Geophysical Journal International

*Geophys. J. Int.* (2024) **237,** 788–804 https://doi.org/10.1093/gji/ggae040
Advance Access publication 2024 January 27
GJI Applied and Marine Geophysics

# Multifidelity adaptive sequential Monte Carlo for geophysical inversion

M. Amaya ,[1] G. Meles ,[1] S. Marelli[2] and N. Linde[1]

[1]*Institute of Earth Sciences, University of Lausanne,*1015 *Lausanne, Switzerland. E-mail:* macarena.amaya@unil.ch
[2]*Chair of Risk, Safety and Uncertainty Quantification, ETH Zürich,* 8093 *Zürich, Switzerland*

Accepted 2024 January 26. Received 2023 December 18; in original form 2023 August 29

## SUMMARY

In the context of Bayesian inversion, we consider sequential Monte Carlo (SMC) methods that provide an approximation of the posterior probability density function and the evidence (marginal likelihood). These particle approaches build a sequence of importance sampling steps between gradually tempered distributions evolving from the prior to the posterior PDF. To automate the definition of the tempering schedule, adaptive SMC (ASMC) allows tuning the temperature increments on-the-go. One general challenge in Bayesian inversions is the computational burden associated with expensive, high-fidelity forward solvers. Lower-fidelity surrogate models are interesting in this context as they can emulate the response of expensive forward solvers at a fraction of their cost. We consider surrogate modelling within ASMC and introduce first an approach involving surrogate modelling only, in which either prior samples are used to train the surrogate, or the surrogate model is retrained by updating the training set during the inversion. In our implementation, we rely on polynomial chaos expansions for surrogate modelling, principal component analysis for model parametrization and a ground-penetrating radar cross-hole tomography problem with either an eikonal or finite-difference time-domain solver as high-fidelity solver. We find that the method based on retraining the surrogate during the inversion outperforms the results obtained when only considering prior samples. We then introduce a computationally more expensive multifidelity approach including a transition to the high-fidelity forward solver at the end of the surrogate-based ASMC run leading to even more accurate results. Both methods result in speed-ups that are larger than one order of magnitude compared to standard high-fidelity ASMC inversion.

**Key words:** Ground penetrating radar; Inverse theory; Statistical methods; Tomography.

## 1 INTRODUCTION

Bayesian inversion provides a comprehensive characterization of the uncertainty associated with inferred model parameters (Sambridge & Mosegaard 2002). General formulations of Bayesian inversion problems can be computationally expensive to solve as implementations typically require a large number (possibly millions) of forward problem evaluations to obtain a sufficient number of independent samples from the posterior distribution. In addition to their computational costs, standard Markov chain Monte Carlo (MCMC) methods tend to perform rather poorly when dealing with high-dimensional parameter spaces and/or highly non-linear forward problems, failing to adequately explore the posterior probability density function (PDF).

Exploration can be enhanced by using tempering-based MCMC methods such as parallel tempering (Earl & Deem 2005) that relies on temperature-dependent posteriors, known as power posteriors. Tempering with high temperatures increases the exploration capacity; a property exploited in simulated annealing for global optimization (Kirkpatrick *et al.* 1983). Schemes based on tempering have been implemented to solve challenging geophysical Bayesian inversions, for example, geoacustic inversions (Yardim *et al.* 2010; Dettmer *et al.* 2010, 2011) and earthquake source inversions (Minson *et al.* 2013; Vasyura-Bathke *et al.* 2020). Annealed importance sampling (AIS, Neal 2001) and sequential Monte Carlo (SMC, Doucet & Johansen 2011) are two tempering-based methods that approximate the posterior distribution by a weighted sample of particle states. Both AIS and SMC perform a sequence of importance sampling steps targeting power posteriors with gradually decreasing temperatures. An important advantage of both methods with respect to MCMC methods is that they also allow estimating the evidence, the normalizing constant in Bayes' theorem and a crucial parameter in Bayesian model selection (Kass & Raftery 1995; Schöniger *et al.* 2014). The difference between the two methods is that SMC incorporates resampling of the particle population when the variance of the importance weights becomes high. To address the challenge of finding a suitable sequence of temperatures, Zhou *et al.* (2016) proposed an adaptive version of SMC that we will refer to as ASMC, which automatically adjusts the temperature decrease between adjacent power posteriors. Such adaptive SMC algorithms have recently been introduced and successfully applied in geophysical applications demonstrating performances superior to state-of-the-art MCMC methods (Amaya *et al.* 2021; Davies *et al.* 2023). Furthermore, ASMC has shown a stronger capacity to explore multi-modal posteriors (Amaya *et al.* 2022).

In this study, we consider forward solvers with different fidelity levels within ASMC to reduce the computational cost of the inversion. High-fidelity modelling is used herein to refer to schemes that achieve high-accuracy, but typically involve a high computational burden. On the other hand, low-fidelity models (also known as surrogate, meta- or proxy models) are mathematical tools that seek to emulate the behaviour of expensive high-fidelity forward solvers at negligible cost per run. The use of surrogate modelling for geophysical inversion purposes has been somehow limited despite their potential to greatly reduce the computational cost associated with Bayesian inversion (Linde *et al.* 2017; Wagner *et al.* 2021; Meles *et al.* 2022). Common types of surrogate models include Kriging or Gaussian process modelling (Santner *et al.* 2003), and polynomial chaos expansions (Xiu & Karniadakis 2002; Lüthen *et al.* 2022). Polynomial chaos expansion (PCE) is a type of surrogate that approximates models by their spectral representations in a suitable basis of polynomial functions (Xiu & Karniadakis 2002; Blatman & Sudret 2011), and has been successfully implemented within various Bayesian inversion frameworks (Marzouk *et al.* 2007; Marzouk & Xiu 2009; Wagner *et al.* 2020; Meles *et al.* 2022).

In Bayesian inversion, the combination of forward solvers with different fidelity levels has shown to be advantageous (Peherstorfer *et al.* 2018). For instance, evolving surrogate solvers within SMC has already been considered, for instance, in biology to predict dynamics of gene expression (Catanach *et al.* 2020). In solid mechanics, the combination of gradually increasing the spatial resolution in forward modelling and ASMC tempering was proposed by Koutsourelakis (2009), a scheme further automatized by Latz *et al.* (2018). The concept of multifidelity modelling was introduced in uncertainty quantification by Ng & Eldred (2012) and has been applied to different surrogate modelling frameworks (Park *et al.* 2017). We use the term multifidelity inversion to refer to Bayesian inversion strategies that accelerate the computation by using low-fidelity solvers, while ensuring accuracy and convergence by including a smaller number of simulations with the high-fidelity solver. For example, two-stage MCMC reserves the high-fidelity evaluations for samples that have been pre-accepted by the low-fidelity solver (Christen & Fox 2005). In this study, we propose an ASMC algorithm that operates within such a multifidelity framework.

The novelty of this work lies in the inclusion of surrogate modelling at different fidelity levels, including an ultimate transition to the high-fidelity solver, within ASMC. The method is initialized with a low-fidelity solver, trained by realizations from the prior, and updates the surrogate as the inversion progresses. To achieve this, high-fidelity simulations are gathered and appended to the training set cumulatively, such that the surrogate training set grows and contains more specified samples (more representative of the posterior PDF) as the temperature decreases. To account for the transitions between different fidelity levels, importance sampling steps are performed. In our multifidelity framework, the sequence of updated surrogates ends with a final transition to the high-fidelity solver. We test the method in a cross-hole ground-penetrating radar (GPR) tomography setting similar to Meles *et al.* (2022). We first consider an eikonal solver as the high-fidelity model, and we compare the inversion results when using (i) the high-fidelity solver only, (ii) a surrogate trained exclusively by realizations from the prior and (iii) a sequence of surrogate updates. We then evaluate the method using a more physically realistic and computationally demanding finite-difference time-domain (FDTD) solver. We demonstrate that our multifidelity ASMC inversion can successfully reproduce the results obtained by traditional high-fidelity inversion, but at a much lower computational cost.

## 2 METHOD

### 2.1 Bayesian inference and model selection

A probabilistic formulation of the inverse problem is offered by Bayes' theorem, which expresses the posterior PDF $\pi(\boldsymbol{\xi}|\mathbf{y})$ over certain parameters of interest $\boldsymbol{\xi}$ (e.g. pixel values of physical properties in a regular grid) given a set of observations $\mathbf{y}$:

$$\pi(\boldsymbol{\xi}|\mathbf{y}) = \frac{\pi(\boldsymbol{\xi})p(\mathbf{y}|\boldsymbol{\xi})}{\pi(\mathbf{y})}. \tag{1}$$

All the prior knowledge on the system is encapsulated in the prior PDF $\pi(\boldsymbol{\xi})$, and the likelihood function $p(\mathbf{y}|\boldsymbol{\xi})$ quantifies how likely it is that a given model realization gave rise to the observations given a prescribed error model. The normalizing constant $\pi(\mathbf{y})$, known as *evidence*, or *marginal likelihood*, quantifies the agreement between the observed data and the conceptual model, referring here to the parametrization of the subsurface together with its prior PDF (Brunetti *et al.* 2017). The evidence is needed to compare or rank different conceptual models (Kass & Raftery 1995). However, this quantity can be challenging to compute for high-dimensional problems, as it is a multidimensional integral over the prior (Schöniger *et al.* 2014):

$$\pi(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\xi})\pi(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}. \tag{2}$$

#### 2.1.1 *Model parametrization and model reduction*

When considering spatially correlated fields, the number of parameters to be inferred can be reduced using dimensionality-reduction techniques (e.g. Linde *et al.* 2015). The forward operator $\mathcal{F}(\boldsymbol{\xi})$ computes the observational output from a set of model parameters $\boldsymbol{\xi}$. If we consider a change of coordinates $\boldsymbol{\xi} = g(\boldsymbol{\theta}_{\mathrm{full}})$ and we assume there is no modelling error in $\mathcal{F}(\boldsymbol{\xi})$ or loss of information in the decomposition, we can formulate the forward problem as:

$$\mathbf{y} = \mathcal{F}(g(\boldsymbol{\theta}_{\mathrm{full}})) + \epsilon_y, \tag{3}$$

where $\mathbf{y}$ represents the observed data corresponding to the set of parameters $\boldsymbol{\theta}_{\mathrm{full}}$ and $\boldsymbol{\epsilon}_y$ the observational noise. The forward operator $\mathcal{F}(g(\boldsymbol{\theta}_{\mathrm{full}}))$ can be expressed as a function composition $\mathcal{M} = \mathcal{F} \circ g$. In practice, one achieves dimensionality-reduction by using a truncated subset of $\boldsymbol{\theta}_{\mathrm{full}}$ coordinates. For example, in this work we rely on principal component analysis (PCA Jolliffe & Cadima 2016). If the function composition over the truncated set of PCA coordinates $\theta$ adequately approximate the forward operator, then:

$$\mathbf{y} = \mathcal{M}(\boldsymbol{\theta}) + \epsilon_y + \epsilon_{\mathrm{PCA}}, \tag{4}$$

where $\boldsymbol{\epsilon}_{\mathrm{PCA}}$ is the error in the computed forward response associated with the truncated projection. In this study, we rely on surrogate forward models that, when dealing with a large number of parameters, require the use of dimensionality reduction techniques to be computationally feasible to implement (Torre *et al.* 2019). We consider a PCA decomposition that identifies the main features from a set of random realizations from the prior similarly to Meles *et al.* (2022). It has a simple implementation and has been used in many geoscientific inversions (Reynolds *et al.* 1996; Sarma *et al.* 2006; Kitanidis & Lee 2014; Jiang & Jafarpour 2021). However, the method proposed in this paper is not exclusively linked to this choice and alternative dimensionality reduction techniques could be considered.

