-frequency data across the entire prior PDF, while accurate modelling of high-frequency data requires training on a more restricted subset of the input space.

In the following, we present two algorithms for waveform data inversion, each with a different emphasis on the training of surrogate models. The first approach relies exclusively on prior-based training, which we refer to as Full-Bandwidth-Prior-Training (FBPT). The second approach is based on the progressive enrichment of the training set with samples that are more representative of the posterior PDF, which we refer to as Progressively-Expanded-Posterior-Training (PEPT).

2.3 Full-bandwidth-prior-training (FBPT) methodology

In recent work, Meles *et al.* (2022) proposed a surrogate strategy for traveltime inversion based on PCA applied to both the input, permittivity distributions and the output, traveltime data. It is natural to consider a similar approach for waveform inversion. However, the transition from traveltime data to waveform gathers is not straightforward in terms of output parametrization. To ensure effective output analysis, the FBPT strategy involves monochromatic passband filtering of the waveform data before performing PCA. Through this important pre-processing step, the FBPT strategy—though applied to waveform data—closely mirrors the approach outlined in Meles *et al.* (2022). We use Π to refer to input-related PCA concepts and Λ for output-related concepts. To enhance readability and facilitate comparison between the two schemes, we will also use 'F' and 'P' to represent the quantities associated with FBPT and PEPT, respectively.

The non-iterative FBPT process involves:

(i) **Parametrization**: The first step is to define appropriate parametrizations for both the input and output domains. For the input, we consider a training set I_F with N_F input samples, and parametrize it via its first Π_F PCs. For the output we collect a sample data set O_F of waveform data associated with I_F. We then consider frequency bins f_i from a comprehensive set Ω_F capturing a relevant portion of the spectrum of the source term. In a pre-processing step, the data O_F is passed through a passband filter for each $f_i \in \Omega_F$. Such filtered data are denoted in the following as $\hat{O}_F$. Sets of $\widehat{\Lambda}_F$ PCs are derived for a number Λ_γ of time-domain minigathers, which are composed of subsets of γ adjacent common-source traces of $\hat{O}_F$. These PCs are then used to parametrize the output. The data O_F is thus transformed into the projection onto $\Lambda_{FT} = \widehat{\Lambda}_F \times \Lambda_\gamma \times |\Omega_F|$ PCs with $|\Omega_F|$ indicating the cardinality of Ω_F. Each frequency bin results in $\Lambda_F = \widehat{\Lambda}_F \times \Lambda_\gamma$ output PCs.

(ii) **Model training**: The second step is to train a surrogate model to map from input to output. We follow Meles *et al.* (2022) and rely on PCE modelling (see Appendix A for more details). The surrogate model, indicated here as PCE_F, estimate the projection on $\Lambda_F = \widehat{\Lambda}_F \times \Lambda_\gamma$ output PCs of each frequency bin of waveform data. Projection of a data set D on the chosen output PCs is indicated here as D_F.

(iii) **MCMC inversion**: In the third step, PCE_F is used as a surrogate model to perform MCMC inversion of the observed data OBS projected on the considered output PCs, indicated as $\mathrm{OBS_F}$, and estimate the corresponding posterior distribution P_F. Here, we use an Metropolis-Hastings algorithm with a Gaussian proposal distribution. The proposal scales are determined by a diagonal covariance matrix, where each diagonal element is proportional to the prior marginal variance of the corresponding input parameter, scaled by a factor indicated as μ_F, determined through trial and error, aiming for an acceptance rate of around 20 per cent (Hastings 1970; Marelli & Sudret 2014).

Table 1. FBPT symbols and quantities,

Symbol	Quantity
I_F	Input training set
N_F	Number of samples of I_F
O_F	Output associated with I_F
Π_F	Number of input PCs
f_i	i-th frequency bin
Ω_F	Set of frequency bins
$\lvert\Omega_F\rvert$	Cardinality of Ω_F
$\hat{O}_F$	Passbanded filtered O_F
Λ_γ	Number of minigathers per frequency bin
γ	Number of traces per minigather
$\widehat{\Lambda}_F$	Number of PCs per minigather
Λ_F	Number of PCs per frequency bin
Λ_{FT}	Total number of PCs
PCE_F	PCE model
OBS	Observed traces to be inverted
$\mathrm{OBS_F}$	Projection of OBS on Λ_{FT} PCs
μ_F	Proposal scale factor - k-th iteration

A list of the most relevant symbols used in the FBPT algorithm can be found in Table 1.

In the FBPT scheme, a single surrogate model is used to predict the output of a waveform simulation across the prior PDF. Given the complexity of the input–output relationship, the training set should effectively represent the full range of the prior. It can be difficult to capture the variability of the prior using N_F samples. In these cases, using N_F samples from an *inflated* prior can be advantageous. This approach helps to avoid undersampling in the peripheral regions of the prior. While it may affect the accuracy in well sampled areas, it ensures that the surrogate model is less prone to bias while still being sufficiently accurate across the entire prior. In this work, prior inflation is achieved by sampling from a Gaussian field with the same mean and correlation structure as the original prior, but with the standard deviations of each principal component increased by a factor α_F.

Algorithm 1 Full-Bandwidth-Prior-Training (FBPT) scheme

1: Define a comprehensive set of frequency bins Ω_F;
2: For a training set I_F with N_F input samples from the prior PDF, possibly inflated by the factor α_F, a corresponding data set O_F of waveform data is collected and passed through a passband filter for each frequency bin in Ω_F to obtain $\hat{O}_F$.
3: Parametrize input I_F via Π_F PCs;
4: Parametrize $\hat{O}_F$, represented by Λ_γ minigathers consisting of γ adjacent receiver traces, via $\Lambda_{FT} = \Lambda_F \times |\Omega_F|$ PCs with $|\Omega_F|$ indicating the cardinality of Ω_F and $\Lambda_F = \widehat{\Lambda}_F \times \Lambda_\gamma$, where $\widehat{\Lambda}_F$ is the number of PCs per minigather.
5: Train PCE$_F$ using projections of $I_F, \hat{O}_F$ on Π_F/Λ_{FT} input/output PCs. Using frequency depending output PCs results in the projection of a generic data set D on a reduced data set D_F;
6: Use PCE$_F$ as a Π_F dimensional surrogate model to invert the Λ_F dimensional data set OBS$_F$, that is the projection of observed gather OBS and estimate posterior P_F.

Note that each output component—corresponding to a specific frequency and a specific minigather—requires its own surrogate model. The term *PCE model* thus refers to the collection of all these individual output models.

2.4 Progressively-expanded-posterior-training (PEPT) methodology

As an alternative to FBPT, the PEPT scheme adopts a sequential model refinement approach. It starts by constructing a PCE surrogate model that inverts only the low-frequency content of the data to constrain low-wavelength components in the input space. This surrogate model is then used to perform an initial MCMC inversion. The resulting posterior samples are employed to refine the surrogate model, assuming that the target posterior lies within the initial posterior. This assumption is reasonable because PCE is unbiased over its training set, only low-frequency data are used and the inclusion of model errors in the likelihood function leads to a more dispersed posterior in the first inversion due to the relatively large model errors. Using samples from this more constrained region of the input domain, we then retrain the surrogate model with an expanded frequency bandwidth. Multiple iterations are carried out to achieve a final surrogate model that adequately simulate each data component of interest in the posterior region. Once the desired level of accuracy is attained, the resulting final surrogate model is employed to conduct a final MCMC inversion to approximate the posterior of interest. This iterative PEPT strategy demands additional considerations.

(i) **Parametrization**: As for FBPT, the first step is to define an appropriate parametrization for the input and output domains. To ensure a fair comparison, the same input parametrization will be applied to both the FBPT and PEPT strategies and the same number of output PCs, given a frequency bin, will be used for both methods.

After this preliminary step, a process follows that is repeated K times:

(i) **Subset selection**: In each iteration $k, 1 \leq k \leq K$, different subsets of the input and output domains are utilized. This is achieved by progressively increasing the number of input principal components ($\Pi_{P_1}, \ldots, \Pi_{P_K}$) and frequency bins ($\Omega_1, \ldots, \Omega_K$) on which the data is projected in the pre-processing step to obtain $\hat{O}_k$. Note that throughout the entire process, PEPT uses the same input PCs as FBPT with the number of PCs used increasing progressively. As in FBPT, the total number of minigathers (Λ_γ) and the corresponding output principal components (Λ_P) can also be adjusted. In this paper, to ensure a fair comparison between the two methods, we apply the same choices for both across all iterations of PEPT; however, this is by no means a requirement. Given the number of minigathers (Λ_γ) and the corresponding PCs ($\widehat{\Lambda}_P$), the total number of output PCs per frequency bin is $\Lambda_P = \widehat{\Lambda}_P \times \Lambda_\gamma$, and the total number of output PCs $\Lambda_{PT_k} = \Lambda_P \times |\Omega_k|$, where $|\Omega_k|$ indicates the cardinality of Ω_k, that is, the number of frequency bins. In terms of dimensionality, the input and output domains used for inversion in PBET will fully align with those in FBPT during the final iteration, where $\Pi_{P_n} = \Pi_F$ and $\Omega_n = \Omega_F$ and therefore $\Lambda_{FT} = \Lambda_{PT_K}$. At this final stage, both FBPT and PBET utilize the same set of input PCs. However, despite having the same dimensionality, the output PCs will differ between the two methods, as for FBPT they are defined based on samples drawn from the prior, while for the PEPT they are based on samples drawn from progressively refined posteriors. Updating the output PCs is key to ensure optimal representation of the relevant data variability in high-posterior regions.

(ii) **Model training**: In each iteration k, $1 \leq k \leq K$, a surrogate model is trained to map the relevant input and output subdomains. The training set needs to be large enough to achieve sufficient accuracy while not too large to keep computational costs low. Initially, samples are drawn from $I_1, \hat{O}_1$. In subsequent iterations, training focuses primarily on the progressively refined, narrower subdomains explored in the previous MCMC inversion detailed below. The surrogate model, indicated here as PCE$_k$, estimate the projection of a data set on $\Lambda_P = \widehat{\Lambda}_P \times \Lambda_\gamma$ output PCs per each considered frequency bin of waveform data. Although PEPT uses the same input PCs as FBPT, the marginal distributions used during training are updated based on the progressively refined posteriors.

(iii) **MCMC inversion**: PCE$_k$ is used as a surrogate model to perform MCMC inversion of the observed data OBS projected on the considered output PCs, indicated as OBS$_k$, and estimate the corresponding posterior distribution P_k. To enhance the efficiency and convergence properties of the MCMC algorithm, we adjust the proposal distribution at each PBET iteration based on the correlations among the different input components, as inferred from the previous posterior sample. Unlike the adaptive scheme by Haario *et al.* (2001), where the proposal distribution is dynamically adjusted as the chains progress, we modify the proposal distribution only during the PEPT iterations, relying instead on a standard MH scheme. The proposal scales are determined by a diagonal covariance matrix, where each diagonal element is proportional to the prior marginal variance of the corresponding input parameter, scaled by a factor indicated as μ_{Pk} determined through trial and error, aiming for an acceptance rate of around 20 per cent. Once established in preliminary tests, this scaling is applied to each inversion run discussed here. Given this adaptation of the proposal scheme, the covariance structure is shared between the PEPT and FBPT schemes only at the initial iteration of the PEPT. At the final iteration of PEPT, the same data as in FBPT are inverted, but the tuning of their respective MCMC schemes may differ significantly. To minimize burn-in, it is possible to initialize the chains using draws from MCMC runs of previous iterations. For instance, the first iteration produces a

Table 2. PEPT symbols and quantities.