## 2.2 Introduction to SMC

SMC methods (Chopin 2002; Del Moral *et al.* 2006) are a group of algorithms in which $N$ particles evolve in parallel towards an approximation of the posterior PDF. The un-normalized power posteriors $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$ are distributions in which the likelihood function is raised to the power of an inverse temperature $\alpha_t$:

$$\gamma_t(\boldsymbol{\theta}_t|\mathbf{y}) \equiv \pi(\boldsymbol{\theta}_t)p(\mathbf{y}|\boldsymbol{\theta}_t)^{\alpha_t}. \tag{5}$$

The SMC particles target in a sequential manner a series of power posteriors with gradually increasing inverse temperatures (that is, decreasing temperatures), starting with the prior PDF ($\alpha = 0$) and ending with the posterior PDF ($\alpha = 1$). To approximate the intermediate power posteriors, a set of $K$ intermediate MCMC steps are performed. After $K$ steps, a transition is made to the next power posterior (defined by an increment on the current $\alpha_t$) using an importance sampling step (IS Hammersley & Handscomb 1964). In this step, the approximation of the present power posterior forms the importance distribution for the next one, resulting in an incremental weight for each particle $i$ given by:

$$w_t^i = \frac{\gamma_t(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}{\gamma_{t-1}(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}, \tag{6}$$

where $\boldsymbol{\theta}_{t-1}^i$ is the state of the particle after the $K$ MCMC steps taken to approximate $\gamma_{t-1}$. These incremental weights are multiplied throughout the run, for each particle, in the form of normalized particle weights:

$$W_t^i = \frac{W_{t-1}^i w_t^i}{\sum_{j=1}^N W_{t-1}^j w_t^j}, \tag{7}$$

where $W_{t-1}^i$ is the normalized weight of the previous IS step. At the end of the run, the final particle states (each one is a set of model parameter values $\boldsymbol{\theta}$) with associated weights $W_T$ approximate the posterior PDF. Fig. 1 shows a graphical representation of the SMC method using $N = 4$ particles. Since the estimation is performed through a series of importance sampling steps, the estimates are unbiased regardless of the number of $K$ MCMC steps (e.g. there is no need to reach 'burn-in' as in MCMC methods). However, if using a too low $K$ or even $K = 0$, the variance of the estimates will be very large rendering the resulting estimates unreliable. For a detailed treatment, we refer to Neal (2001) and Del Moral *et al.* (2006).

### *2.2.1 Resampling*

SMC methods uses resampling (Doucet & Lee 2018), a re-organization of the particle states with replications that are proportional to the particle weights (red-dashed lines in Fig. 1). During resampling, the states of particles with high normalized weights are likely to be reproduced, whereas the states of particles with low normalized weights are likely to be discarded. After each resampling step, the normalized weights of all particles are re-initialized to $W_t^i = 1/N$. Resampling keeps the variance of the normalized particle weights from growing indefinitely by favouring particle trajectories with significant posterior probabilities. However, the resampling process increases the variance of the SMC estimates unnecessarily if it is performed too often (Douc & Cappé 2005). To make decisions on when to perform resampling, the effective sample size (*ESS*) is monitored throughout the run:

$$ESS_t = \frac{(\sum_{i=1}^N W_{t-1}^i w_t^i)^2}{\sum_{h=1}^N (W_{t-1}^h)^2 (w_t^h)^2}. \tag{8}$$

The *ESS* quantifies the effective number of particles in the particle approximation and when the $ESS_t/N \in [0, 1]$ gets lower than a threshold value $ESS^*/N$, a resampling step is performed. We rely on systematic resampling because of its simplicity and satisfactory performance (Doucet & Johansen 2011).

### *2.2.2 Evidence estimation*

The normalizing constant in Bayes' theorem (eq. 1) is the key quantity in Bayesian model selection and is obtained as a byproduct of the SMC method. Del Moral *et al.* (2006) show that the ratio of the normalizing constants of two consecutive power posteriors, $Z_t = \int \gamma_t(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}$ and $Z_{t-1} = \int \gamma_{t-1}(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}$, can be approximated as:

$$\frac{Z_t}{Z_{t-1}} \approx \sum_{i=1}^N W_{t-1}^i w_t^i. \tag{9}$$

As the prior PDF integrates to one ($Z_0 = 1$), the evidence (eq. 2) can be expressed as the product of consecutive normalizing constant ratios and, hence, approximated using the evolving particle weights as:

$$\pi(\mathbf{y}) = Z_T = \frac{Z_T}{Z_0} = \prod_{t=1}^T \frac{Z_t}{Z_{t-1}} \approx \prod_{t=1}^T \sum_{i=1}^N W_{t-1}^i {w_t}^i. \tag{10}$$

### *2.2.3 Adaptive sequential Monte Carlo (ASMC)*

Pre-selecting a suitable sequence of $\alpha_t$ is very challenging. The ASMC method by Zhou *et al.* (2016) aims at building a sequence of gradual intermediate IS steps bridging the prior and the posterior, such that the quality of the importance sampling steps quantified by the variance of the incremental weights in eq. (6), is controlled. In ASMC, the $\alpha_t$-increments are automatically defined on-the-go based on the conditional ESS (*CESS*):

$$CESS = N \frac{\left(\sum_{i=1}^N W_{t-1}^i w_t^i\right)^2}{\sum_{j=1}^N W_{t-1}^j \left(w_t^j\right)^2}. \tag{11}$$

In practice, a binary search is made over a range of possible increments to find the one that provides incremental weights $w_t$ that give the *CESS* that is the closest to a pre-defined value ($CESS_{\mathrm{op}}$). The $CESS_{\mathrm{op}}$ parameter controls the quality of the importance sampling steps and, hence, the quality of the resulting particle approximation. The $CESS_{\mathrm{op}}/N$ can take values between 0 and 1, the higher it is, the better the quality of the importance distribution, but the larger is the number of resulting power posteriors in the sequence $L$. As the relationship between $L$ and $CESS_{\mathrm{op}}$ is non-linear, it is challenging to recommend suitable $CESS_{\mathrm{op}}$ values. Results by Amaya *et al.*

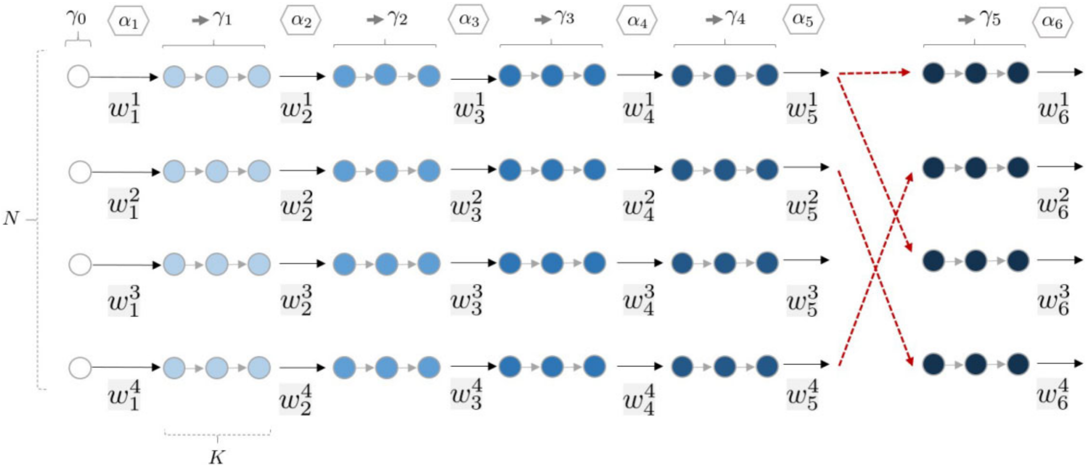

**Figure 1.** Schematic representation of SMC for $N = 4$ particles. The circles represent the particles evolving states with the initial samples drawn from the prior PDF (white circles) and $K = 3$ MCMC steps are used to approximate the power posteriors $\gamma_t$ with increasing inverse temperatures $\alpha_t$. Importance sampling steps are performed to transition between power posteriors, in which incremental weights $w_t^i$, contributing to the particle normalized weights $W_t^i$, are computed. The red-dashed lines indicate the re-organization of the particle states performed in a resampling step.

(2021) suggest that $CESS_{\text{op}}/N > =0.99$ provides a suitably smooth sequence. A derivation of eqs (8) and (11) can be found in Zhou *et al.* (2016).

In the $K$ MCMC steps used for each considered power posterior, the proposed models obtained, if using symmetric proposal distributions, are accepted or rejected according to the Metropolis rule (Metropolis *et al.* 1953). It states that the probability to accept the proposed candidate model $\boldsymbol{\theta}_p$ over the current one $\boldsymbol{\theta}_c$ is:

$$\Gamma(\boldsymbol{\theta}_p, \boldsymbol{\theta}_c) = \min\left(1, \frac{\pi(\boldsymbol{\theta}_p)p(\mathbf{y}|\boldsymbol{\theta}_p)}{\pi(\boldsymbol{\theta}_c)p(\mathbf{y}|\boldsymbol{\theta}_c)}\right). \tag{12}$$

We consider candidate particle states $\boldsymbol{\theta}_p$ that are generated by adding zero-mean and uncorrelated random Gaussian perturbations to the current state $\boldsymbol{\theta}_c$. The standard deviation of the proposal distribution is chosen to be proportional to the standard deviation of the prior, with the constant of proportionality $\phi$ determining the proposal scale. Compared with MCMC algorithms, an important advantage of ASMC is that $\phi$ can be tuned throughout the run to ensure that the acceptance rate *AR* is kept within a suitable range while still ensuring unbiased estimates. This is a consequence of the fact that the $K$ MCMC steps are only used to improve the quality of the importance sampling distribution.

### 2.3 ASMC with surrogate updates

This section focuses on the novel contribution of this work, which is a multifidelity framework allowing the inclusion of surrogate modelling in ASMC. The aim is to significantly reduce computational times compared to ASMC using high-fidelity solvers only while still obtaining approximations that are close to the results obtained in that case. If simply replacing a high-fidelity forward solver with a low-fidelity solver, the resulting posterior PDF is different from the original problem.

As the inverse temperature reduces the influence of the likelihood term, errors induced by a low-fidelity forward model are less significant at the early stages of the ASMC algorithm, implying that a less accurate surrogate can be sufficient. We propose to build a sequence of power posteriors that combines inverse temperature increments with updates of increasingly more accurate surrogate models as needed when the inverse temperature rises, while always including an estimation of the associated model error. By gradually improving the surrogate and ultimately transfer to the high-fidelity solver, we aim to decrease the difference between the posterior PDF-estimates obtained with those of using the high-fidelity forward solver only. We follow (Meles *et al.* 2022) and use PCE surrogate models (see Appendix A).

If we consider a sequence of surrogate solver updates $s_j$, with $j \in [1, J]$ and $J$ the number of total surrogate updates, for a power posterior $\gamma_t^{s_j}$, the likelihood is computed using the surrogate solver $\hat{\mathcal{M}}^{s_j}(\boldsymbol{\theta})$ and the covariance accounting for the three error sources in eq. (A6) (Appendix A). The covariances associated with the data noise $\mathbf{C}_y$ and the PCA projection $\mathbf{C}_{\text{PCA}}$ are then constant throughout the inversion, while the covariance associated to the PCE $\mathbf{C}_{\text{PCE}}$ is recalculated for each update of the surrogate model. Each time the surrogate is updated, we introduce an IS step to account for the transition between surrogates. It is performed to account for the change in the forward operator and $\mathbf{C}_{\text{PCE}}$. If the surrogate $s_j$ with covariance $\mathbf{C}_{\text{PCE}}^{s_j}$ is replaced with the updated surrogate $s_{j+1}$ with covariance $\mathbf{C}_{\text{PCE}}^{s_{j+1}}$, then the incremental weight resulting from the IS step will be:

$$w[s_j, s_{j+1}]_t^i = \frac{\gamma_t^{s_{j+1}}(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}{\gamma_{t-1}^{s_j}(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}. \tag{13}$$

Similarly to before, the $\alpha_t$ of the following power posterior in the sequence $\gamma_t^{s_{j+1}}$ can be optimized based on the *CESS*. The algorithm selects from a range of $\alpha_t = F\alpha_{t-1}$ with $F \in [0.1, 2]$, the one that provides the highest *CESS*. The incremental weights in eq. (13) contribute to the particle importance weights and evidence in the same way as the standard ASMC incremental weights in eq. (6). In contrast to those IS steps, we find that the optimal $\alpha_t$ is smaller than $\alpha_{t-1}$, because there is generally a reduction of the modelling errors when updating the surrogate.

To gain intuition about the impact of these different IS steps, it is useful to consider the multiplication of the incremental weights following Neal (2001, their eq. 5). If we consider an ASMC inversion in which the surrogate $s_0$ is unchanged, the sequence of incremental

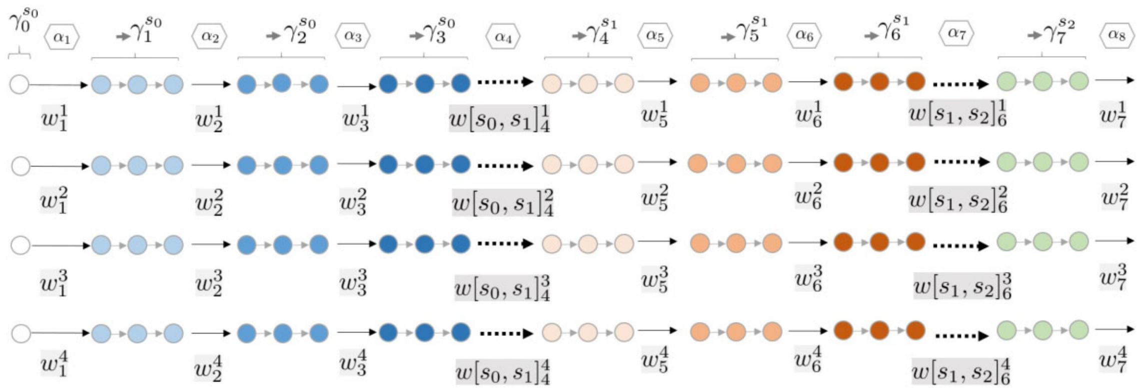

**Figure 2.** Modified schematic diagram of ASMC (Fig. 1) with surrogate updating (ASMC–SURR). The black-dashed lines represent a retraining of the surrogate together with the corresponding importance sampling step to transition from one surrogate to the following one. The change in the colour denotes a change of surrogate, whereas the change of the colour tone denotes a change of temperature when keeping the surrogate solver fixed.

weight ratios (eq. 6) for a particle is:

$$\prod_{t=1}^{T} w_t = \frac{\gamma_1^{s_0}(\boldsymbol{\theta}_0|\mathbf{y})}{\gamma_0^{s_0}(\boldsymbol{\theta}_0|\mathbf{y})}\frac{\gamma_2^{s_0}(\boldsymbol{\theta}_1|\mathbf{y})}{\gamma_1^{s_0}(\boldsymbol{\theta}_1|\mathbf{y})}\frac{\gamma_3^{s_0}(\boldsymbol{\theta}_2|\mathbf{y})}{\gamma_2^{s_0}(\boldsymbol{\theta}_2|\mathbf{y})}\frac{\gamma_4^{s_0}(\boldsymbol{\theta}_3|\mathbf{y})}{\gamma_3^{s_0}(\boldsymbol{\theta}_3|\mathbf{y})}\frac{\gamma_5^{s_0}(\boldsymbol{\theta}_4|\mathbf{y})}{\gamma_4^{s_0}(\boldsymbol{\theta}_4|\mathbf{y})}\frac{\gamma_6^{s_0}(\boldsymbol{\theta}_5|\mathbf{y})}{\gamma_5^{s_0}(\boldsymbol{\theta}_5|\mathbf{y})} \cdots \frac{\gamma_T^{s_0}(\boldsymbol{\theta}_{T-1}|\mathbf{y})}{\gamma_{T-1}^{s_0}(\boldsymbol{\theta}_{T-1}|\mathbf{y})}, \tag{14}$$

where we have removed the particle index $i$ to simplify the notation. Eq. (14) illustrates the bridge created by the IS steps; the sequence starts with the prior $\gamma_0$ as an importance distribution and ends with the posterior $\gamma_T$ as the target distribution. When considering surrogate updates, the sequence will also include the corresponding IS steps, for example:

$$\prod_{t=1}^{T} w_t = \frac{\gamma_1^{s_0}(\boldsymbol{\theta}_0|\mathbf{y})}{\gamma_0^{s_0}(\boldsymbol{\theta}_0|\mathbf{y})}\frac{\gamma_2^{s_0}(\boldsymbol{\theta}_1|\mathbf{y})}{\gamma_1^{s_0}(\boldsymbol{\theta}_1|\mathbf{y})}\frac{\gamma_3^{s_1}(\boldsymbol{\theta}_2|\mathbf{y})}{\gamma_2^{s_0}(\boldsymbol{\theta}_2|\mathbf{y})}\frac{\gamma_4^{s_1}(\boldsymbol{\theta}_3|\mathbf{y})}{\gamma_3^{s_1}(\boldsymbol{\theta}_4|\mathbf{y})}\frac{\gamma_5^{s_1}(\boldsymbol{\theta}_4|\mathbf{y})}{\gamma_4^{s_1}(\boldsymbol{\theta}_4|\mathbf{y})}\frac{\gamma_6^{s_2}(\boldsymbol{\theta}_5|\mathbf{y})}{\gamma_5^{s_1}(\boldsymbol{\theta}_5|\mathbf{y})} \cdots \frac{\gamma_T^{s_J}(\boldsymbol{\theta}_{T-1}|\mathbf{y})}{\gamma_{T-1}^{s_J}(\boldsymbol{\theta}_{T-1}|\mathbf{y})}, \tag{15}$$

where the third and sixth ratio are associated with the transition between different surrogates (eq. 13). To update the surrogate, we collect samples of the high-fidelity response as the ASMC sequence progresses. These samples are used to retrain the PCE surrogate after a pre-defined number of inverse temperature updates $T_{\mathrm{up}}$. The training set grows as the samples are added to the previous training set. As ASMC targets power posteriors that gradually approach the posterior PDF, the sampling gets more focused in high-likelihood regions of the parameter space. Consequently, the surrogate will not only perform better due to the increased size of the training set, but it will also become more accurate in regions of the parameter space having significant posterior probabilities. A schematic representation of ASMC incorporating surrogate updates can be found in Fig. 2; we refer to this method as ASMC–SURR.

### 2.3.1 *Multifidelity modelling by transitioning to the high-fidelity solver*

The ASMC–SURR algorithm progresses while improving the surrogate until $\alpha_t = 1$. At this point, it provides an approximation of the posterior PDF and the evidence associated with the last surrogate used in the sequence. One can finish the inversion with these approximations (that is, the ASMC–SURR method), or go further by considering a switch to the original high-fidelity solver (indicated in the following by HF). The incremental weight resulting from this step considering one particle is:

$$w[s_J, HF]_t^i = \frac{\gamma_t^{\mathrm{HF}}(\boldsymbol{\theta}_{t-1}|\mathbf{y})}{\gamma_{t-1}^{s_J}(\boldsymbol{\theta}_{t-1}|\mathbf{y})}, \tag{16}$$

where $s_J$ is the last surrogate of the sequence. In this step the inverse temperature $\alpha_{t-1}$, corresponding to the power posterior in the denominator $\gamma_{t-1}^{s_J}$ is equal to 1. The inverse temperature $\alpha_t$ associated to $\gamma_t^{\mathrm{HF}}$ can again be optimized. We seek an optimal value from $\alpha_t = F_{\mathrm{HF}}\alpha_{t-1}$ with $F_{\mathrm{HF}} \in [0.1, 1]$, such that it provides the highest possible *CESS*. We refer to $F_{\mathrm{HF}}$ as the $\alpha$-correction. Subsequently, the ASMC algorithm proceeds with high-fidelity updates until $\alpha_t = 1$ is reached. We refer to this algorithm as ASMC–SURR–HF. This is a multifidelity inversion method in the sense of Peherstorfer *et al.* (2018), in that it targets the same posterior and evidence as if one would always use the high-fidelity solver. A flow chart describing the basics of both ASMC–SURR and ASMC–SURR–HF are found in Fig. 3. A detailed description of the two algorithms is given in Appendix B.

### 2.3.2 *Performance assessment*

In our test cases, performing one of the $K$ MCMC steps using surrogate solvers compared to using the high-fidelity solver implies a gain in computational time of 10–1000 times. To assess the computational demand of the ASMC–SURR inversions, we assume that the dominant parameter is the required number of simulations using the high-fidelity solver (*HF sim*) and that the computational cost of the surrogate solver is negligible. We consider also the relative number of *HF sim* required for ASMC–SURR compared to the number of *HF sim* required when always using the high-fidelity solver (*HF sim relative*).

To evaluate the accuracy of the posterior estimates, we rely on the logarithmic scoring rule (Good 1992; Krüger *et al.* 2021). For each estimated parameter, the logarithmic score is the negative logarithm of the estimated posterior PDF evaluated at a reference (true) value. When comparing two posterior estimates, the one with the lower score is preferred. A kernel density estimate of the posterior samples is used to approximate the posterior univariate distributions; we use a Gaussian kernel with a bandwidth smoothing window calculated using the Scott's rule (Scott 2015). We consider as a performance metric the average logarithmic score of the model parameters (*Mean log-score*). We also present the structural similarity index measure (*SSIM*, Wang *et al.* 2004), that quantifies the similarity between two images (it can vary between −1 to 1, with 1

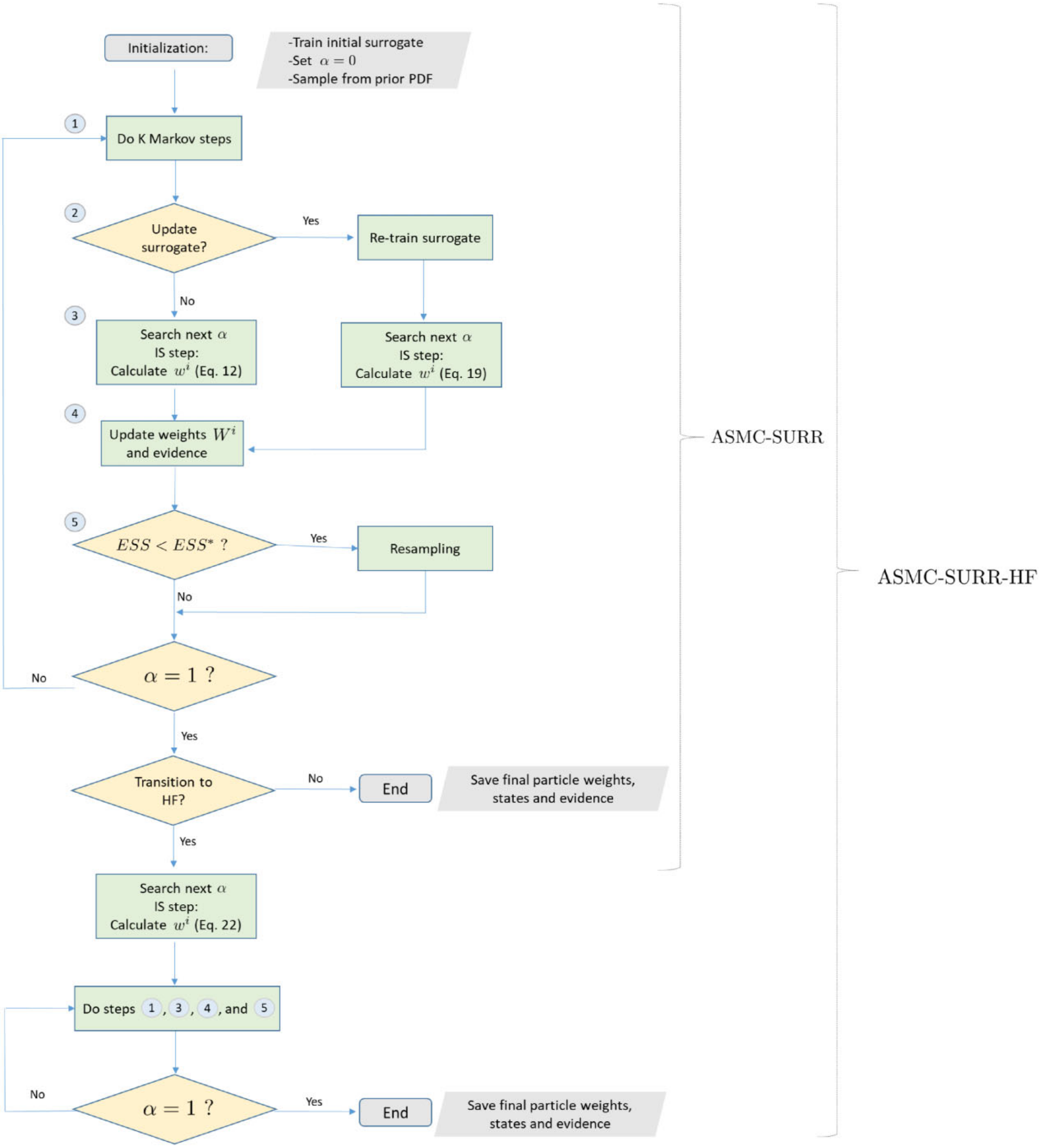

**Figure 3.** Simplified flow chart representing the main steps to perform ASMC inversion with surrogate updates with (ASMC–SURR–HF) and without (ASMC–SURR) transitioning to the high-fidelity solver. After the initialization, the sampling of the adaptive sequence of power posteriors at gradually increasing $\alpha_s$ starts, with importance sampling steps performed between consecutive power posteriors (see Section 2.2.3). The surrogate can be updated followed by an importance sampling step to transition between surrogates (see Section 2.3). If the *ESS* falls below the threshold *ESS**, resampling is performed (see Section 2.2.1). The particle normalized weights and evidence (see Section 2.2.2) are updated after the importance sampling steps. When $\alpha = 1$, we can chose between finishing the run (ASMC–SURR) or transitioning to the high-fidelity solver (ASMC–SURR–HF, see Section 2.3.1).

being the highest similarity). We consider the SSIM with respect to the posterior mean model of the full high-fidelity inversion. On the output domain, we compute the weighted mean root-mean-squared-error between the data and the output of the final particle states (*RMSE output*).