Symbol	Quantity
k	PEPT Iteration number
I_k	Input training set - k-th iteration
N_k	Number of samples of I_k
O_k	Output associated with I_k
Π_k	Number of input PCs
f_i	i-th frequency bin
Ω_k	Set of frequency bins - k-th iteration
$\lvert\Omega_k\rvert$	Cardinality of Ω_k
$\hat{O}_k$	Passbanded filtered O_k
Λ_γ	Number of minigathers per frequency bin
γ	Number of traces per minigather
$\widehat{\Lambda}_k$	Number of PCs per minigather
Λ_P	Number of PCs per frequency bin
Λ_{Pk}	Total number of PCs - k-th iteration
PCE$_k$	PCE model - k-th iteration
OBS$_k$	Projection of OBS on Λ_{Pk} PCs
μ_k	Proposal scale factor - k-th iteration
P_k	Posterior - k-th iteration
ν_k	Samples extracted from P_k
ρ_p	Samples kept from P_p

posterior for the first Π_{P1} PCs. The final values from this step can serve as the starting point for the first iteration, while the remaining Π_{P2}-Π_{P1} PCs starting values are drawn from the prior.

(iv) **Training set enrichment**: If $k \leq K$, it is necessary to enhance the training set by drawing additional samples from the MCMC posterior to improve the model's ability to describe broader frequency bandwidths of the data. This involves collecting ν_k samples from P_k, adding components from the *null space*, defined as the prior subdomain not considered in the modelling/inversion process (Meles *et al.* 2022), and transforming them back into the physical distribution domain. These new permittivity distributions are then included in I_{k+1}, potentially alongside samples ρ_p, $p \leq k$ from previous iterations and used to compute O_{k+1}. Note that including samples from earlier iterations—associated with broader posteriors may reduce PCE accuracy in the regions of primary interest. Therefore, their use must be carefully evaluated.

A list of the most relevant symbols used in the PEPT algorithm, along with their meanings, can be found in Table 2.

As in FBPT, an inflation factor α_P can be applied during training to mitigate excessive inaccuracies in surrogate modelling performance.

Due to the orthogonality of PCA and Fourier decomposition, both the FBPT and the PEPT approaches ensure that orthonormality between corresponding basis vectors is maintained when considering mutually separated bandwidths. Consequently, the data error and the likelihood structure in the original gather domain are preserved in these specific output quantities (see appendix in Meles *et al.* 2022). Note that reducing the number of PCs in the input space alters the covariance matrix in the likelihood term, whereas restricting the number of PCs in the output domain does not affect the likelihood structure.

3 APPLICATION TO GPR CROSS-HOLE TOMOGRAPHY

We apply the proposed FBPT and PEPT methodologies to a GPR cross-hole tomography problem (Ernst *et al.* 2007b; Kuroda *et al.* 2007; Meles *et al.* 2010). We consider for simplicity variations in

Algorithm 2 Progressively-Expanded-Posterior-Training (PEPT) scheme

Define expanding sets of frequency bins (Ω_1, ... Ω_K);
For a sample I_1 with N_1 input samples (typically with $N_1 \ll N_F$) from the prior PDF, possibly inflated by a factor α_P, collect a dataset O_1 of waveform data and and passed through a passband filter for each frequency bin in Ω_1 to get $\hat{O}_1$.;
Parametrize input I_1 via Π_{P_1} PCs ;
Parametrize $\hat{O}_1$, represented by Λ_γ minigathers consisting of γ adjacent receiver traces, via $\Lambda_{P1} = \Lambda_P \times |\Omega_1|$ PCs with $|\Omega_1|$ indicating the cardinality of Ω_1 and $\Lambda_P = \widehat{\Lambda}_P \times \Lambda_\gamma$, where $\widehat{\Lambda}_P$ is the number of PCs per minigather.
for $k = 1$ **to** K **do**
  Train PCE$_k$ using projections of I_k, $\hat{O}_k$ on Π_{Pk}, Λ_{Pk} input/output PCs. Using frequency depending output PCs associated with f_i, $i \in \Omega_k$ results in the projection of a generic dataset D on a reduced dataset D_k;
  Use PCE$_k$ to invert OBS$_k$, that is, the projection of observed gather OBS and estimate posterior P_k;
  Collect ν_k samples from P_k, add elements from the *null space* and transform back into the physical distribution domain and include them in I_{k+1}.
  Based on empirical accuracy and convergence considerations, samples ρ_p from I_p, $p \leq k$ can also be included in I_{k+1};
  Given I_{k+1} compute with the reference forward solver simulations corresponding output O_{k+1} and pass it through a passband filter for each frequency bin in Ω_{k+1} to get $\hat{O}_{k+1}$;
  Parametrize input I_{k+1} via Π_{Pk+1} PCs;
  Parametrize $\hat{O}_{k+1}$, represented by Λ_γ minigathers consisting of γ adjacent receiver traces, via $\Lambda_{Pk+1} = \Lambda_P \times |\Omega_{k+1}|$ PCs with $|\Omega_{k+1}|$ indicating the cardinality of Ω_{k+1} and $\Lambda_P = \widehat{\Lambda}_P \times \Lambda_\gamma$, where $\widehat{\Lambda}_P$ is the number of PCs per minigather.
end for

relative permittivity, while assuming a constant and known electrical conductivity. We use the recording configuration displayed in Fig. 1(a) with nine 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 0.56 m. Each source is characterized by a 100 MHz Blackman–Harris pulse that is initiated separately and the propagating signals are recorded at all receivers. We employ a 2-D FDTD solver to simulate noise-free propagation in the transverse-electrical (TE) mode (Irving & Knight 2006) with perfectly matched layers. We rely on a generative model representing prior knowledge to create a set of relative permittivity distributions ϵ_r typical of a near-surface geophysical setting (Hunziker *et al.* 2017). We consider 5 m wide, 5 m deep random multi-Gaussian field realizations created using a predefined covariance structure based on the 2-D Matérn geostatistical model (Dietrich & Newsam 1997; Laloy *et al.* 2015). The models are generated using six parameters: mean (14), standard deviation of ϵ_r (3), anisotropy angle (90), anisotropy ratio (0.5), integral scale of the major axis (10 m) and a shape parameter (1). Figs 1(a), (c), (e) display representative samples of ϵ_r-distributions, together with the corresponding gathers in Figs 1(b), (d), (e). Given the permittivity range and the frequency content of the source wavelet, a spatial grid with dx = dz = 4 cm and a time stepping dt of 0.16 ns is used to avoid numerical dispersion and instability. Each generated permittivity field corresponds to a 125 × 125 dimensional space. Following Meles *et al.* (2022), we use the generative model

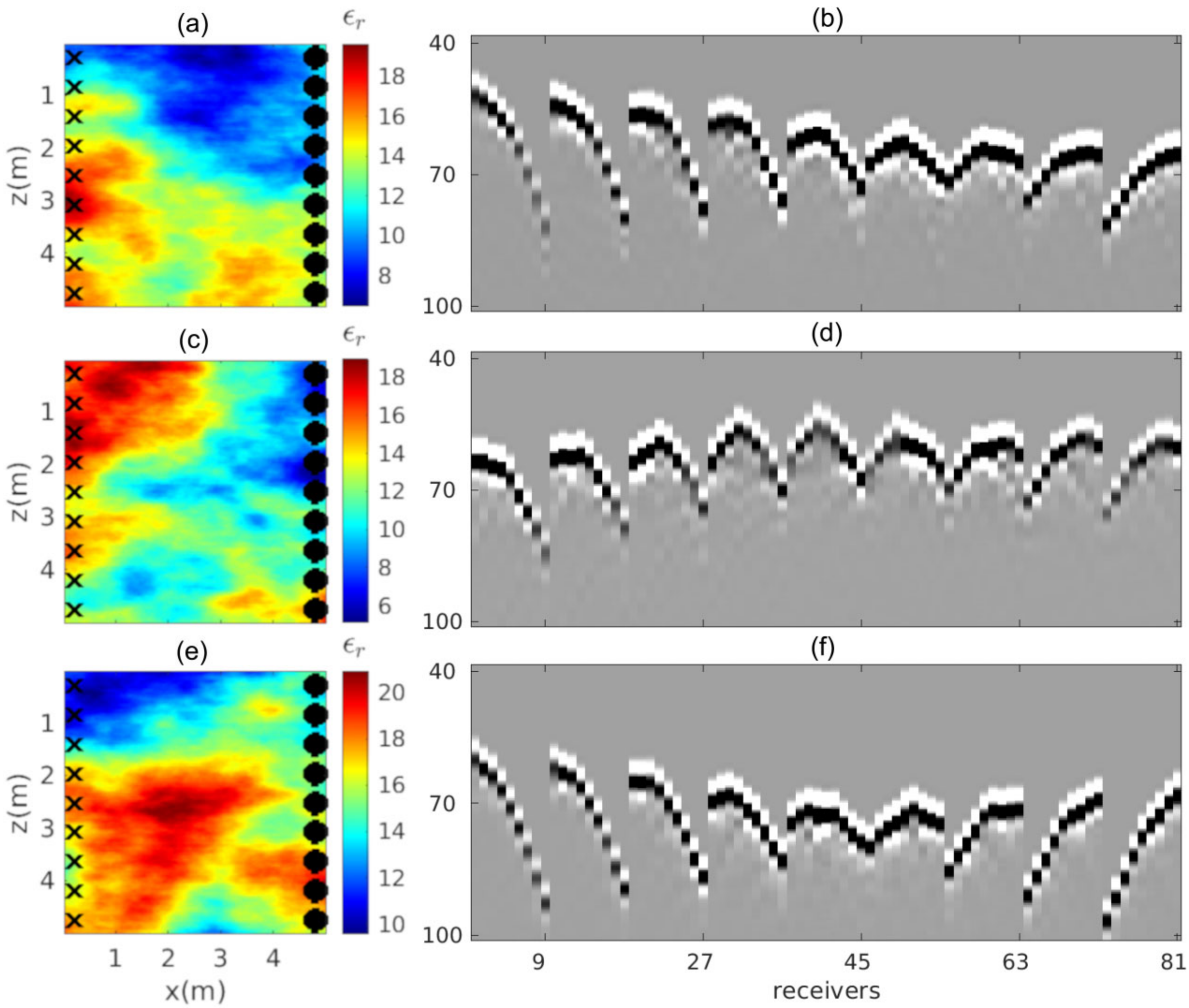

Figure 1. (a, c, e) Selected realizations from the generative model and (b, d, f) the respective waveforms at receiver locations. Source and receiver locations are marked (a, c, e) with crosses and circles, respectively.

to create a total of 1000 random realizations from the prior and learn *a prior*-knowledge-based parametrization of the input domain (i.e. ϵ_r distributions) in terms of principal components, with the first ones linked to low-wavenumber features, and higher modes revealing small-scale details. In terms of wave propagation, the low-frequency content of the data is mainly sensitive to the first principal components. To also capture the high-frequency content in the output, it is necessary to expand and consider the higher modes of the principal components. This is demonstrated in Fig. 2 using the reference ϵ_r distribution in Fig. 1(e) and its truncations on 5, 35 and 100 PCs, and the corresponding data computed via FDTD modelling for low (0–18 MHz), medium (0–45 MHz), high (0–72 MHz) and full frequencies (0–300 MHz) bandwidths. The ability of the truncated input to capture the reference data set depends on the frequency content considered, with 5 PCs being able to only adequately recover low frequencies of the data (Figs 2a–d). As we expand the PCA basis elements up to 35, the residual reduces even when considering an expanded frequency bandwidth (Figs 2e–h). Projecting the input on 100 PC components results in close agreement across the entire bandwidth of the data (Figs 2i–l).

The standard output in full waveform applications consists of all values in the inverted gathers. For the problem considered here, with gathers ranging from 0 to 110 ns and a time sampling interval of 0.32 ns, this results in a total of $9 \times 9 \times 344 = 27864$ values.