## 3 RESULTS

### 3.1 Test case

To evaluate the ASMC–SURR and the ASMC–SURR–HF methods, we consider a synthetic GPR crosshole tomography problem similar

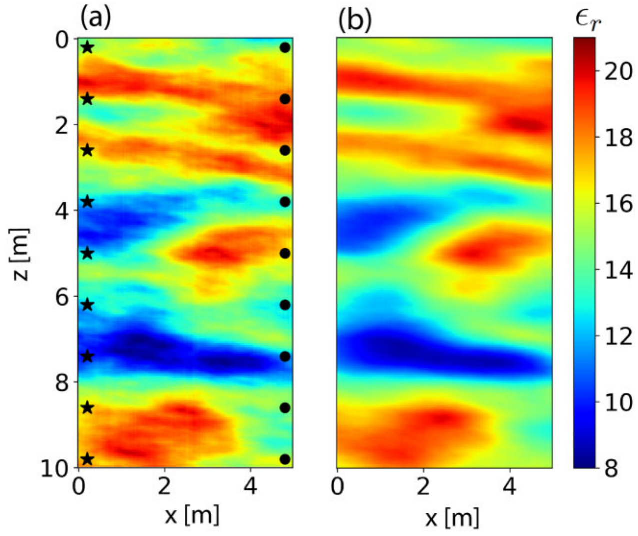

**Figure 4.** (a) Reference model and (b) its 100-PCA representation back-projected into the relative permittivity domain. The GPR sources are represented in (a) with black stars and the receivers with black circles.

to Meles *et al.* (2022) with subsurface relative dielectric permittivity $\epsilon_r$ variations and constant electrical conductivity (Fig. 4a). The dimension of the $\epsilon_r$ models is 125 × 250 with a cell discretization of $dx = dz = 0.04$ m, which results in 5 × 10 m models. The simulated GPR first-arrival traveltimes are obtained using 9 sources and 9 receivers spaced 4.6 m apart in the $x$-direction and 0.6 m in the $z$-direction. Only observations for which the angle between sources and receivers is lower than 45° are considered (Peterson 2001) such that the data set is composed of 69 first-arrival traveltimes. The covariance structure is based on the 2-D Matérn geostatistical model (Dietrich & Newsam 1997; Laloy *et al.* 2015) with a mean of 15, a standard deviation of 2.45, an anisotropy ratio of 0.3, an anisotropy angle of 85°, an integral scale of the major axis of 10 m and a shape parameter of 1.15.

#### *3.1.1 PCA representation*

Following Meles *et al.* (2022), the PCA decomposition is learned from a set of 1000 samples drawn from the generative prior model. Figs 5(a)–(d) show random realizations of the generative model. The number of principal components defines the resolution and the maximum achievable level of reconstruction of the original features. The dimension of the inverse problem is reduced as the inversion is performed in the PCA domain. We rely on 100 principal components to parametrize the input domain; these type of models are in what follows indicated as 100-PCA. The back-projection of the 100-PCA representation of the reference model is shown in Fig. 4(b). Figs 5(e)–(h) show the resulting 100-PCA projection of the model realizations in Figs 5(a)–(d), back projected into the relative permittivity domain.

### 3.2 Inversion

#### *3.2.1 ASMC setting*

We perform ASMC inversions using $N = 50$ particles, $K = 500$ MCMC steps to approximate each power posterior and an optimal conditional effective sample size of $CESS_{\mathrm{op}}/N = 0.99$. When the acceptance rate gets lower than $AR_{\min} = 15$ per cent at $\alpha_{t-1}$, the proposal scale $\phi$ is reduced a factor of $f = 20$ per cent for $\alpha_t$. Resampling is performed when $ESS/N < 0.3$ (Del Moral *et al.* 2006; Amaya *et al.* 2022).

To evaluate the performance of ASMC–SURR and ASMC–SURR–HF, we first explore cases for which the high-fidelity solver is an eikonal forward solver. This choice is made as it is then computationally feasible to estimate the posterior PDF and the evidence using the high-fidelity solver only, thereby, allowing for comparisons with a reference solution. This reference solution is termed ASMC$_{\mathrm{eikonal}}$ and it uses the *time2d* algorithm by Podvin & Lecomte (1991) for forward computations; a finite-difference approximation of the eikonal equation. We then consider the case when we train a surrogate solver using exclusively model realizations drawn from the prior and their corresponding high-fidelity eikonal simulations (ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{prior}}$). This configuration is also evaluated in the multifidelity setting (ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{prior}}$–HF). We compare these results to those obtained when four retraining steps of the surrogate are performed (ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{update}}$) and with a subsequent transition to the high-fidelity solver (ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{update}}$–HF). To allow for a fair comparison, we consider an equal total budget of high-fidelity simulations for training regardless if prior training only or retraining is allowed. The final number of high-fidelity simulations needed for ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{prior}}$–HF and ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{update}}$–HF after switching to the high-fidelity solver depends on the resulting optimized $\alpha$-correction $F_{\mathrm{HF}}$ (see Section 2.3.1). Finally, we consider results obtained when the high-fidelity forward solver is made up of a computationally much more expensive finite-difference time-domain solver (the simulations take 1000 times more time than for the PCE). Table 1 shows the resulting number of power posteriors, high-fidelity and surrogate evaluations for each test.

#### *3.2.2 PCE surrogate training*

The PCE surrogates are trained to learn the relationships between the PCA coefficients and the high-fidelity first-arrival traveltimes. The inversion is performed over the 100 principal components used to parametrize the input domain. Hence, the training set for the PCEs is composed of input models in the 100-PCA space and the corresponding eikonal first-arrival traveltimes computed on the 100-PCA back-projected relative permittivity models as output. In ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{prior}}$, a set of 1000 prior input–output pairs are used to train the surrogate. In ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{updates}}$, the same total budget of 1000 high-fidelity simulations is used. From these 1000, a set of $\tau_0 = 200$ prior input–output pairs are used to train the initial surrogate $s_0$. We consider four surrogate updates ($s_1$, $s_2$, $s_3$, $s_4$) retrained at regular intervals of $T_{\mathrm{up}} = 35$ inverse temperatures. To reach the 1000 high-fidelity budget, 200 new high-fidelity input–output pairs are gathered and appended to the training set matrices during each interval, such that the increasing training set sizes of the four surrogate updates are $\tau_k = 400, 600, 800, 1000$ (see Appendix B for details on the implementation).

#### *3.2.3 Error covariance matrices*

The data used for the various inversions are obtained by simulating the high-fidelity first-arrival traveltimes of the full reference model (Fig. 4a), and then contaminating them with uncorrelated Gaussian noise with a standard deviation of $\sigma = 0.5$ ns. We consider a Gaussian likelihood function (eq. A4) with a covariance matrix that incorporates the different error sources (see Appendix A, eq. A5).

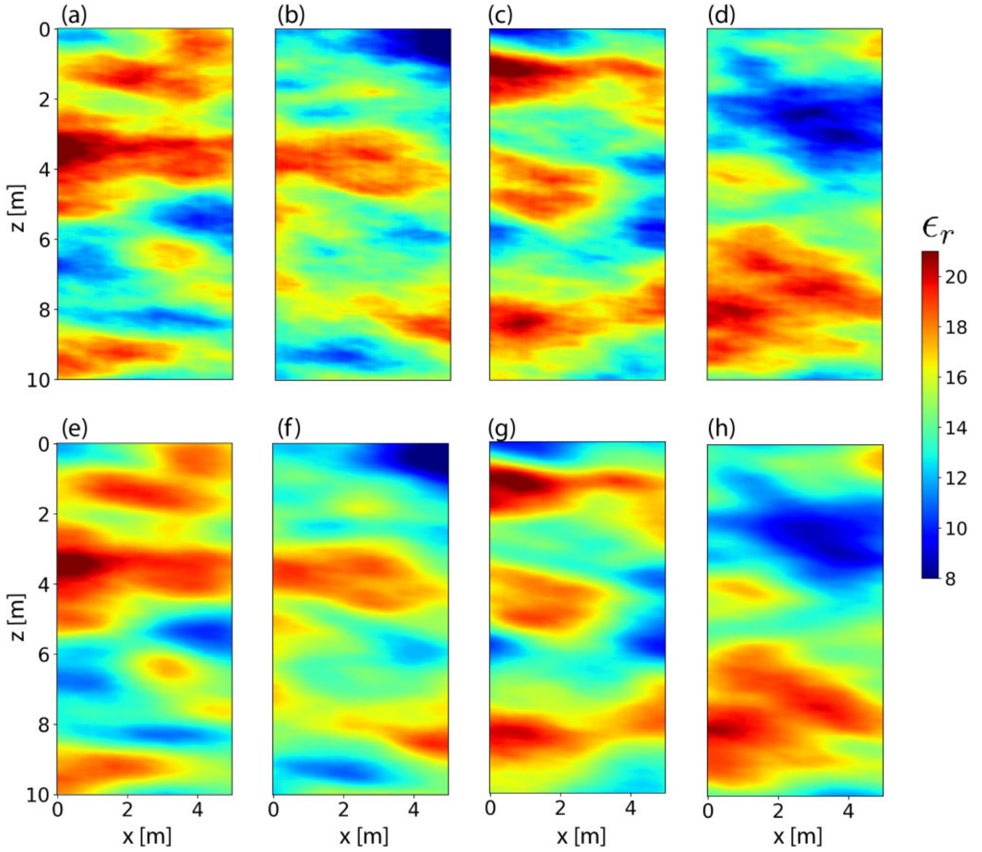

**Figure 5.** (a)–(d) Random samples obtained with the generative prior model and (e)–(h) their 100-PCA representation back-projected into the relative permittivity domain.