For the PCE implementation, the input parametrization is guided by the effective dimensionality analysis discussed above. Based on our analysis, we use the decomposition of the relative permittivity distribution on the first 100 input PCs ($\Pi_F = 100$ in Algorithm 1).

To reduce the output dimension and build an effective surrogate, we apply PCA alongside frequency filtering of the data. Key parameter choices were made by a performance analysis. These choices include the data subdomains and the frequency bandwidths used for defining the output principal components, and the number of principal components considered for modelling and inversion. The selection process involved evaluating the method's behaviour across various configurations to optimize accuracy and computational efficiency.

We use mutually disjoint passband filters and consider each frequency bin individually. We then apply PCA decomposition to minigathers, each consisting of 3 ($\gamma = 3$ in algorithm 1) adjacent receiver traces for each filtered data set. In total, 27 such minigathers ($\Lambda_\gamma = 27$ in algorithm 1) compose the entire data set for one frequency bin, which is trained independently of the other frequencies. For each frequency and minigather, we use two output PCs ($\widehat{\Lambda}_F = 2$ in algorithm 1) per minigather per frequency bin, which allowed us

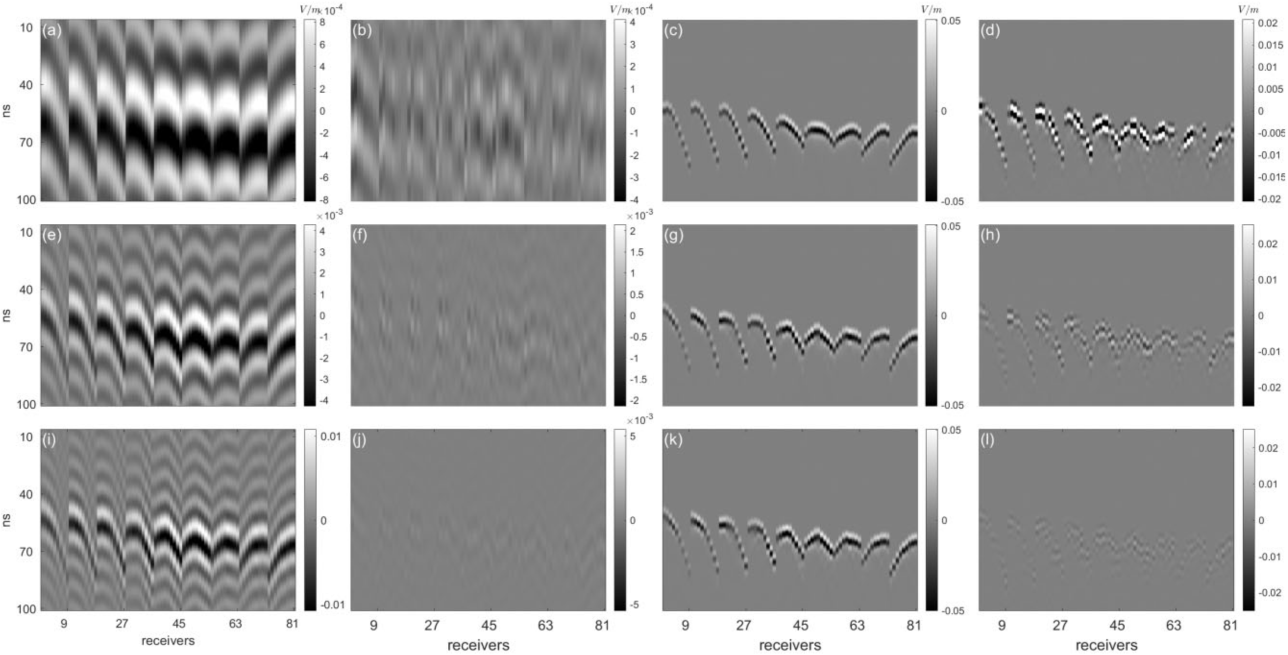

Figure 2. (a) Low-frequency (0–18 MHz) passband-filtered forward solution when the input is the projection of the relative permittivity distribution in Fig. 1(a) onto the first five PCs. (b) For the same frequency bandwidth as (a), the difference with respect to the data corresponding to the forward solution for the reference relative permittivity distribution in Fig. 1(a). (c) and (d) As in (a) and (b), but considering the full bandwidth of the data. (e)–(h) As in (a)–(d), but using 35 PCs and the 0–45 MHz bandwidth. (i)–(l) As in (a)–(d), but using 100 PCs and the 0–72 MHz bandwidth.

to capture a considerable amount of the variance in the data while retaining high accuracy of the surrogate modelling. The first two components capture, on average, about 60 per cent of the total variance across the various frequency bins and minigathers considered here.

For the FBPT scheme, a total of 16 equally spaced frequency bins, represented by the parameter Ω_F in Algorithm 1, are used. These bins correspond to a minimum and maximum frequency of 9 and 144 MHz, respectively. The parameter Ω_F in Algorithm 1 is defined as $\Omega_F = \{9, 18, 27, \ldots, 144\}$. This corresponds to a total of 864 ($\Lambda_{FT} = 864$ in algorithm 1) output data (2 PCs × 27 minigathers × 16 bins) to be inverted. A visual comparison of the full-bandwidth, passband-filtered and projected data sets used in the FBPT is shown in Fig. 3.

The training is performed on a total of 2900 samples from a broad distribution centred on the prior by a factor $\alpha_F = 2$. Expanding the input space leads to a decrease in *overall* accuracy across the effective prior PDF when the number of training sets is fixed. However, we find that the enlargement avoids the introduction of bias on specific target distributions. The enlargement factor α_F was chosen based on trial-and-error analysis. A summary of the parameters used in the FBPT experiment can be found in Table 3.

The parameters chosen for the PEPT experiment, most notably the number of PCs in the input and output, and the inverted frequencies, were determined through a trial-and-error approach, guided by heuristic considerations and empirical observations.

For the PEPT strategy, we begin with a first iteration based on draws from the same inflated prior used for the FBPT scheme, using 900 samples and the first three frequency bins, denoted as Ω_1 in Algorithm 2, corresponding to 9, 18 and 27 MHz. For each frequency bin, we consider the first two PCs of the output domain for each minigather consisting of three adjacent receivers ($\Lambda_\gamma = 27$ and $\widehat{\Lambda}_P = 2$) in Algorithm 2. This results in a total of 162 waveform attributes used in the inversion ($\Lambda_{P1} = 162$ in Algorithm 2). Since we are inverting low-frequency data in this initial stage, we consider only the first 15 PCs of the input domain ($\Pi_{P1} = 15$ in Algorithm 2), as the remaining components have minimal impact on the response and are treated as part of the 'null space'. We then train a PCE model mimicking the relationship between the first 15 input domain PCs and each of the analysed 162 output attributes. From the posterior obtained in this first iteration, we extract $\nu_1 = 500$ samples that are used when retraining the PCE surrogate, with $\rho_1 = 0$ (see Algorithm 2). This initial iteration of the algorithm employs an inflated prior-based PCE model. Afterwards, three additional iterations are performed, each leveraging the results of the previous inversion to enhance the training set in the region of significant posterior probability. For the second and third iteration, $\nu_{2,3} = 500$ and $\rho_{2,3} = 0$. For the fourth and final iteration, we used $\nu_4 = 500$ and $\rho_3 = 100$. These iterations focus on model expansion and refinement, similar to the first PEPT iteration but with a posterior-based PCE model, and with adjustments to input/output dimensions and MCMC parameters. While each iteration refines the posterior-based PCE model based on prior inversion outcomes, all other parameters, including input/output dimensions and MCMC settings, are predetermined by the iteration number, independent of the specific data set. The entire inversion process is adaptive yet fully automated. A summary of the parameters used in each PEPT experiment iteration can be found in Table 4.

We evaluate the FBPT and PEPT schemes considering waveform data generated by the ϵ_r distribution in Fig. 1(a) and compare with MCMC inversion of the corresponding traveltimes using an eikonal solver (Hansen *et al.* 2013) using the UQLab Matlab package (Wagner *et al.* 2021).

Uncorrelated Gaussian noise with a standard deviation equal to 2 per cent of the maximum amplitude of the gather is added to the synthetic FDTD data prior to PCA decomposition and inversion. The traveltimes are contaminated with uncorrelated Gaussian noise with a standard deviation of 0.5 ns.

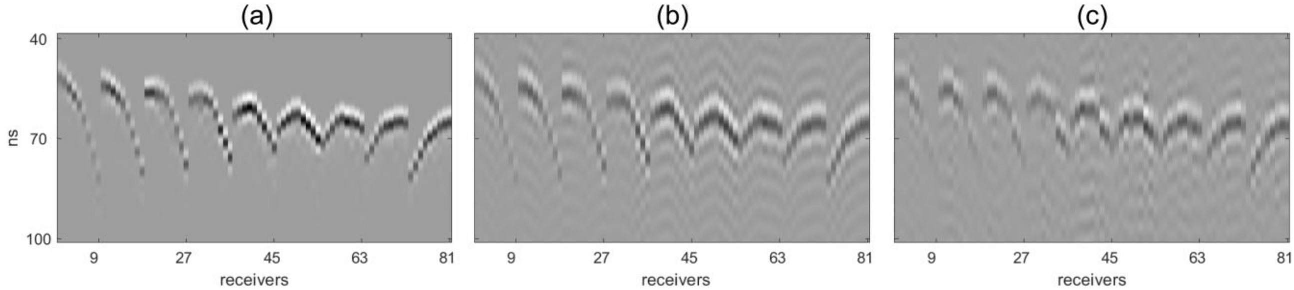

Figure 3. (a) Full-bandwidth response for the relative permittivity distribution shown in Fig. 1(a). (b) Passband-filtered response of (a) within the 0–144 MHz bandwidth. (c) Projection of (b) onto the 864 output PCs considered in the FBPT scheme.

Table 3. Parameters used in the FBPT algorithm.

Iteration	Π_F	Λ_{FT}	#Chains	MCMC Steps	Bandwidth (bins)	Training Set (size)
1	100	864	10	30000	$\Omega = \{1, 2, .., 16\}$	$I_F/O_F(2900)$

Table 4. Parameters used for the PEPT algorithm.

Iteration	Π_k	Λ_{Pk}	#Chains	MCMC steps	Bandwidth (bins)	Training set (size)
1	15	162	10	7500	$\Omega_1 = \{1, 2, 3\}$	$I_1/O_1(900)$
2	50	324	10	25000	$\Omega_2 = \{1, 2, .., 6\}$	$I_2/O_2(500)$
3	80	486	10	80 000	$\Omega_3 = \{1, 2, .., 9\}$	$I_3/O_3(500)$
4	100	864	10	100 000	$\Omega_4 = \{1, 2, .., 16\}$	$I_3/O_3(100) \cup I_4/O_4(500)$

Similarly to Meles *et al.* (2022), we account for the modelling and truncation error in the covariance operator (Hansen *et al.* 2014; Madsen & Hansen 2018). While in Meles *et al.* (2022), the use of the modelling error term only slightly improved the inversion results, here it is absolutely crucial. This can be appreciated by analysing the magnitude of the modelling error with respect to the noise level for the different frequency bins. Figs 4(a) and (b) show the modelling error terms for the FBPT and the final iteration of the PEPT schemes (for readability purposes, only one out of nine output components is shown). Fig. 4(c) further shows, as a function of the frequency bin, the mean value of the diagonal terms of the modelling error covariance matrices (red and blue solid lines), the variance of the output of the corresponding training sets (red and blue dashed lines) and the noise level (black dashed line). These figures indicate that the modelling error associated with FBPT is considerably smaller than the noise level only for the first four frequency bins and that it reaches the variance of the training set around the sixth bin. In contrast, the modelling error associated with the last iteration of the PEPT scheme, despite being larger than the noise level for most of the considered frequency bins, is consistently an order of magnitude smaller than the variance of the corresponding training set. The important differences in the magnitudes of the noise and modelling error terms highlight the importance of including the latter in the likelihood to ensure reliable inversion.

When analysing Fig. 4(c), it is important to recognize that the curves correspond to different validation sets: the FBPT scheme is validated over the prior inflated by the factor $\alpha_F = 2$, while the PEPT scheme focuses on a much narrower region centred around the posterior distribution. The substantial reduction in *relative* modelling error observed with the PBET scheme arises from training on samples obtained through inversions of progressively expanded bandwidths. This strategy concentrates on a progressively smaller portion of the prior, ensuring convergence near the target and maximizing accuracy in the subsequent development of surrogate models in the most critical regions of the input space. This can be seen in Fig. 5 for two arbitrarily chosen input parameters (whose exact values are indicated by the green dot), the progressive focusing around the target of posteriors as a function of the PEPT iteration is illustrated. For the first iteration, the target lies in a peripheral zone of the prior on which the training takes place (blue isosurfaces in Fig. 5a). As such, we can expect somewhat low accuracy in the corresponding surrogate modelling. Still, inverting the low-frequency content of the data only, as prescribed by the PEPT scheme, allows focusing around the true parameter value (red isosurfaces in Figs 5a and b). Further training leads to improved accuracy, which in turn results in an even more focused posterior as provided by the second iteration of the PEPT scheme (green isosurfaces in Fig. 5b). This trend continues throughout the PEPT scheme, ultimately leading to the reduced modelling error term discussed above and correspondingly to accurate inference.

For the FBPT scheme, we use a MH algorithm based on a Gaussian proposal distribution scaled with respect to the prior distribution. For the PEPT first iteration, we follow the same scheme as for the FBPT in terms of proposal distribution. For the remaining iterations, the correlations among the input components are inferred using the previous iteration and used to build corresponding covariance proposal matrices. Examples of such matrices are shown in Figs 6(b) and (d). While all are diagonally dominant, the matrices exhibit notable variability along the diagonal and in the off-diagonal elements. In particular, capturing the relative variations along the diagonal is crucial to enable adequate posterior exploration in the later inversion stages. As the iterations progress, the first PCs in the posterior becomes very well defined, which implies that maintaining the same scales as for FBPT would lead to very low

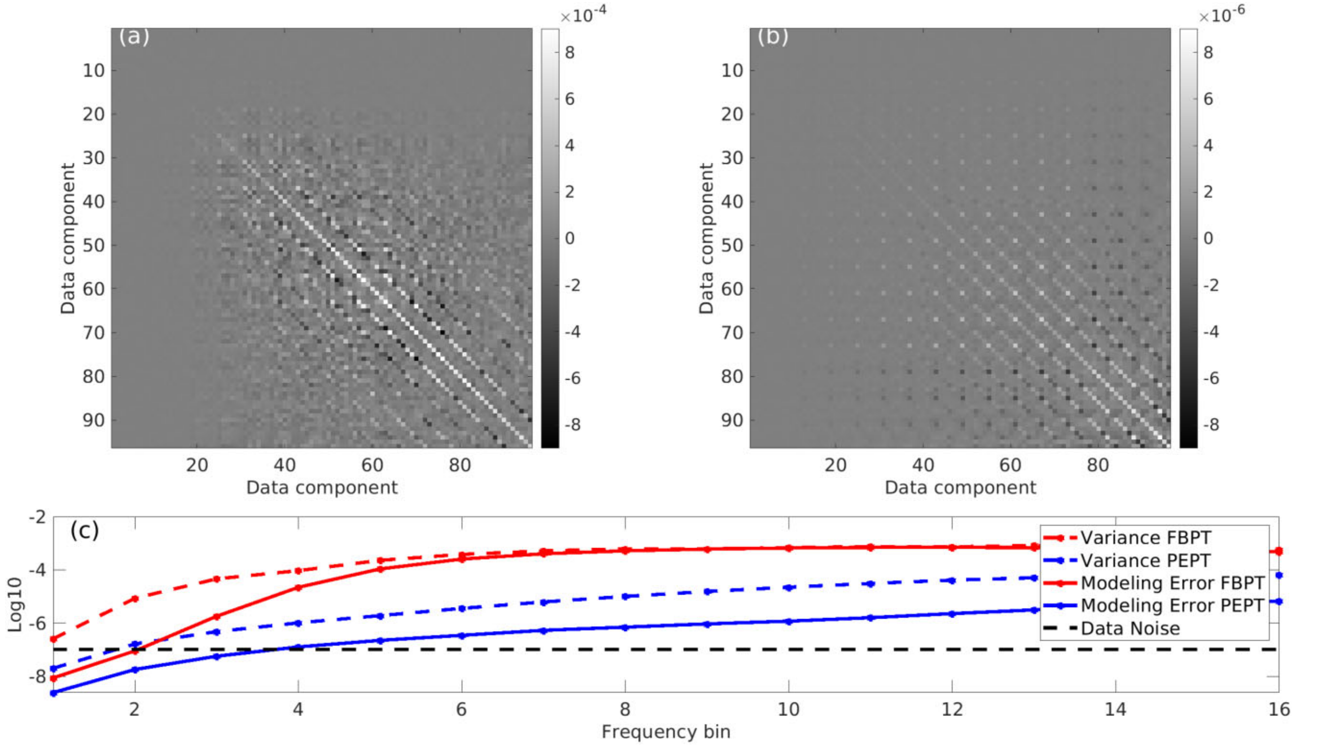

Figure 4. (a) Modelling error matrix for the FBPT scheme. For readability, only 96 out of 864 output values are shown. (b) Similar to (a), but for the fourth and final iteration of the PEPT scheme. (c) Mean values of the diagonal elements of the modelling error terms in (a) and (b) (red and blue solid lines, respectively), variance of the corresponding training sets (red and blue dashed lines, respectively) and variance of the noise terms (black dashed line) as a function of the frequency bin.

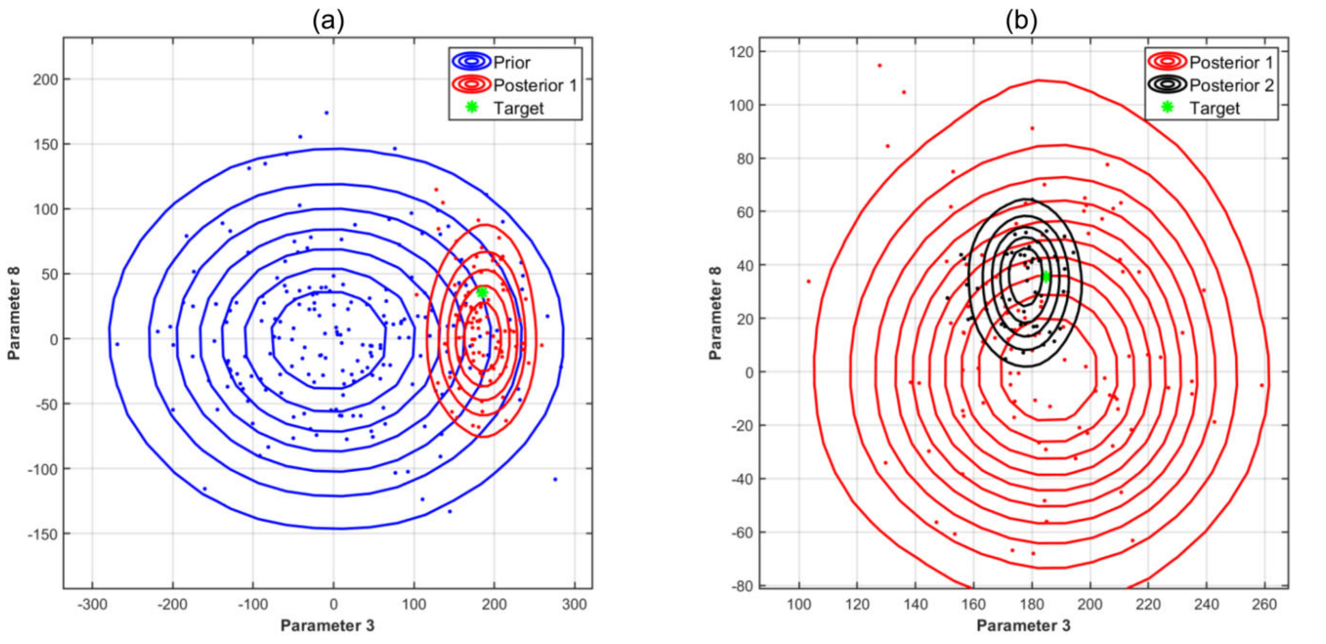

Figure 5. For the PEPT scheme, joint PDFs between the third and eight principal components for the prior (blue isosurfaces), first (red isosurfaces), and second (black isosurfaces) posteriors. The green dot corresponds to the true values of the target input distribution. Blue and red dots correspond to training elements from the prior and posterior of the 0-th PEPT iteration. The black squares in (a) and (b) cover the same areas of the parameter space.

acceptance rates. At each iteration the number of inverted parameters expands, the size of the matrices increases with each iteration, ranging from 15 to 100 and the magnitudes decrease.

The inversion results are assessed by comparing the posterior mean and Maximum *A Posteriori* (MAP) realization, including elements from the 'null space', and computing the Root Mean Square Error (RMSE), structural similarity (SSIM) and log score maps (Gneiting & Raftery 2007; Friedli *et al.* 2022). The log score evaluates the accuracy of probabilistic predictions by quantifying how well a predicted posterior probability distribution matches the

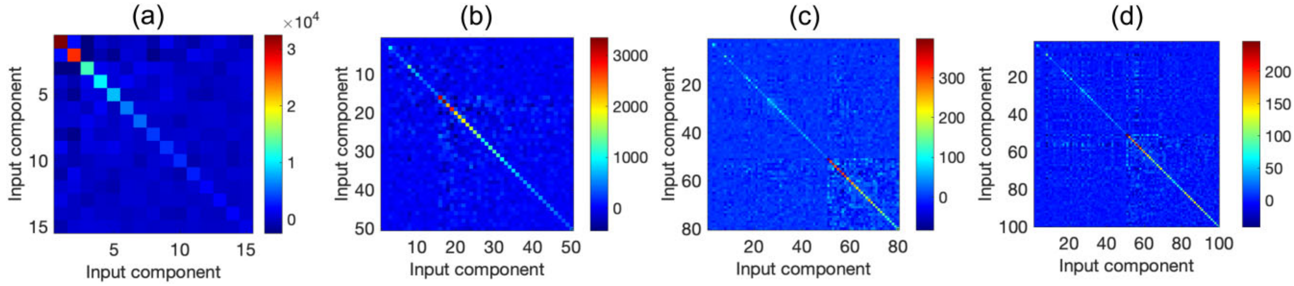

Figure 6. (a) Covariance matrix based on the validation set used in the first iteration of the PEPT scheme. (b)–(d) For the inversion associated with the permittivity distribution in Fig. 1(a), covariance matrices of the MCMC model proposals for the second to fourth PEPT iterations, showing a strong decrease in magnitude after each surrogate model update. These matrices highlight the significant differences between the subsequent PEPT matrices and what would have been used if the covariance matrices were not updated between the different iterations of the PEPT algorithm.

true values. It refers to the negative logarithm of the inferred posterior PDF evaluated at the true value. This involves using kernel density estimation (KDE) on posterior samples to estimate the PDF, and then calculating the log of the density at the true value across the domain to generate the log score maps. Mathematically, it is expressed as:

$$\log S(\hat{P}, \boldsymbol{x}_{\text{true}}) = -\log(\hat{P}(\boldsymbol{x}_{\text{true}})),$$

where $\hat{P}$ is the estimated posterior PDF, and $\boldsymbol{x}_{\text{true}}$ is the true value. A lower log score indicates that the true value has a high probability according to the posterior PDF, implying higher predictive accuracy. Conversely, a higher log score suggests that the true value has a low probability, indicating poor predictive performance, for example, due to too wide uncertainty bounds or biased predictions with small uncertainty bounds. The log score provides a single numerical value for each model parameter that can be used to compare the effectiveness of different probabilistic inversion schemes. To enhance interpretability, we present the results in the pixel domain rather than in terms of the principal components used in the inversion.