**Table 1.** Lengths of the ASMC runs and their performance metrics. The total number of MCMC steps per particle is given by the number of power posteriors ($L$, number of inverse temperatures) arising from the adaptive nature of the algorithm (Section 2.2.3) multiplied by the $K$ MCMC steps used to approximate each power posterior. All runs are performed using $N = 50$, $CESS_{\mathrm{op}}/N = 0.99$ and $K = 500$ (except for the reduction to $K = 50$ in the high-fidelity stage of ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{updates}}$–HF).

| | ASMC$_{\mathrm{eikonal}}$ | ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{prior}}$ | ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{prior}}$–HF | ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{updates}}$ | ASMC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{updates}}$–HF | ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{updates}}$ | ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{updates}}$–HF |
|---|---|---|---|---|---|---|---|
| $L$ | 192 | 173 | 251 | 180 | 203 | 178 | 191 |
| *HF sim*[$\times 10^3$] | 4800 | 1 | 1950 | 1 | 576 | 1 | 33.5 |
| *HF sim relative* | 100 per cent | 0.02 per cent | 40.63 per cent | 0.02 per cent | 12 per cent | - | - |
| *PCE sim*[$\times 10^3$] | 0 | 4325 | 4325 | 4500 | 4500 | 4450 | 4450 |
| *Resampling* | 2 | 2 | 4 | 2 | 3 | 2 | 3 |
| *SSIM* | 1 | 0.84 | 0.92 | 0.91 | 0.93 | - | - |
| *Mean σ* | 0.82 | 0.83 | 0.82 | 0.81 | 0.78 | 0.86 | 0.68 |
| *Mean log-score* | 1.46 | 1.73 | 1.50 | 1.60 | 1.52 | 1.23 | 1.09 |
| *RMSE output* | 0.52 ns | 0.84 ns | 0.48 ns | 0.65 ns | 0.49 ns | 0.72 ns | 0.55 ns |
| *Log-evidence* | −110.8 | −121.3 | −123.7 | −114.3 | −116.5 | −122.7 | −132.1 |

Figs 6(a) and (b) show the data error covariance matrix $\mathbf{C}_y$ and the PCA projection error covariance matrix $\mathbf{C}_{\mathrm{PCA}}$, respectively. The covariance matrix used in the likelihood calculation is the sum of the contributions $\mathbf{C}_y$ and $\mathbf{C}_{\mathrm{PCA}}$ in the cases cases when the forward simulations are performed using the high-fidelity solver (ASMC$_{\mathrm{eikonal}}$). Fig. 6(c) shows the covariance matrix of the PCE errors $\mathbf{C}_{\mathrm{PCE}}$ for the surrogate used in AMSC$_{\mathrm{eikonal}}$–SURR$_{\mathrm{prior}}$. When performing simulations using the PCE surrogate, we assume the covariance matrix in the likelihood function to be the sum of the three contributions $\mathbf{C}_y$, $\mathbf{C}_{\mathrm{PCA}}$ and $\mathbf{C}_{\mathrm{PCE}}$. The diagonals of the covariance matrices for this case are shown in Fig. 6(d). The lower the number of principal components used, the higher the errors associated to the PCA projection. The choice of 100 principal components was made to ensure that the influence of the PCA projection errors in the likelihood function is negligible compared to the data and surrogate modelling error (see Fig. 6a).

### 3.2.4 *High-fidelity reference solution with eikonal solver*

The inversion using the high-fidelity forward solver (ASMC$_{\mathrm{eikonal}}$) resulted in $L = 192$ power posteriors, a total number of 4.8 million high-fidelity forward simulations and two resampling steps (Table 1). Figs 7(a) and (h) show the corresponding weighted posterior mean and standard deviation, respectively, while Fig. 8(a) shows

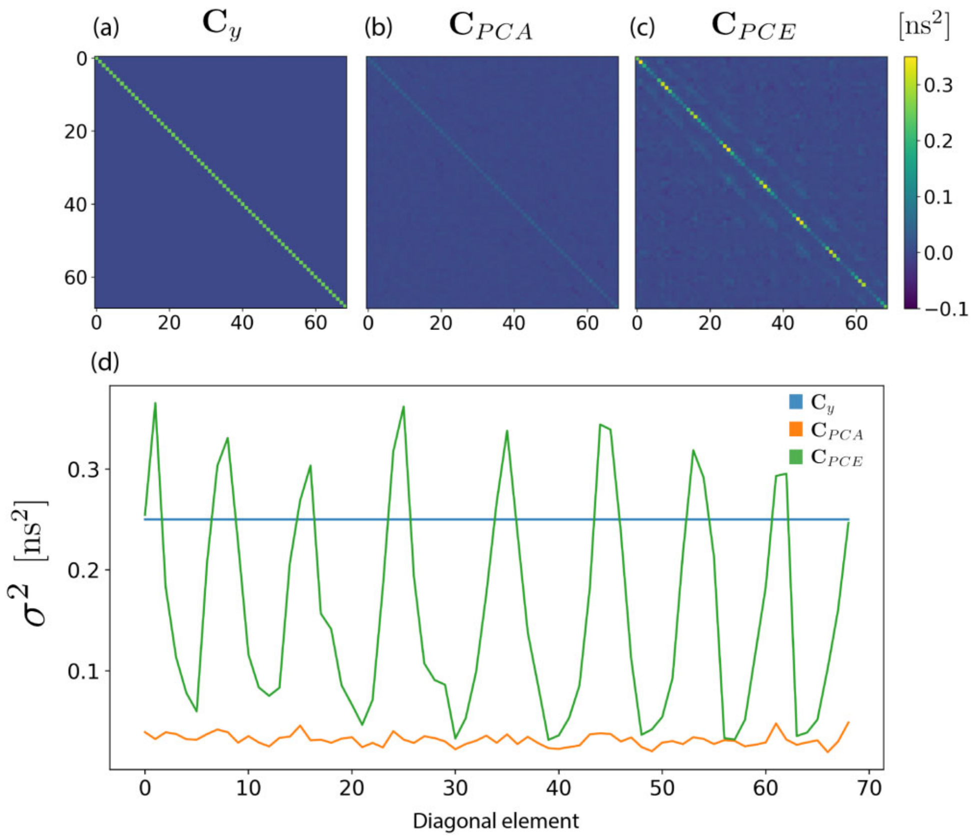

**Figure 6.** (a) Data error covariance matrix, (b) PCA projection error covariance matrix, (c) PCE surrogate modelling error covariance matrix for a surrogate trained using 1000 prior samples and (d) their diagonal elements.

the logarithmic score values on the relative permittivity domain. Table 1 provides the performance metrics: the *SSIM* equals one as we consider this posterior mean as the reference for the following tests, the mean of the standard deviation *Mean* $\sigma$, the *Mean log-score*, the *RMSE output* and the *Log-evidence* estimation. The *RMSE output* is 0.516 ns, which is close to the 0.5 ns data noise level (Fig. 6). The evidence estimate is −110.81.

### *3.2.5 Surrogate-based ASMC solutions with eikonal high-fidelity solver*

Figs 7(b)–(e) and (g)–(j) show the posterior means and standard deviations for the considered ASMC–SURR and ASMC–SURR–HF runs. The logarithmic score values are shown in Figs 8(b)–(e). The resemblance among the mean images and with respect to the reference solution suggest that the posterior mean approximations are overall of similar quality. By focusing on the performance metrics given in Table 1, we see for the surrogate-only inversions (ASMC–SURR) that updating the surrogates during the inversion ($\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{update}}$) outperforms the case when the surrogate is based on prior realizations only ($\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{prior}}$). Comparing these two cases, it is seen that the *SSIM* is 0.91 versus 0.84, the *Mean log-score* 1.6 versus 1.73 and the *RMSE output* 0.65 ns versus 0.84 ns. Both runs underestimate the evidence when compared to the high-fidelity inversion, but the $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{update}}$ approximation is closer to the reference (−114.3 versus 121.3). For an equal budget, this suggests that when high-fidelity simulations are only used to train the surrogate, there is a notable improvement offered by retraining the surrogate as the ASMC inversion progresses instead of only performing training using samples from the prior.

When considering the multifidelity approach involving a transition to the high-fidelity solver (ASMC–SURR–HF), Table 1 shows that there are significant improvements in the values of *SSIM*, *Mean log-score* and *RMSE output* compared to the ASMC–SURR results (e.g. the *RMSE output* decreases from 0.84 ns to 0.48 ns, and from 0.65 ns to 0.49 ns). On the other hand, the metrics in Table 1 suggest that $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{prior}}$–HF and $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{update}}$–HF results provide very similar results. The main difference between the two inversions is that the $\alpha$-correction factor is $F_{\mathrm{HF}} = 0.53$ for $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{update}}$–HF compared to $F_{\mathrm{HF}} = 0.15$ for $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{prior}}$–HF. This implies that the high-fidelity stage of the inversion starts at $\alpha = 0.53$ and $\alpha = 0.15$, respectively. This translates in a computing cost of the high-fidelity stage that is more than three times higher for $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{prior}}$–HF, as many more high-fidelity solutions are needed to reach $\alpha = 1$. This clearly shows that the price to pay for transitioning to the high-fidelity forward solver with a poor surrogate is reflected in a low corresponding $\alpha$-correction factor $F_{\mathrm{HF}}$. Hence, updating the surrogate sequence results in a smoother transition to the high-fidelity solver and, hence, lower computing times.

Figs 9(a)–(c) show the evolution of *CESS*/*N* (eq. 11) for $\mathrm{ASMC_{eikonal}}$, $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{prior}}$–HF and $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{update}}$–HF, respectively. The values are close to the optimal $CESS_{\mathrm{op}} = 0.99$ except for the IS steps used to transition between forward solvers when the *CESS*/*N* is much lower. The diagonal elements of the modelling error covariance matrix $\mathbf{C}_{\mathrm{PCE}}$ are shown in Fig. 10(a). As the training set grows in $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{update}}$–HF, the mean error level in the $\mathbf{C}_{\mathrm{PCE}}$ diagonals decreases. After

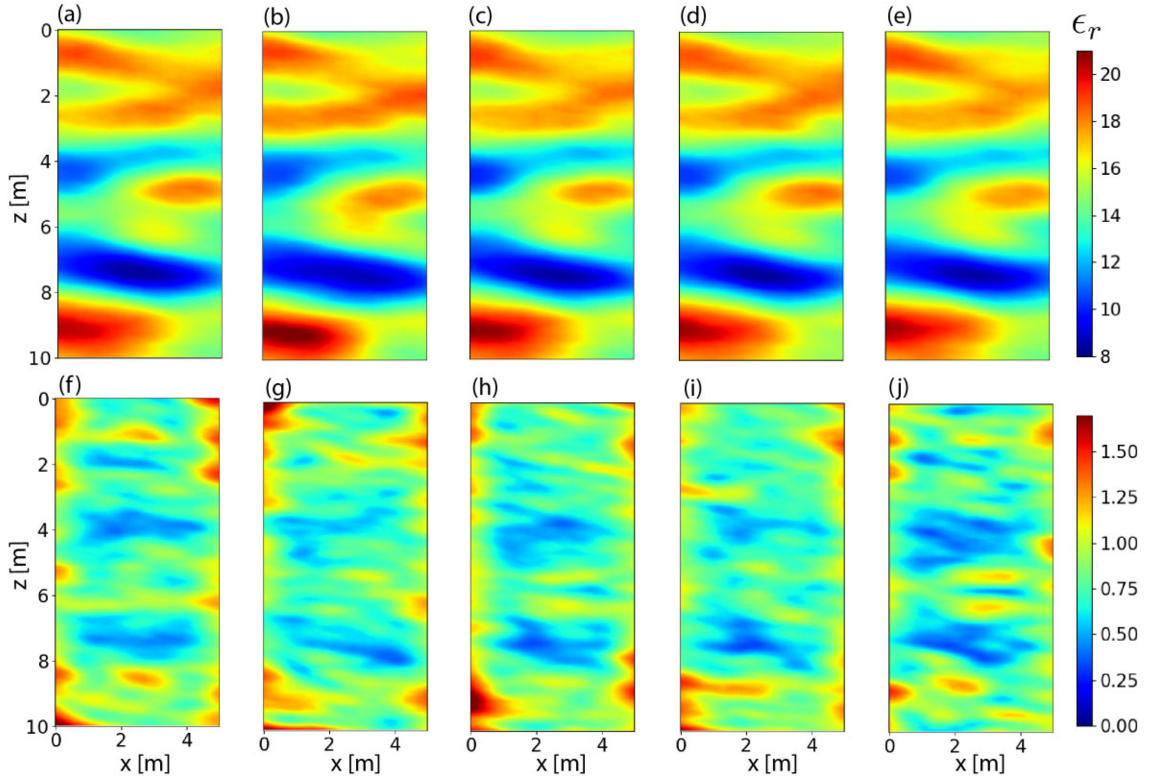

**Figure 7.** Posterior weighted mean model and standard deviation in, respectively, (a) and (f) $\mathrm{ASMC}_{\mathrm{eikonal}}$, (b) and (g) $\mathrm{ASMC}_{\mathrm{eikonal}}$–$\mathrm{SURR}_{\mathrm{prior}}$, (c) and (h) $\mathrm{ASMC}_{\mathrm{eikonal}}$–$\mathrm{SURR}_{\mathrm{prior}}$–HF, (d) and (i) $\mathrm{ASMC}_{\mathrm{eikonal}}$–$\mathrm{SURR}_{\mathrm{update}}$, (e) and (j) $\mathrm{ASMC}_{\mathrm{eikonal}}$–$\mathrm{SURR}_{\mathrm{updates}}$–HF. Note that the standard deviations only consider those associated with the first 100 PCA components that are considered in the inversions (see Meles *et al.* 2022).