The results for the test permittivity distribution in Fig. 1(a) are summarized in Fig. 7. It is seen that the eikonal inversion results capture some of the main features of the target distribution (Figs 7a–d), with a mean value of σ_{ϵ_r} throughout the domain close to 1.55. Conversely, both the FBPT (Figs 7e–h) and PEPT (Figs 7i–l) surrogate-based waveform strategies provide much smaller σ_{ϵ_r} (0.37 and 0.43 for the FBPT and PEPT schemes, respectively). However, for the FBPT scheme, this reduction in uncertainty is accommodated by high log score values due to bias introduced by the surrogate modelling. The PEPT strategy, however, with its progressively improving modelling accuracy in the region of high posterior probability is capable of reducing both uncertainty and log score values compared to the eikonal inversions, thus emerging as the most suitable of the three considered algorithms. The reduction in bias in PEPT compared to FBPT, as well as its lower uncertainty relative to the eikonal scheme, can be better appreciated by analysing the performance of the three methods at specific points (Fig. 8). To highlight these differences, we evaluate posterior marginals at selected locations across the permittivity distribution [points A–E in Fig. 8(a)]. In Figs 8(b)–(f), we show the marginal posteriors for the three inversion schemes. The eikonal inversion results in broad distributions, indicating high uncertainty, though the true values consistently retain non-zero probability. In contrast, FBPT significantly reduces uncertainty but introduces notable bias, as seen in Figs 8(b)–(d), where the reference values have very low probability, as also shown in the corresponding log score diagram in Fig. 7(h). Finally, PEPT achieves both low uncertainty and an unbiased result, with the reference values always exhibiting significant probability at each selected point. The results for the test permittivity distributions shown in Figs 1(c) and (e) are given in Appendix B.

To speed up the inversion process in the PEPT scheme, results from different MCMC runs are combined for components shared between consecutive iterations, with the final samples from one iteration serving as the initial samples for the next. For completeness, we also performed an inversion using the final PEPT model for the distribution in Fig. 1(a), starting with samples randomly drawn from the prior. In the following, we refer to PEPT as combining samples from consecutive iterations as PEPT(0), and to the application of the final surrogate model obtained by applying PEPT(0) to samples from the prior distribution as PEPT(1). The results for PEPT(1) (Fig. 9) are very similar to those of PEPT(0) (Fig. 7).

Additional metrics of the inversion results in the input and output domains for the tests associated with the permittivity distribution in Fig. 1(a) are summarized in Tables 5 and 6.

A visual comparison of the full-bandwidth, passband-filtered and projected data sets used in the fourth PEPT iteration (Fig. 10) demonstrates that the same number of output PCs can more accurately approximate the reference data when associated with a significantly more focused input domain [see Fig. 3(c) versus Fig. 10(c)]. Similar to the FBPT strategy, the first two components capture, on average, about 60 per cent of the total variance across the various frequency bins and minigathers considered here. However, updating the output PCs as we focus in high-posterior regions is crucial for optimally capturing relevant data details. While the relative variance captured by the employed PCs may be similar between FBPT and PEPT, the overall agreement is notably improved when PEPT is applied.

Finally, a note on convergence: two of the waveform algorithms considered here, FBPT and PEPT(0), produced chains that achieved convergence for all parameters according to the Gelman-Rubin criterion (Gelman & Rubin 1992). Unsurprisingly, the FBPT algorithm achieved convergence in fewer steps due to its larger modelling error covariance matrix. In contrast, the PEPT(1) algorithm exhibited weaker convergence properties, with only one-fifth of the parameters reaching convergence. While longer chains could potentially improve convergence, a more in-depth analysis of this aspect falls outside the primary scope of this study.

4 DISCUSSION

We have introduced a Bayesian FWI scheme that is accelerated by sequential model refinement through progressive frequency bandwidth expansion (PEPT), and applied it to GPR data. Our findings

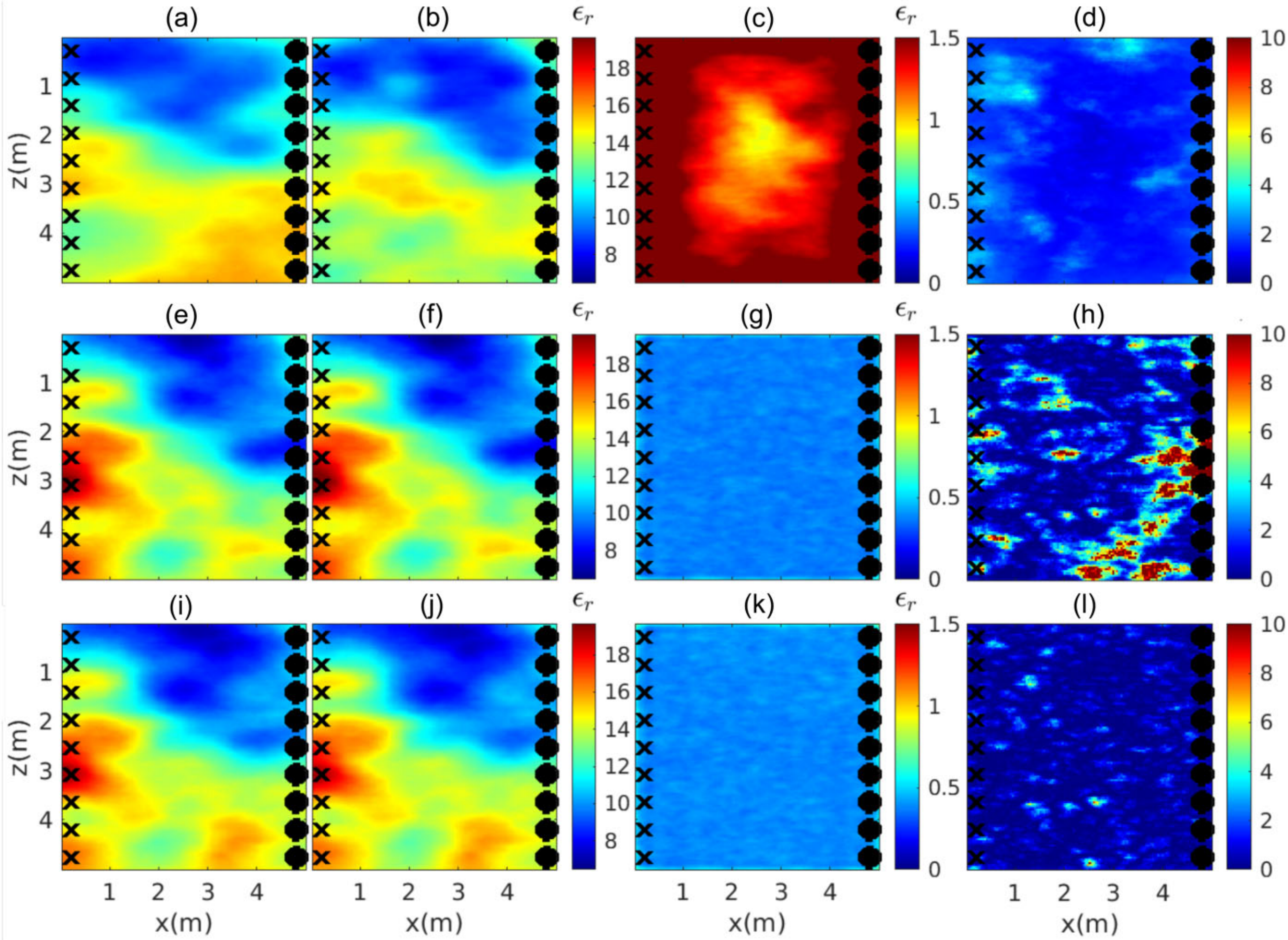

Figure 7. (a) Posterior MEAN ϵ_r, (b) MAP ϵ_r, (c) σ_{ϵ_r} and (d) log score maps evaluated with respect to the true model shown in Fig. 1(a) as provided by traveltime inversion with an eikonal solver. (e–h): as for (a–d), but for surrogate modelling of waveform data inverted by the FBPT scheme. (i–l), as for (a–d), but for waveform data inverted by the PEPT(0) scheme.

offer compelling evidence for the effectiveness of this approach compared to other approaches using an eikonal forward solver or training one surrogate model only as in the FBPT approach, yet several key areas warrant further investigation.

Identifying optimal frequency bands could significantly enhance the robustness of the PEPT scheme. Another important factor is the number of PCs used in each iteration. The computational efficiency and accuracy of the inversion are influenced by the choice of PCs, which in turn depend on the inverted frequency bandwidth and the experimental design. In theory, inclusion of higher frequencies would require more PCs (limited to a maximum of 100 in this study). Interestingly, the polynomial complexity of FBPT and PEPT differs markedly when applied to the same input domain (i.e. in the last PEPT iteration), particularly in terms of the required truncation scheme. To reach the relatively modest accuracy reported in this paper, FBPT required a maximum polynomial degree of five, with higher degrees proving computationally prohibitive. In contrast, PEPT achieved significantly better accuracy with a maximum degree of only two. While the polynomial degree in FBPT had to be capped due to the associated computational cost, PEPT attained sufficient accuracy well before computational constraints became a limiting factor. The relative simplicity of the PCE associated with PEPT also results in significantly faster execution times. Although the PEPT scheme is less computationally demanding in terms of PCE complexity and could accommodate even more PCs, we limited the input dimensionality to ensure a fair comparison. Future work could explore adaptive approaches that dynamically adjust the number of PCs based on the frequency range and the complexity of the subsurface model, potentially enhancing performance. Another promising direction involves applying an additional PCA to exploit correlations among the currently used PCs under the posterior PDF. If successful, this approach could further accelerate the PEPT scheme, whose computational cost scales linearly with the number of output components.

The size of the training set used for surrogate model construction also plays an important role in determining accuracy. While larger training sets capture more variability, they come at a higher computational cost. Research in how to determine optimal training set sizes and efficient sampling strategies would be beneficial to strike the right balance between accuracy and computational efficiency. Additionally, the way input and output parameters are defined significantly impacts the surrogate model's accuracy and, by extension, the entire inversion process. In cases involving complex geological priors, advanced dimensionality reduction techniques beyond PCA may be required. Deep generative models (DGMs) present

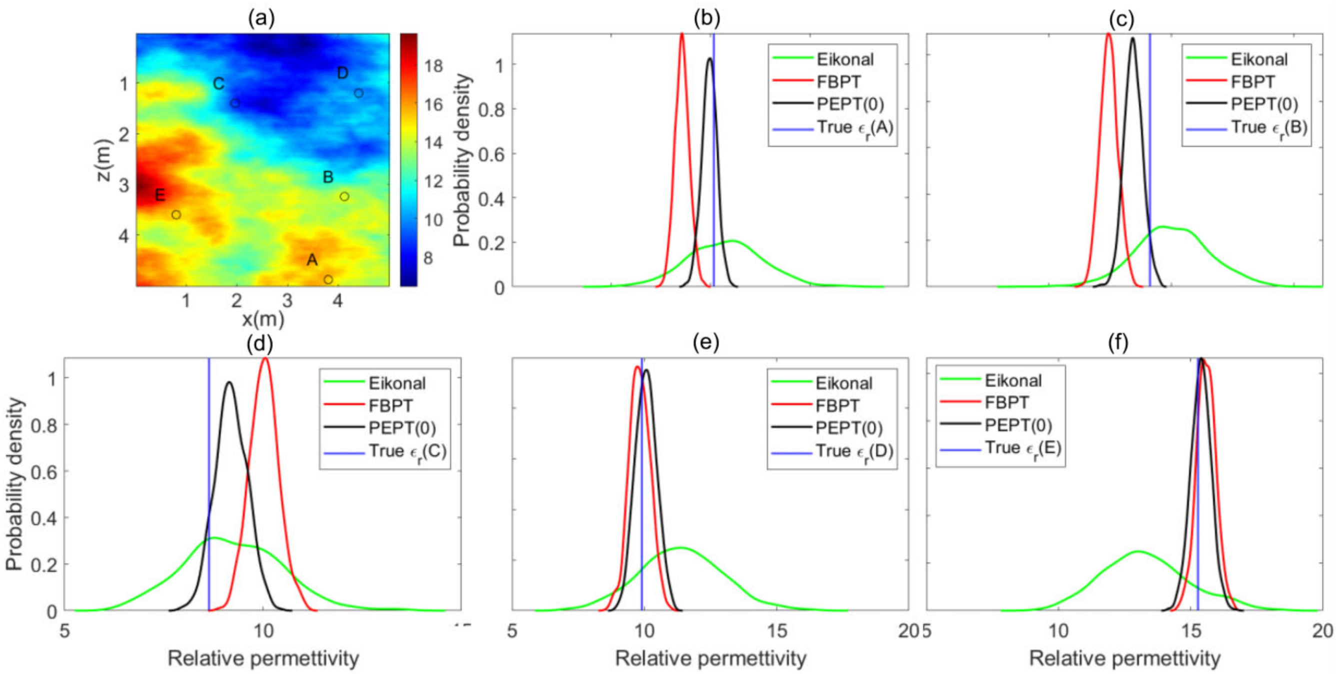

Figure 8. For selected points A–E distributed across the permittivity distribution (a), (b–f) show the overlay of the corresponding posterior marginals for the eikonal, FBPT and PEPT(0) inversion schemes.

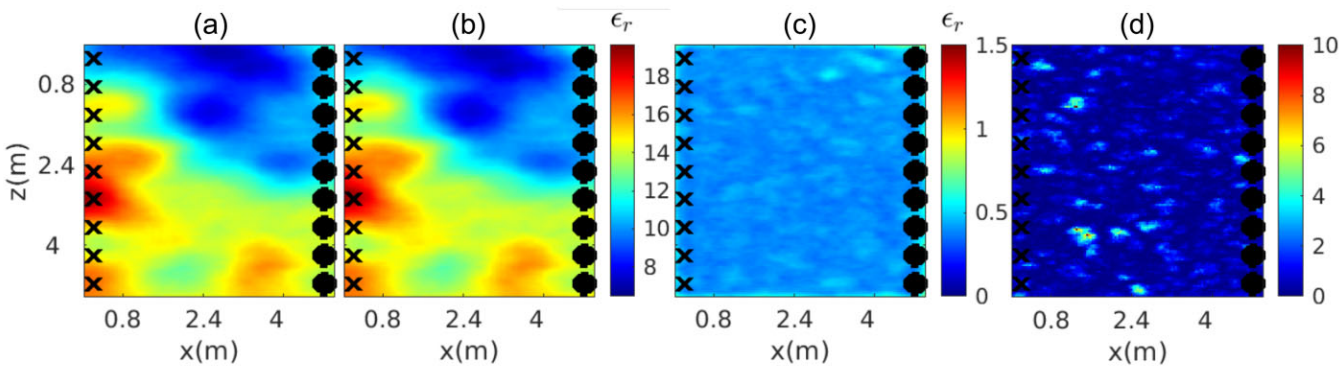

Figure 9. (a) Posterior MEAN ϵ_r, (b) MAP ϵ_r, (c) σ_{ϵ_r} and (d) log score maps evaluated with respect to the true model shown in Fig. 1(a) using the PEPT(1) scheme.

Table 5. Summary of the inversion results in the data domain for the permittivity distribution in Fig. 1(a).

ALGORITHM:	DATA RMSE MEAN (V m^{-1})	DATA RMSE MAP (V m^{-1})
EIKONAL	0.277	0.626
FBPT	0.393	0.380
PEPT(0)	0.046	0.048
PEPT(1)	0.053	0.056

Table 6. Summary of the inversion results in the input domain for the permittivity distribution in 1(a).

ALGORITHM:	RMSE MEAN$_{\epsilon_r}$	RMSE MAP$_{\epsilon_r}$	SSIM MEAN$_{\epsilon_r}$	SSIM MAP$_{\epsilon_r}$	$\bar{\sigma}_{POST}$	$\overline{\log(S)}$
EIKONAL	1.44	1.50	0.64	0.61	1.54	1.72
FBPT	0.71	0.76	0.80	0.80	0.38	2.56
PBET(0)	0.37	0.38	0.82	0.81	0.44	0.48
PBET(1)	0.39	0.39	0.80	0.80	0.40	0.50

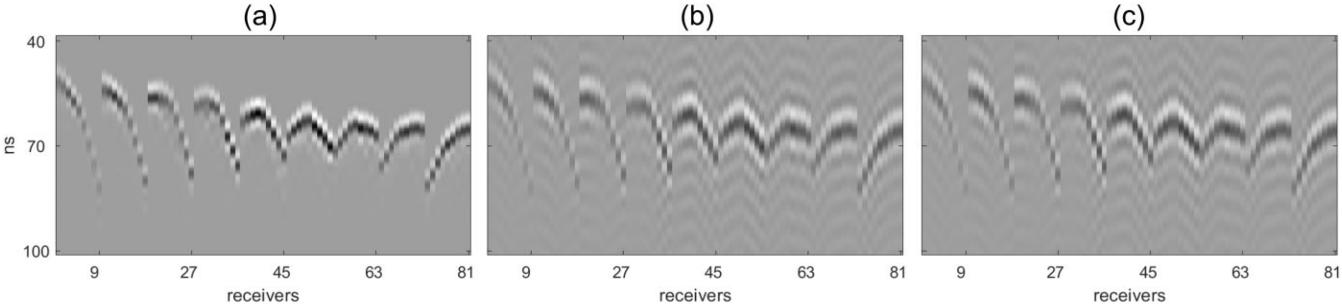

Figure 10. (a) Full-bandwidth response for the relative permittivity distribution shown in Fig. 1(a). (b) Passband-filtered response of (a) within the 0–144 MHz bandwidth. (c) Projection of (b) onto the 864 output PCs considered in the PEPT(0) scheme.

a promising tool for this, although they pose challenges for PCE-based surrogate modelling.

In this study, we did not invert for electrical conductivity and we assumed known source characteristics, which simplified the problem by reducing the number of parameters. In seismic applications, complexity increases significantly due to multiple parameters, such as compressional wave velocity, shear wave velocity and density, resulting in a greater number of input variables compared to ground-penetrating radar (GPR) scenarios. Nevertheless, our approach of progressively building more accurate surrogates is robust and should be tested across different wave phenomena and acquisition settings, including seismic surface data.

As an alternative to training a single surrogate model across the entire prior, as done in the FBPT scheme, one could consider training multiple surrogate models on different subdomains of the prior. This would avoid the need for progressive model refinement as prescribed by the PEPT scheme. However, given the relatively vastness of the prior space with respect to the posterior achieved via FWI, this approach would require the creation of millions of surrogate models, making such an approach highly inefficient.

In terms of computational expenses, all the numerical experiments in this paper were performed on a workstation equipped with 16GB of DDR4 RAM and a 3.5GHz Quad-Core processor, running MATLAB on Linux. The three methods we examined—eikonal, FBPT and PEPT—incurred similar computational costs, averaging around 10 hr. We opted not to conduct an FDTD-based MCMC inversion due to its prohibitively high computational demands (Hunziker *et al.* 2019). If we had followed the same number of iterations as in the final PEPT strategy, the execution time would have been roughly 200 times longer than that required for a full PEPT inversion. In contrast to FDTD-based inversion, our strategies can be executed on standard computing hardware within a practical timeframe, making them accessible for a variety of applications. Additionally, the algorithm is comparatively simple and can be easily adapted to meet specific user requirements.

5 CONCLUSIONS

We have introduced a novel strategy for accelerating MCMC inversion of waveform data using surrogate modelling. Our approach employs a sequential, data-driven method to circumvent the challenge of training a single surrogate model capable of capturing the complexity of the input–output relationship across the entire prior PDF. We leverage the strong relationship between the low-wavenumber and low-frequency components of both parameter space and data domain. The input and output functions are parametrized using appropriate principal component schemes to derive an initial surrogate model with limited coverage but sufficient accuracy. This preliminary surrogate model serves as a basis for refining modelling accuracy across an expanded bandwidth range through retraining with posterior samples from MCMC targeting limited bandwidths of the data. As the frequency content of the inverted data is expanded, so is the dimension of the input space used to model it, as shorter wavelength structures become more and more relevant in explaining the data. Through an iterative process, we ultimately obtain a final surrogate model capable of inverting the entire data set for the definitive MCMC inversion. Unlike traveltime inversion via surrogate modelling, where modelling errors are typically small compared to noise, surrogate-based waveform inversion schemes can encounter larger modelling errors, especially for high frequencies. Hence, incorporating these errors in the covariance term of the likelihood is necessary. Our strategy yields significantly better results than those provided by an eikonal solver at comparatively low computational costs, while a surrogate-based waveform modelling approach trained only on samples from the prior PDF leads to biased estimates with misleadingly small uncertainty estimates.

ACKNOWLEDGMENTS

The authors would like to sincerely thank Prof Hervé Chauris for his careful editorial guidance. We also gratefully acknowledge the constructive comments provided by the anonymous reviewer, which have significantly helped us improve the quality and clarity of this paper.

DATA AVAILABILITY

The data and scripts underlying this paper will be shared on reasonable request to the authors.

REFERENCES

Adler, J. & Öktem, O., 2024. Deep Bayesian inversion, in *Data-driven Models in Inverse Problems,* pp. 359–412.

Aki, K. & Richards, P., 2002. *Quantitative Seismology (second edition),* University Science Books.

Aleardi, M., 2020. Combining discrete cosine transform and convolutional neural networks to speed up the Hamiltonian Monte Carlo inversion of pre-stack seismic data, *Geophys. Prospect.,* **68**(9), 2738–2761.

Amaya, M., Meles, G., Marelli, S. & Linde, N., 2024. Multifidelity adaptive sequential Monte Carlo for geophysical inversion, *J. geophys. Int.,* **237**(2), 788–804.

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

Billings, S.A., 2013. *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-temporal Domains,* John Wiley & Sons, Inc.

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

Blatman, G. & Sudret, B., 2011b. Principal component analysis and least angle regression in spectral stochastic finite element analysis, in *Proc. 11th Int. Conf. on Applications of Stat. and Prob. in Civil Engineering (ICASP11),* pp. 669–676, Zurich, Switzerland.

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. Association for Computing Machinery, New York.

Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics,* **60**(5), 1457–1473.

Chew, W.C. & Lin, J., 1995. A frequency-hopping approach for microwave imaging of large inhomogeneous bodies, *IEEE Microwave Guided Wave Lett.,* **5**(12), 439–441.

Chib, S. & Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm, *Am. Statistician,* **49**(4), 327–335.

Cui, T., Marzouk, Y.M. & Willcox, K.E., 2015. Data-driven model reduction for the Bayesian solution of inverse problems, *Int. J. Numer. Methods Eng.,* **102**(5), 966–990.

Dhabaria, N. & Singh, S.C., 2024. Hamiltonian Monte Carlo based elastic full-waveform inversion of wide-angle seismic data, *J. geophys. Int.,* **237**(3), 1384–1399.