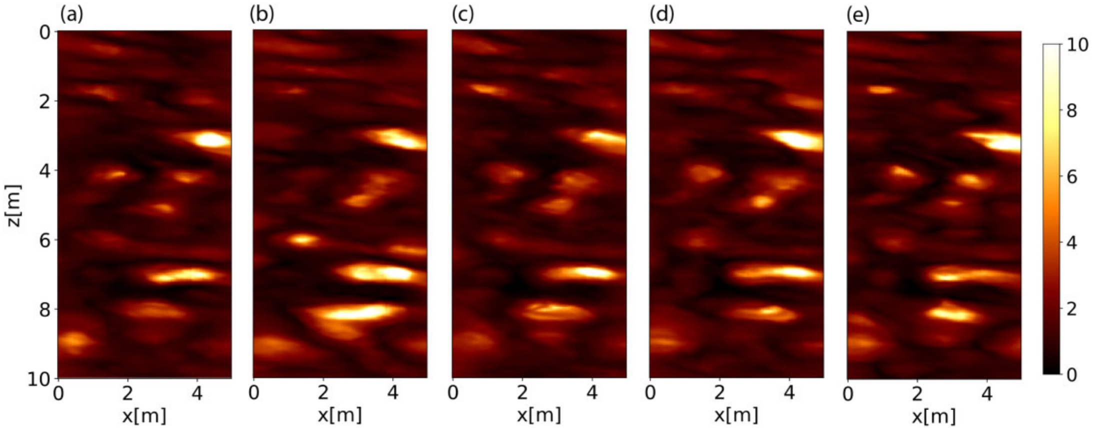

**Figure 8.** Logarithmic score values for (a) $\mathrm{ASMC}_{\mathrm{eikonal}}$, (b) $\mathrm{ASMC}_{\mathrm{eikonal}}$–$\mathrm{SURR}_{\mathrm{prior}}$, (c) $\mathrm{ASMC}_{\mathrm{eikonal}}$–$\mathrm{SURR}_{\mathrm{prior}}$–HF, (d) $\mathrm{ASMC}_{\mathrm{eikonal}}$–$\mathrm{SURR}_{\mathrm{updates}}$, (e) and (l) $\mathrm{ASMC}_{\mathrm{eikonal}}$–$\mathrm{SURR}_{\mathrm{updates}}$–HF.

the third update, the surrogate presents a lower error level than the surrogate trained using prior samples.

### *3.2.6 Surrogate-based ASMC solutions with FDTD high-fidelity solver*

To demonstrate the gain offered when considering more realistic and computationally more expensive high-fidelity forward solvers than the eikonal solver, we consider now a high-fidelity two-dimensional finite-difference time-domain (FDTD) forward solver simulating propagation in the transverse-electric mode (Irving & Knight 2006). Following Meles *et al.* (2022), automatic determination of traveltimes is performed by applying a threshold based on the relative maximum amplitude of each source-receiver pair. Running the ASMC inversion using only this expensive solver would require several months of computing (one parallelized simulation of 50 particles takes around 3.5–5 min); see also (Hunziker *et al.* 2019). We test our proposed method using FDTD as the high-fidelity solver ($\mathrm{ASMC}_{\mathrm{FDTD}}$–$\mathrm{SURR}_{\mathrm{update}}$) while keeping the same ASMC inversion setting as for the $\mathrm{ASMC}_{\mathrm{eikonal}}$–$\mathrm{SURR}_{\mathrm{update}}$ case. The inversion needed 178 power posteriors and two resampling steps (Table 1). The correction $F_{\mathrm{HF}}$ for the transition to the high-fidelity stage is 0.73,

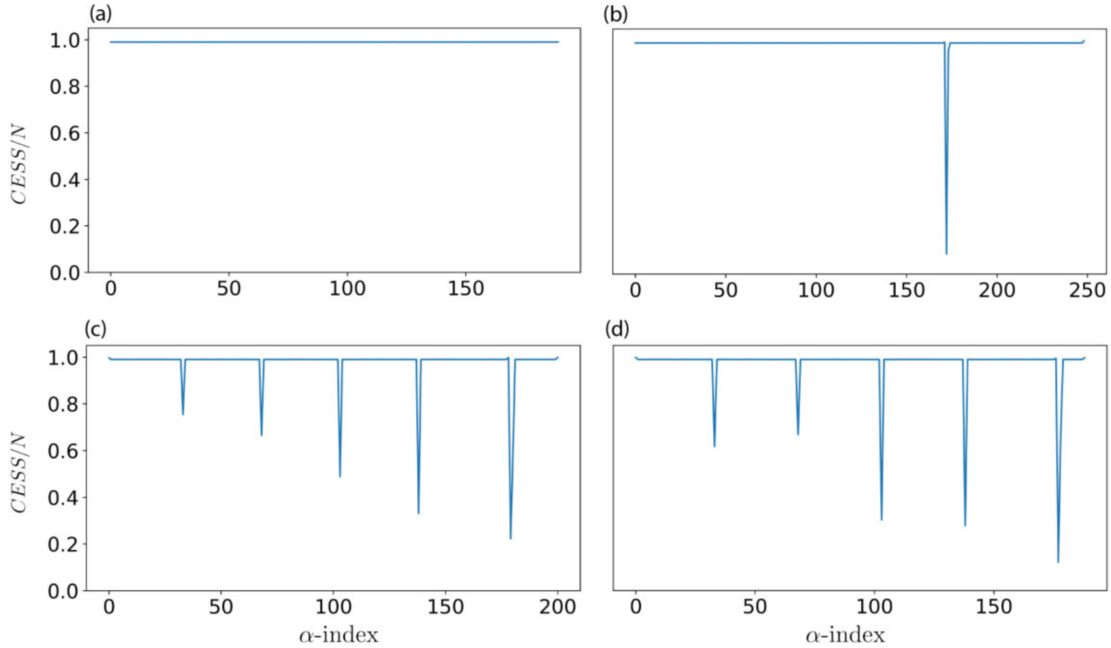

**Figure 9.** Conditional effective sample size over the number of particles (*CESS*/*N*) versus inverse temperature index for (a) $\mathrm{ASMC_{eikonal}}$, (b) $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{prior}}$–HF, (c) $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{update}}$–HF and (d) $\mathrm{ASMC_{FDTD}}$–$\mathrm{SURR_{update}}$–HF. The low values correspond to *CESS*-decreases occurring when updating the surrogate model or when transitioning to the high-fidelity forward solver.

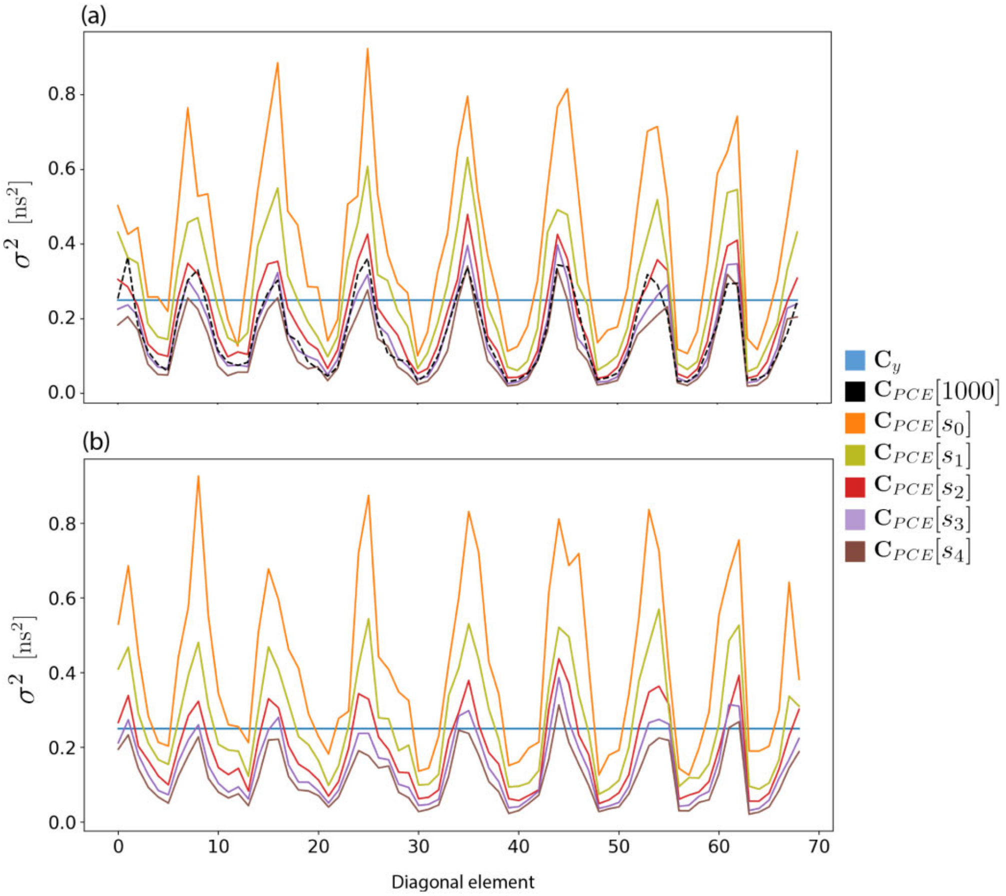

**Figure 10.** Evolution of the diagonal of the covariance modelling error matrix $\mathbf{C}_{\mathrm{PCE}}$ for the four surrogate updates in (a) $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{update}}$ and (b) $\mathrm{ASMC_{FDTD}}$–$\mathrm{SURR_{update}}$. The initial surrogate $s_0$ is trained using 200 samples, and 200 more are added to the training set in each update, such that the fourth surrogate update $s_4$ is trained using 1000 samples. The blue horizontal line corresponds to the diagonal of the data error covariance matrix $\mathbf{C}_y$, and the black-dashed line in (a) is the $\mathbf{C}_{\mathrm{PCE}}$ diagonal of the eikonal surrogate trained with 1000 prior samples used in $\mathrm{ASMC_{eikonal}}$–$\mathrm{SURR_{prior}}$.

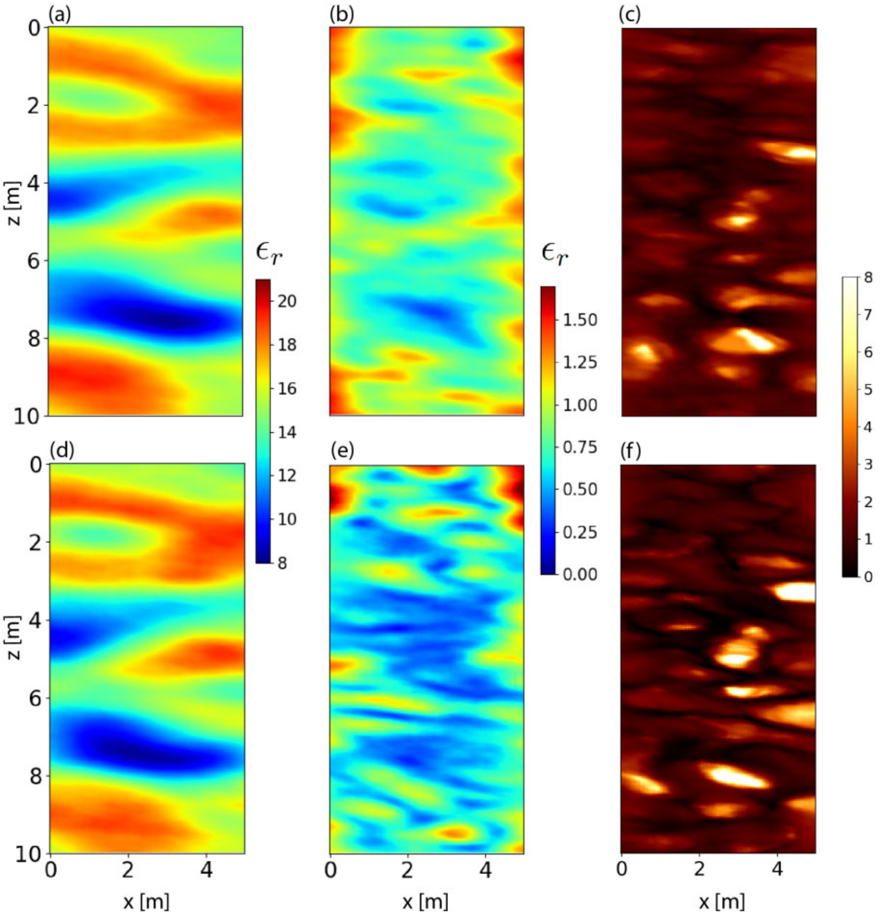

**Figure 11.** Posterior weighted mean model for (a) ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$ and (d) ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$–HF, posterior standard deviation for (b) ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$ and (e) ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$–HF, and logarithmic score value for (c) ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$ and (f) ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$–HF. Note that the standard deviations only consider those associated with the first 100 PCA components (see Meles *et al.* 2022).