Dietrich, C.R. & Newsam, G.N., 1997. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix, *SIAM J. Sci. Comput.,* **18**(4), 1088–1107.

Ernst, J.R., Green, A.G., Maurer, H. & Holliger, K., 2007a. Application of a new 2d time-domain full-waveform inversion scheme to crosshole radar data, *Geophysics,* **72**(5), J53–J64.

Ernst, J.R., Maurer, H., Green, A.G. & Holliger, K., 2007b. Full-waveform inversion of crosshole radar data based on 2-D finite-difference time-domain solutions of Maxwell's equations, *IEEE Trans. Geosci. Remote Sens.,* **45**(9), 2807–2828.

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. & Villasenor, A., 2013. Multiscale full waveform inversion, *J. geophys. Int.,* **194**(1), 534–556.

Friedli, L., Linde, N., Ginsbourger, D. & Doucet, A., 2022. Lithological tomography with the correlated pseudo-marginal method, *J. geophys. Int.,* **228**(2), 839–856.

Gebraad, L., Boehm, C. & Fichtner, A., 2020. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. geophys. Res.: Solid Earth,* **125**(3), e2019JB018428.

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

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B., 1995. *Bayesian Data Analysis,* Chapman and Hall/CRC.

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,** 5901110. doi:

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

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

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.

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.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov Chains and their applications, *Biometrika,* **57**(1), 97–109.

Hunziker, J., Laloy, E. & Linde, N., 2017. Inference of multi-gaussian relative permittivity fields by probabilistic inversion of crosshole ground-penetrating radar data, *Geophysics,* **82**(5), H25–H40.

Hunziker, J., Laloy, E. & Linde, N., 2019. Bayesian full-waveform tomography with application to crosshole ground penetrating radar data, *J. geophys. Int.,* **218**(2), 913–931.

Irving, J. & Knight, R., 2006. Numerical modeling of ground-penetrating radar in 2-D using MATLAB, *Comput. Geosci.,* **32**(9), 1247–1258.

Jolliffe, I.T. & Cadima, J., 2016. Principal component analysis: a review and recent developments, *Phil. Trans. R. Soc. A: Math. Phys. Eng. Sci.,* **374**(2065), 20150202.

Khan, A. *et al.*, 2016. Single-station and single-event marsquake location and inversion for structure using synthetic Martian waveforms, *Phys. Earth planet. Inter.,* **258,** 28–42.

Kleijnen, J.P. & Mehdad, E., 2014. Multivariate versus univariate kriging metamodels for multi-response simulation models, *Eur. J. Operation. Res.,* **236**(2), 573–582.

Klotzsche, A., Vereecken, H. & van der Kruk, J., 2019. Review of crosshole ground-penetrating radar full-waveform inversion of experimental data: Recent developments, challenges, and pitfalls, *Geophysics,* **84**(6), H13–H28.

Kuroda, S., Takeuchi, M. & Kim, H.J., 2007. Full-waveform inversion algorithm for interpreting crosshole radar data: a theoretical approach, *Geosci. J.,* **11**(3), 211–217.

Laloy, E., Linde, N., Jacques, D. & Vrugt, J.A., 2015. Probabilistic inference of multi-Gaussian fields from indirect hydrological data using circulant embedding and dimensionality reduction, *Water Resources Res.,* **51**(6), 4224–4243.

Linde, N., Ginsbourger, D., Irving, J., Nobile, F. & Doucet, A., 2017. On uncertainty quantification in hydrogeology and hydrogeophysics, *Adv. Water Resources,* **110,** 166–181.

Liu, B., Ren, Y., Liu, H., Xu, H., Wang, Z., Cohn, A.G. & Jiang, P., 2021. Gprinvnet: Deep learning-based ground-penetrating radar data inversion for tunnel linings, *IEEE Trans. Geosci. Remote Sens.,* **59**(10), 8305–8325.

Liu, J.S. & Liu, J.S., 2001. *Monte Carlo strategies in scientific computing,* Vol. **10,** Springer.

Lüthen, N., Marelli, S. & Sudret, B., 2021. Sparse polynomial chaos expansions: Literature survey and benchmark, *SIAM/ASA J. Uncertainty Quant.,* **9**(2), 593–649.

Lüthen, N., Marelli, S. & Sudret, B., 2022. Automatic selection of basis-adaptive sparse polynomial chaos expansions for engineering applications, *Int. J. Uncertainty Quant.,* **12**(3).

Madsen, R.B. & Hansen, T.M., 2018. Estimation and accounting for the modeling error in probabilistic linearized amplitude variation with offset inversion, *Geophysics,* **83**(2), N15–N30.

Mai, C.V. & Sudret, B., 2017. Surrogate models for oscillatory systems using sparse polynomial chaos expansions and stochastic time warping, *SIAM/ASA J. Uncertainty Quant.,* **5**(1), 540–571.

Marelli, S. & Sudret, B., 2014. UQLab: A framework for uncertainty quantification in Matlab, in *Vulnerability, Uncertainty, and Risk (Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom),* pp. 2554–2563. ASCE, Reston, VA.

Marelli, S., Lüthen, N. & Sudret, B., 2021. UQLab user manual–Polynomial chaos expansions, Tech. rep., *Chair of Risk, Safety and Uncertainty Quantification,* ETH Zurich, Switzerland, Report # UQLab-V1.4-104.

Meles, G.A., Van der Kruk, J., Greenhalgh, S.A., Ernst, J.R., Maurer, H. & Green, A.G., 2010. A new vector waveform inversion algorithm for simultaneous updating of conductivity and permittivity parameters from combination crosshole/borehole-to-surface GPR data, *IEEE Trans. Geosci. Remote Sens.,* **48**(9), 3391–3407.

Meles, G.A., Greenhalgh, S., van der Kruk, J., Green, A. & Maurer, H., 2012. Taming the non-linearity problem in GPR full-waveform inversion for high contrast media, *J. Appl. Geophys.,* **78,** 31–43.

Meles, G.A., Linde, N. & Marelli, S., 2022. Bayesian tomography with prior-knowledge-based parametrization and surrogate modeling, *J. geophys. Int.,* **231,** 673–691. doi: 10.1093/gji/ggac214.

Meles, G.A., Amaya, M., Levy, S., Marelli, S. & Linde, N., 2024. Bayesian tomography using polynomial chaos expansion and deep generative networks, *J. geophys. Int.,* **237**(1), 31–48.
Métivier, D., Vuffray, M. & Misra, S., 2020. Efficient polynomial chaos expansion for uncertainty quantification in power systems, *Electric Power Syst. Res.,* **189,** 106791.
Mulder, W. & Plessix, R.-E., 2008. Exploring some issues in acoustic full waveform inversion, *Geophys. Prospect.,* **56**(6), 827–841.
Peherstorfer, B., Willcox, K. & Gunzburger, M., 2018. Survey of multifidelity methods in uncertainty propagation, inference, and optimization, *SIAM Rev.,* **60**(3), 550–591.
Pratt, R.G., Shin, C. & Hick, G., 1998. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion, *J. geophys. Int.,* **133**(2), 341–362.
Reynolds, A.C., He, N., Chu, L. & Oliver, D.S., 1996. Reparametrization 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.
Schär, S., Marelli, S. & Sudret, B., 2024. Emulating the dynamics of complex systems using autoregressive models on manifolds (mNARX), *Mech. Syst. Signal Process.,* **208,** 110956.
Stotzka, R., Wuerfel, J., Mueller, T.O. & Gemmeke, H., 2002. Medical imaging by ultrasound computer tomography, in SPIE *Medical Imaging 2002: Ultrasonic Imaging and Signal Processing,* Vol. **4687,** pp. 110–119, International Society for Optics and Photonics.
Sun, J., Innanen, K., Zhang, T. & Trad, D., 2023. Implicit seismic full waveform inversion with deep neural representation, *J. geophys. Res.: Solid Earth,* **128**(3), e2022JB025964.
Taillandier, C., Noble, M., Chauris, H. & Calandra, H., 2009. First-arrival traveltime tomography based on the adjoint-state method, *Geophysics,* **74**(6), WCB1–WCB10.
Tant, K. M.M., Galetti, E., Mulholland, A., Curtis, A. & Gachagan, A., 2018. A transdimensional Bayesian approach to ultrasonic travel-time tomography for non-destructive testing, *Inverse Probl.,* **34**(9), 095002.
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,** 126903.
Ulrych, T.J., Sacchi, M.D. & Woodbury, A., 2001. A Bayes tour of inversion: A tutorial, *Geophysics,* **66**(1), 55–69.
Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics,* **74,** WCC1–WCC26.
Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A. & Zhou, W., 2017. An introduction to full waveform inversion, in *Encyclopedia of exploration geophysics,* pp. R1–1, Society of Exploration Geophysicists.
Wagner, P.-R., Nagel, J., Marelli, S. & Sudret, B., 2021. UQLab user manual–Bayesian inversion for model calibration and validation, Tech. rep., *Chair of Risk, Safety and Uncertainty Quantification,* ETH Zurich, Switzerland Report UQLab-V1.4-113.
Wang, W., Yang, F. & Ma, J., 2018. Velocity model building with a modified fully convolutional network, in *SEG International Exposition and Annual Meeting,* pp. SEG–2018, SEG.
Williamson, P., 1990. Tomographic inversion in reflection seismology, *J. geophys. Int.,* **100**(2), 255–274.
Wu, Y., Lin, Y. & Zhou, Z., 2018. Inversionnet: Accurate and efficient seismic waveform inversion with convolutional neural networks, in *SEG International Exposition and Annual Meeting,* pp. 419–433, SEG.
Xiu, D. & Karniadakis, G.E., 2002. The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.,* **24**(2), 619–644.
Yaghoubi, V., Marelli, S., Sudret, B. & Abrahamsson, T., 2017. Sparse polynomial chaos expansions of frequency response functions using stochastic frequency transformation, *Probab. Eng. Mech.,* **48,** 39–58.
Yang, F. & Ma, J., 2019. Deep-learning inversion: A next-generation seismic velocity model building method, *Geophysics,* **84**(4), R583–R599.
Zhang, X., Lomas, A., Zhou, M., Zheng, Y. & Curtis, A., 2023. 3-D Bayesian variational full waveform inversion, *J. geophys. Int.,* **234**(1), 546–561.
Zhao, D., Lei, J. & Liu, L., 2008. Seismic tomography of the Moon, *Chinese Sci. Bull.,* **53**(24), 3897–3907.
Zhou, B. & Greenhalgh, S.A., 2003. Crosshole seismic inversion with normalized full-waveform amplitude data, *Geophysics,* **68**(4), 1320–1330.

APPENDIX A:

For this study, we rely on PCE-based surrogate modelling due to its efficiency, flexibility and ease of deployment (Xiu & Karniadakis 2002; Blatman & Sudret 2011a; Métivier *et al.* 2020; Lüthen *et al.* 2021). PCE approximates functions in terms of linear combinations of orthonormal multivariate polynomials $\Psi_{\boldsymbol{\alpha}}$:

$$\tilde{\mathcal{M}}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}), \tag{A1}$$

where M is the dimension of $\boldsymbol{x}$ and $\mathcal{A}$ is a subset of $\mathbb{N}^M$ defined by a truncation scheme to be set based on accuracy requirements and available computational resources (Xiu & Karniadakis 2002).

Training 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 $\mathcal{F}(\boldsymbol{u})$). To alleviate this computational challenge, we use sparse regression techniques to identify and retain only the most relevant polynomial basis terms. However, while such approaches significantly reduce the number of required model evaluations and enable the use of higher polynomial degrees, the fundamental concerns related to the choice of truncation scheme—especially its capacity to capture the model's complexity—still remain. In this paper we showed how training the surrogate to approximate the model response only in regions of interest can decrease modelling complexity and thus increase accuracy. Once derived, the surrogate response can be evaluated at a negligible cost by direct computation of eq. (A1) and its accuracy estimated using a validation set or cross-validation techniques (Blatman & Sudret 2011a; Marelli *et al.* 2021).

APPENDIX B:

In the main body of this paper, we have presented results using the eikonal, FBPT and PEPT(0) inversion schemes, along with the corresponding ϵ_r distribution shown in Fig. 1(a). To demonstrate the general character of these results, we provide here in Figs B1 and B2, respectively, the results obtained when considering the permittivity distributions in Figs 1(c) and (e). Note that all the examples presented here share the same prior and training parameters. These results are consistent with what is shown in Fig. 7. The traveltime inversion results exhibit high uncertainty [Figs B1(c) and B2(c)]. This uncertainty decreases dramatically when the FBPT scheme is applied [Figs B1(g) and B2(g)], but these results may imply a significant bias, as indicated in the log score maps [Figs B1(h) and B2(h)]. Finally, uncertainty is similarly reduced when the PEPT(0) algorithm is used [Figs B1(k) and B2(k)], but in this case, the bias is drastically reduced, as evidenced by the low log score values [Figs B1(l) and B2(l)]. Moreover, the trend observed in the summary of output-input MEAN/MAP fitting in Tables 5 and 6 for the first numerical experiment is further confirmed in Tables B1 to B4, which correspond to the numerical tests discussed in this appendix.

Table B1. Metrics of the inversion results in the data domain for the permittivity distribution in 1(c).

ALGORITHM:	DATA RMSE MEAN ($\mathrm{V\,m^{-1}}$)	DATA RMSE MAP ($\mathrm{V\,m^{-1}}$)
EIKONAL	0.292	0.301
FBPT	0.378	0.400
PEPT(0)	0.058	0.059

Table B2. Metrics of the inversion results in the input domain for the permittivity distribution in Fig. 1(c).

ALGORITHM:	RMSE MEAN$_{\epsilon_r}$	RMSE MAP$_{\epsilon_r}$	SSIM MEAN$_{\epsilon_r}$	SSIM MAP$_{\epsilon_r}$	$\bar{\sigma}_{POST}$	$\overline{\log(S)}$
EIKONAL	1.87	1.72	0.63	0.62	1.35	1.98
FBPT	0.62	0.66	0.78	0.78	0.37	1.74
PBET(0)	0.45	0.46	0.79	0.79	0.39	0.67

Table B3. Metrics of inversion results in the data domain for the permittivity distribution in Fig. 1(e).

ALGORITHM:	DATA RMSE MEAN ($\mathrm{V\,m^{-1}}$)	DATA RMSE MAP ($\mathrm{V\,m^{-1}}$)
EIKONAL	0.201	0.267
FBPT	0.652	0.612
PEPT(0)	0.044	0.048

Table B4. Metrics of the inversion results in the input domain for the permittivity distribution in Fig. 1(e).

ALGORITHM:	RMSE MEAN$_{\epsilon_r}$	RMSE MAP$_{\epsilon_r}$	SSIM MEAN$_{\epsilon_r}$	SSIM MAP$_{\epsilon_r}$	$\bar{\sigma}_{POST}$	$\overline{\log(S)}$
EIKONAL	1.01	1.51	0.72	0.68	1.18	1.46
FBPT	0.69	0.69	0.80	0.80	0.37	2.31
PBET(0)	0.34	0.35	0.82	0.82	0.39	0.41

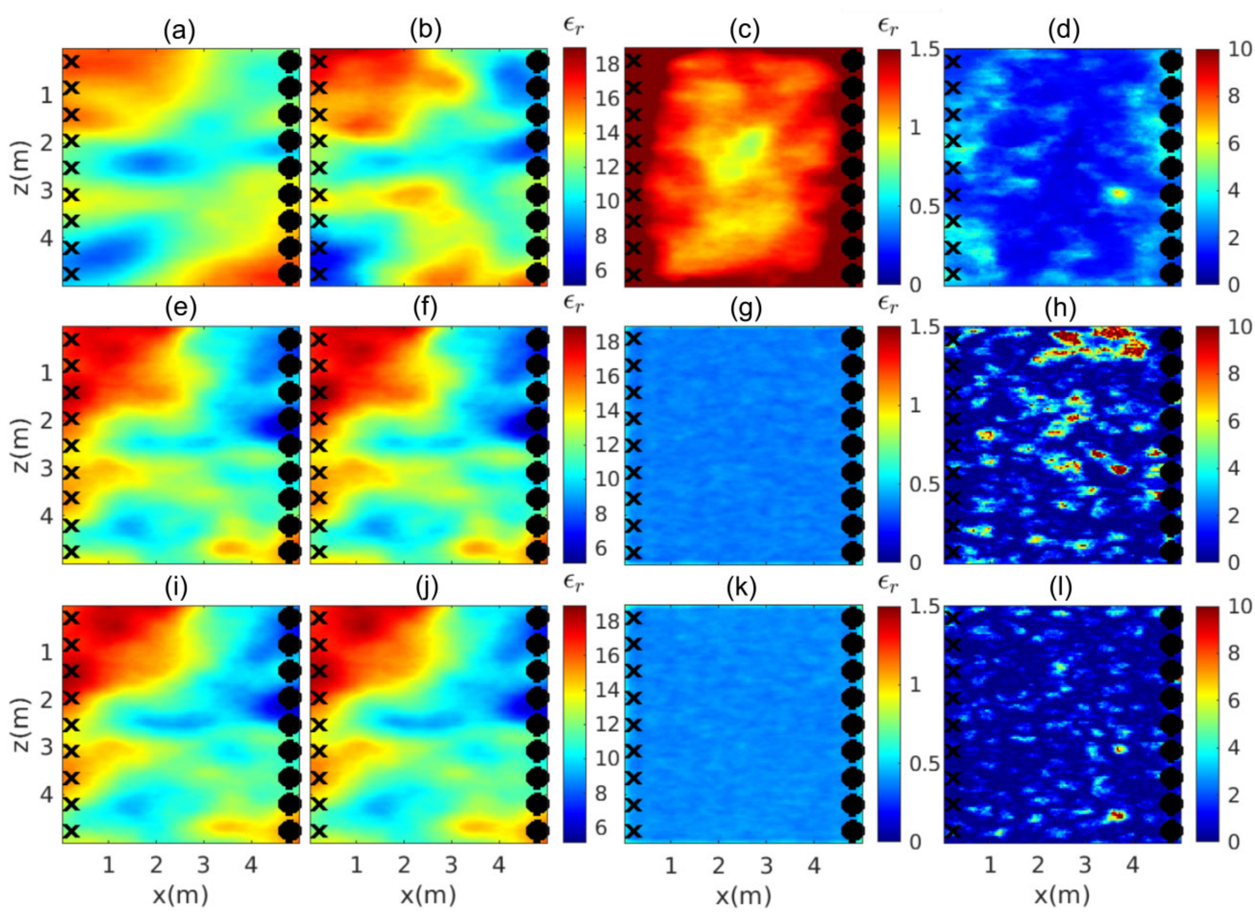

Figure B1. (a) Posterior MEAN ϵ_r, (b) MAP ϵ_r, (c) σ_{ϵ_r} and (d) log score maps evaluated with respect to the true model shown in Fig. 1(c) as provided by traveltime inversion with an eikonal solver. (e–h): as for (a–d), but for surrogate modelling of waveform data inverted by the FBPT scheme. (i–l), as for (a–d), but for waveform data inverted by the PEPT(0) scheme.

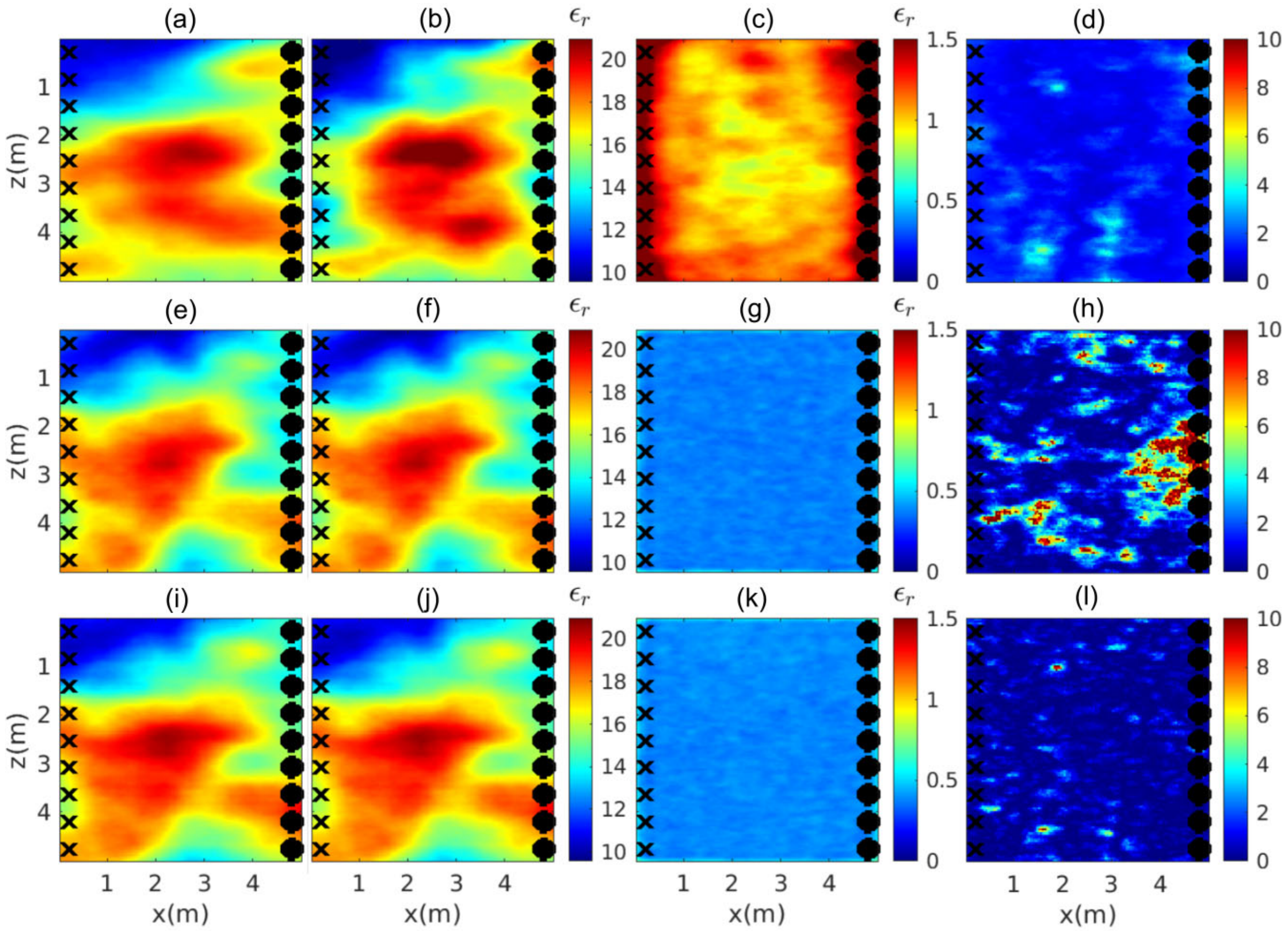

Figure B2. (a) Posterior MEAN ϵ_r, (b) MAP ϵ_r, (c) σ_{ϵ_r} and (d) log score maps evaluated with respect to the true model shown in Fig. 1(e) as provided by traveltime inversion with an eikonal solver. (e–h): as for (a–d), but for surrogate modelling of waveform data inverted by the FBPT scheme. (i–l), as for (a–d), but for waveform data inverted by the PEPT(0) scheme.