which resulted in only 13 extra inverse temperatures (ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$–HF). Performing $K = 500$ Markov steps to approximate the 13 power posteriors is still very computationally demanding. In ASMC it is possible to decrease the $K$, with the consequence of having a less appropriate importance distribution. By choosing $K = 50$ for the steps involving HF makes it feasible to run the high-fidelity stage of the inversion. Fig. 11 shows the posterior weighted mean, standard deviation and logarithmic score values for ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$ and ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$–HF. Fig. 10(b) shows the evolution of the covariance matrices in the PCE updates and Fig. 11 shows the *CESS*/*N* values with the low values corresponding to transitions between forward solvers, respectively. The performance metrics in Table 1 show the improvement achieved by adding the high-fidelity stage. For example, the *Mean log-score* decreases from 1.23 to 1.09, and the *RMSE output* from 0.72 to 0.55 ns. Even if the results are not completely comparable with the eikonal cases, as we are using different data and forward solver, the metrics are close and even better than for the eikonal reference test (e.g. the *SSIM* for ASMC$_{\mathrm{FDTD}}$–SURR$_{\mathrm{update}}$–HF is 0.638 compared to 0.62 for ASMC$_{\mathrm{eikonal}}$ and the *Mean log-score* is 1.09 compared to 1.46). The evidence estimations are different than for the eikonal case, but this is expected as the high-fidelity forward solvers used to create the observed data are different. In the FDTD case, it is difficult to say anything conclusive as we do not have a reference value of the evidence using the high-fidelity solver only.

## 4 DISCUSSION

Over the past 20 yr, surrogate modelling has emerged as a prominent computational paradigm due to the massive reduction in computational costs achieved by approximating complex forward models with simple and easy-to-estimate functions. The accuracy of surrogate modelling critically depends on the training set used to learn the input–output relationship, and should ideally be lower than the data noise level. The construction of a globally accurate surrogate applicable to any prior realization can be challenging. However, surrogate accuracy can be refined in regions of high posterior probability that can be identified by low-fidelity models, whose inaccuracies are accounted for by their corresponding modelling error covariance (Li & Marzouk 2014; Hansen *et al.* 2014; Wagner *et al.* 2021; Rossat *et al.* 2022). We show that the quality of the results obtained by combining ASMC with surrogate solvers in the ASMC–SURR method is enhanced when updating the surrogate compared to training the surrogate from the prior only. With ASMC–SURR there is an outstanding speed-up compared to inversion based on the high-fidelity solver. By accounting for the modelling error in the likelihood function, we obtain less informative, but still reliable results that recover the main features of the reference model. The ASMC–SURR algorithm can thus be used to obtain posterior estimations for applications that do not require the highest degree of accuracy or for poorly informative acquisition configurations. The multifidelity inversion ASMC–SURR–HF provides more accurate results, while still achieving a substantial decrease of the computational effort. If the application requires an accurate estimation of the posterior PDF, it is recommended to incorporate this transition to high-fidelity. In our multifidelity setting, we have demonstrated the importance of having a high quality surrogate when performing the transition, such that less time is spent on the resulting high-fidelity stage. In our example, the resulting computational gain due to the surrogate updates exceeds a factor of three with respect to surrogate training with prior samples only.

We have assessed the computational gain by only considering the number of high-fidelity simulations as the evaluation of the high-fidelity solver represents the main source of computational demand. Nevertheless, training costs must also be taken into account, which will vary according to the desired accuracy (and thus, the truncation scheme used), the dimensionality of the input and output domains, and the size of the training set. For our examples (100-D input and 69 output parameters), the training cost for 1000 and 200 training sets was approximately one hour and 30 minutes, respectively. Each output (in this contribution the traveltime for any given source–receiver pair) requires its own PCE model, with the training cost for each PCE depending non-linearly on the sample size. Note that PCA can also be applied in the output domain, thus reducing the number of PCEs to be trained and, hence, alleviate the overall computational burden (Meles *et al.* 2022). The relevance of the PCE training cost as compared to the whole inversion depends on the length of the ASMC process and the computational burden of high-fidelity modelling. In our ASMC–SURR–HF examples, the training represented less than 10 per cent of the total time of the run.

The ASMC parameters were chosen based on manual testing. It is challenging to provide recommendations for choosing appropriate parameter values as they depend on the problem and the computational resources available. Unlike $K$, the impact of $CESS_{\mathrm{op}}$ on the total number of forward simulations is non-linear and challenging to predict in advance. For this reason, it is recommended to initially select a sufficiently large $CESS_{\mathrm{op}}$ and then adjust the value of $K$, with $CESS_{\mathrm{op}}$ exceeding $0.99N$ (Amaya *et al.* 2021). In this study we started with this minimum recommended $CESS_{\mathrm{op}}$ and increased $K$ in order to increase the quality of the resulting power posterior approximations. On the other hand, the more complex the posterior distribution, the more particles are needed. The total number $N$ of particles effectively used, however, can be limited by the number of cores available for parallel computation. In this study we use a relatively small number of $N = 50$ particles, but the posterior approximation can be improved by increasing $N$ if enough cores are available. In this algorithm we have parallelized the $N$-forward simulations corresponding to one MCMC step when using the high-fidelity solver. The simulations of the trained PCEs are vectorized, meaning that regardless of the number of particles the PCE simulation time will be almost the same. Therefore, the advantage of using PCEs within ASMC inversions is not only the computing time reduction compared to the high-fidelity solvers, but also that less resources are needed, as the high-fidelity solvers (ideally) should run in parallel with as many cores as the number of particles. A future consideration to enhance algorithm efficiency is to use a higher number of particles in the surrogate phase of the inversion, followed by a reduction by resampling in the transition to the high-fidelity phase.

Our regularly spaced surrogate updates during the ASMC inversion led to improvements compared to training based on prior samples only. The design of alternative and more elaborate surrogate retraining schemes is a topic for future research. In general terms, the better the surrogate emulates the high-fidelity response, the more accurate the inversion estimations are and the lower the price to pay (in computing time) when transitioning to the high-fidelity solver. We considered samples gathered in a cumulative way yielding a training set that grows with each update. This approach was chosen for two reasons. First, it makes use of the information given by all high-fidelity simulations created for training purposes. Secondly, maintaining the previous samples in the training reduces the difference in the estimated PCE error covariance matrices between surrogate updates, thereby, lowering the variance of the incremental weights in the IS-based surrogate transitions of the ASMC algorithm. In the future, ASMC schemes with surrogate updates could be considered for monitoring scenarios where the physical geometries remain relatively constant. In these cases, the last trained surrogate of the inversion could be used as the starting solver for the data acquired in the following time lapse.

The quality of ASMC results are sensitive to the quality of the intermediate importance-sampling steps as reflected in the variance of the importance weights. When the importance sampling steps are of poor quality, few particles contribute effectively to the estimations. The quality of the importance sampling steps, when transitioning between forward solvers, is less controlled compared to the gradual increments of the inverse temperatures. The importance sampling distribution should be ideally more disperse than the target and centred on the same values (Geweke 1989). However, when transitioning between solvers, enough overlap between the importance and target distribution is difficult to ensure. Moreover, it is challenging to accurately estimate the model errors associated with the surrogates as we do not know with certainty how a surrogate will perform in a different region of the parameter space than the one used for training. The errors in the intermediate normalizing ratios used to update the evidence accumulate throughout the run and this can explain why the evidences are underestimated. To quantify the uncertainty associated with the evidence estimations in ASMC, repeated runs or a single-run estimator can be used (Amaya *et al.* 2021).

Even if adaptive SMC schemes have shown to be advantageous and convenient, there is concern regarding bias (decreasing asymptotically with the number of particles) that arises from the adaptive selection of temperatures (Beskos *et al.* 2016; Latz *et al.* 2018). When the computational resources are available and the study requires ascertained unbiased evidence estimations, one alternative is to perform a second non-adaptive SMC run using the temperature sequence obtained in a first ASMC run (Dai *et al.* 2022). In Appendix C, we provide the metrics obtained when rerunning the eikonal tests with the previously optimized tempering schedule. For our considered cases, there appear to be no significant bias caused by the adaptive selection of temperatures (*cf.* Tables 1 and C1).

In this study, we focus on the relative improvement and computational gain of the proposed method using multiple surrogates and multifidelity, compared to standard approaches where either only the high-fidelity solver is used, or surrogates are trained only with prior samples. We tested the method on a GPR traveltime tomography problem relying on surrogates obtained through polynomial chaos expansion and with parameter dimensionality-reduction offered by PCA. Nevertheless, the proposed method is not limited to these choices and it can be extended to other types of problems (e.g. interferometric seismic noise tomography (Nicolson *et al.* 2012) and alternative surrogate models [e.g. Gaussian process regression

(Santner *et al.* 2003; Rasmussen 2004) or deep learning (Jin *et al.* 2020)].

## 5 CONCLUSION

Using a GPR crosshole tomography example, we demonstrate that surrogate modelling can accelerate the ASMC inversion by orders of magnitude while still providing results that are similar to those obtained when using the high-fidelity solver only. We demonstrate that a multifidelity approach combining surrogate modelling with a transition to a final high-fidelity stage improves the accuracy of the posterior estimates further while still achieving a substantial reduction of computational cost. The choice of including the high-fidelity stage or not, for a given problem, will in practice depend on the computing budget and the demands in terms of estimation accuracy. When using surrogates only and considering the same high-fidelity budget for training, we find that retraining the surrogate as the ASMC inversion progresses provides better estimates than training from the prior only. When incorporating the transition to the high-fidelity stage, the advantage of retraining the surrogate leads to a substantial reduction of the computational cost. The evidences estimated with surrogate updating were found to be more accurate than the ones obtained when training only with prior samples. We emphasize that the method is not limited to PCE-based surrogate models, PCA parametrizations or GPR examples, but that it can be adapted to other surrogates, model parametrizations and geophysical problem settings.

## ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (project number: 184574). We acknowledge the detailed and insightful reviews provided by Jan Dettmer and Malcolm Sambridge.

## DATA AVAILABILITY

Our ASMC code and test examples are available in the following repository: https://doi.org/10.5281/zenodo.10075203.

## REFERENCES

Amaya, M., Linde, N. & Laloy, E., 2021. Adaptive sequential Monte Carlo for posterior inference and model selection among complex geological priors, *Geophys. J. Int.,* **226**(2), 1220–1238.

Amaya, M., Linde, N. & Laloy, E., 2022. Hydrogeological multiple-point statistics inversion by adaptive sequential Monte Carlo, *Adv. Water Resour.,* **166,** doi:10.1016/j.advwatres.2022.104252.

Beskos, A., Jasra, A., Kantas, N. & Thiery, A., 2016. On the convergence of adaptive sequential Monte Carlo methods, *Ann. Appl. Probab.,* **26**(2), 1111–1146.

Blatman, G., 2009. Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis, *PhD thesis*, Université Blaise-Pascal, Clermont-Ferrand II.

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

Brunetti, C., Linde, N. & Vrugt, J.A., 2017. Bayesian model selection in hydrogeophysics: application to conceptual subsurface models of the South Oyster Bacterial Transport Site, Virginia, USA, *Adv. Water Resour.,* **102,** 127–141.

Catanach, T.A., Vo, H.D. & Munsky, B., 2020. Bayesian inference of stochastic reaction networks using multifidelity sequential tempered Markov chain Monte Carlo, *Int. J. Uncertain. Quant.,* **10**(6), 515–542.

Chopin, N., 2002. A sequential particle filter method for static models, *Biometrika,* **89**(3), 539–552.

Christen, J.A. & Fox, C., 2005. Markov chain Monte Carlo using an approximation, *J. Comput. Graph. Stat.,* **14**(4), 795–810.

Dai, C., Heng, J., Jacob, P.E. & Whiteley, N., 2022. An invitation to sequential Monte Carlo samplers, *J. Am. Stat. Assoc.,* **117**(539), 1587–1600.

Davies, L., Ley-Cooper, A., Sutton, M. & Drovandi, C., 2023. Bayesian detectability of induced polarisation in airborne electromagnetic data, *Geophys. J. Int.,* **235**(3), 2499–2523.

Del Moral, P., Doucet, A. & Jasra, A., 2006. Sequential Monte Carlo samplers, *J. R. Stat. Soc., B,* **68**(3), 411–436.

Dettmer, J., Dosso, S.E. & Osler, J.C., 2010. Bayesian evidence computation for model selection in non-linear geoacoustic inference problems, *J. acoust. Soc. Am.,* **128**(6), 3406–3415.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2011. Sequential trans-dimensional Monte Carlo for range-dependent geoacoustic inversion, *J. acoust. Soc. Am.,* **129**(4), 1794–1806.

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.

Douc, R. & Cappé, O., 2005. Comparison of resampling schemes for particle filtering, in *ISPA 2005: Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis,* pp. 64–69, IEEE.

Doucet, A. & Johansen, A.M., 2011. A tutorial on particle filtering and smoothing: fifteen years later, *The Oxford Handbook of Nonlinear Filtering,* eds Crisan, D. & Rozovskii, B., pp. 656–704, Oxford Univ. Press.

Doucet, A. & Lee, A., 2018. Sequential Monte Carlo methods, in *Handbook of Graphical Models,* pp. 165–189.

Earl, D.J. & Deem, M.W., 2005. Parallel tempering: theory, applications, and new perspectives, *Phys. Chem. Chem. Phys.,* **7**(23), 3910–3916.

Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration, *Econometrica,* **57**(6), 1317–1339.

Good, I.J., 1992. Rational decisions, in *Springer Series in Statistics,* eds Kotz, S. & Johnson, N.L., pp. 365–377, Springer.

Hammersley, J.M. & Handscomb, D.C., 1964. *General Principles of the Monte Carlo Method,* pp. 50–75, Springer Netherlands.

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.

Hunziker, J., Laloy, E. & Linde, N., 2019. Bayesian full-waveform tomography with application to crosshole ground penetrating radar data, *Geophys. J. 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.

Jiang, A. & Jafarpour, B., 2021. Inverting subsurface flow data for geologic scenarios selection with convolutional neural networks, *Adv. Water Resour.,* **149,** doi:10.1016/j.advwatres.2020.103840.

Jin, Z.L., Liu, Y. & Durlofsky, L.J., 2020. Deep-learning-based surrogate model for reservoir simulation with time-varying well controls, *J. Petrol. Sci. Eng.,* **192,** doi:10.1016/j.petrol.2020.107273.

Jolliffe, I.T. & Cadima, J., 2016. Principal component analysis: a review and recent developments, *Phil. Trans. R. Soc., A,* **374**(2065), doi:10.1098/rsta.2015.0202.

Kass, R.E. & Raftery, A.E., 1995. Bayes factors, *J. Am. Stat. Assoc.,* **90**(430), 773–795.

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P., 1983. Optimization by simulated annealing, *Science,* **220**(4598), 671–680.

Kitanidis, P.K. & Lee, J., 2014. Principal component geostatistical approach for large-dimensional inverse problems, *Water Resour. Res.,* **50**(7), 5428–5443.

Koutsourelakis, P.-S., 2009. A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters, *J. Comput. Phys.,* **228**(17), 6184–6211.

Krüger, F., Lerch, S., Thorarinsdottir, T. & Gneiting, T., 2021. Predictive inference based on Markov chain Monte Carlo output, *Int. Stat. Rev.,* **89**(2), 274–301.

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 Resour. Res.,* **51**(6), 4224–4243.

Latz, J., Papaioannou, I. & Ullmann, E., 2018. Multilevel sequential$^2$ Monte Carlo for Bayesian inverse problems, *J. Comput. Phys.,* **368,** 154–178.

Li, J. & Marzouk, Y.M., 2014. Adaptive construction of surrogates for the Bayesian solution of inverse problems, *SIAM J. Sci. Comput.,* **36**(3), A1163–A1186.

Linde, N., Renard, P., Mukerji, T. & Caers, J., 2015. Geological realism in hydrogeological and geophysical inverse modeling: a review, *Adv. Water Resour.,* **86,** 86–101.

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

Lüthen, N., Marelli, S. & Sudret, B., 2021. Sparse polynomial chaos expansions: literature survey and benchmark, *SIAM/ASA J. Uncertain. 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. Uncertain. Quant.,* **12**(3), doi:10.1615/Int.J.UncertaintyQuantification.2021036153.

Marelli, S. & Sudret, B., 2014. UQLab: A framework for uncertainty quantification in Matlab, in *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management,* pp. 2554–2563, eds Beer, M., Au, S. & Hall, J.W., American Society of Civil Engineers.

Marelli, S., Lüthen, N. & Sudret, B., 2022. UQLab user manual – Polynomial chaos expansions, Tech. rep., Report UQLab-V2.0-104, Chair of Risk, Safety and Uncertainty Quantification, Switzerland.

Marzouk, Y. & Xiu, D., 2009. A stochastic collocation approach to Bayesian inference in inverse problems, *Commun. Comput. Phys.,* **6,** 826–847.

Marzouk, Y.M., Najm, H.N. & Rahn, L.A., 2007. Stochastic spectral methods for efficient bayesian solution of inverse problems, *J. Comput. Phys.,* **224**(2), 560–586.

Meles, G.A., Linde, N. & Marelli, S., 2022. Bayesian tomography with prior-knowledge-based parametrization and surrogate modelling, *Geophys. J. Int.,* **231**(1), 673–691.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.,* **21**(6), 1087–1092.

Minson, S., Simons, M. & Beck, J., 2013. Bayesian inversion for finite fault earthquake source models. I—theory and algorithm, *Geophys. J. Int.,* **194**(3), 1701–1726.

Neal, R.M., 2001. Annealed importance sampling, *Stat. Comput.,* **11**(2), 125–139.

Ng, L. W.-T. & Eldred, M., 2012. Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation, in *Proceedings of the 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference,* AIAA 2012-1852.

Nicolson, H., Curtis, A., Baptie, B. & Galetti, E., 2012. Seismic interferometry and ambient noise tomography in the british isles, *Proc. Geol. Assoc.,* **123**(1), 74–86.

Park, C., Haftka, R.T. & Kim, N.H., 2017. Remarks on multi-fidelity surrogates, *Struct. Multidiscipl. Optim.,* **55,** 1029–1050.

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

Peterson, J.E., Jr, 2001. Pre-inversion corrections and analysis of radar tomographic data, *J. Environ. Eng. Geophys.,* **6**(1), 1–18.

Podvin, P. & Lecomte, I., 1991. Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools, *Geophys. J. Int.,* **105**(1), 271–284.

Rasmussen, C.E., 2004. Gaussian processes in machine learning, in *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, Tübingen, Germany, Revised Lectures,* pp. 63–71, Springer.

Reynolds, A.C., He, N., Chu, L. & Oliver, D.S., 1996. Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data, *SPE J.,* **1**(04), 413–426.

Rossat, D., Baroth, J., Briffaut, M. & Dufour, F., 2022. Bayesian inversion using adaptive polynomial chaos kriging within subset simulation, *J. Comput. Phys.,* **455,** doi:10.1016/j.jcp.2022.110986.

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40**(3), 3–1-3-29.

Santner, T.J., Williams, B.J., Notz, W.I. & Williams, B.J., 2003. *The Design and Analysis of Computer Experiments,* Vol. **1,** Springer.

Sarma, P., Durlofsky, L.J., Aziz, K. & Chen, W.H., 2006. Efficient real-time reservoir management using adjoint-based optimal control and model updating, *Comput. Geosci.,* **10,** 3–36.

Schöniger, A., Wöhling, T., Samaniego, L. & Nowak, W., 2014. Model selection on solid ground: rigorous comparison of nine ways to evaluate Bayesian model evidence, *Water Resour. Res.,* **50**(12), 9484–9513.

Scott, D.W., 2015. *Multivariate Density Estimation: Theory, Practice, and Visualization,* John Wiley & Sons.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* SIAM.

Torre, E., Marelli, S., Embrechts, P. & Sudret, B., 2019. Data-driven polynomial chaos expansion for machine learning regression, *J. Comput. Phys.,* **388,** 601–623.

Vasyura-Bathke, H. *et al.*, 2020. The Bayesian earthquake analysis tool, *Seismol. Res. Lett.,* **91,** 1003–1018.

Wagner, P.-R., Fahrni, R., Klippel, M., Frangi, A. & Sudret, B., 2020. Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels, *Eng. Struct.,* **205,** doi:10.1016/j.engstruct.2019.110063.

Wagner, P.-R., Marelli, S. & Sudret, B., 2021. Bayesian model inversion using stochastic spectral embedding, *J. Comput. Phys.,* **436,** doi:10.1016/j.jcp.2021.110141.

Wang, Z., Bovik, A.C., Sheikh, H.R. & Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.,* **13**(4), 600–612.

Xiu, D. & Karniadakis, G.E., 2002. The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.,* **24**(2), 619–644.

Yardim, C., Gerstoft, P. & Hodgkiss, W.S., 2010. Geoacoustic and source tracking using particle filtering: experimental results, *J. acoust. Soc. Am.,* **128**(1), 75–87.

Zhou, Y., Johansen, A.M. & Aston, J.A., 2016. Toward automatic model comparison: an adaptive sequential Monte Carlo approach, *J. Comput. Graph. Stat.,* **25**(3), 701–726.

# APPENDIX A: SURROGATE MODELLING WITH POLYNOMIAL CHAOS EXPANSION

Surrogate models $\hat{\mathcal{M}}(\boldsymbol{\theta})$ are typically analytical functions that emulate selected quantities of interest (QoI) of expensive forward solvers with much lower computational demand, based on a relatively small size set of training data (the *experimental design*):

$$\hat{\mathcal{M}}(\boldsymbol{\theta}) \approx \mathcal{M}(\boldsymbol{\theta}). \tag{A1}$$

Polynomial chaos expansion (PCE) is a stochastic spectral expansion method that projects a forward operator $\mathcal{M}(\boldsymbol{\theta})$ onto a suitable orthonormal polynomial basis. PCE surrogate models have shown to be both flexible and efficient (Lüthen *et al.* 2021; Lüthen *et al.* 2022). If $\boldsymbol{\theta}$ is a random vector of independent parameters with joint probability density function $f_{\boldsymbol{\theta}}$, and $H$ a stochastic Hilbert space, then any map $\mathcal{M}(\boldsymbol{\theta})$ of finite variance on $H$ can be expanded as a sum of polynomial basis elements $\psi_b(\boldsymbol{\theta})$ (Xiu & Karniadakis 2002):

$$\mathcal{M}(\boldsymbol{\theta}) = \sum_{b \in \mathcal{N}^{\mathcal{M}}} a_b \psi_b(\boldsymbol{\theta}), \tag{A2}$$

where $a_b$ are the coefficients and $\psi_b(\boldsymbol{\theta})$ the multivariate polynomials orthonormal with respect to $f_{\boldsymbol{\theta}}$. In practice, the surrogate model $\hat{\mathcal{M}}(\boldsymbol{\theta})$ is obtained by truncating the series at a maximum allowed polynomial degree and the coefficients $a_b$ are calculated using basis-adaptive sparse regression techniques (Lüthen *et al.* 2022). In this study, the surrogate solvers are trained to learn a direct mapping between the PCA-reduced set of parameters and the output response (Meles *et al.* 2022). The forward problem when using a PCE approximation on a truncated PCA set of coordinates can be formulated as:

$$\mathbf{y} = \hat{\mathcal{M}}(\boldsymbol{\theta}) + \epsilon_y + \epsilon_{\mathrm{PCA}} + \epsilon_{\mathrm{PCE}}, \tag{A3}$$

where $\boldsymbol{\epsilon}_{\mathrm{PCE}}$ is the error associated with the surrogate modelling.

To train a PCE surrogate, a set of model realizations together with their high-fidelity forward responses are required. These input–output pairs compose the PCE training set. We consider herein the input in the reduced PCA domain, and the output to be the simulated response of the PCA models projected onto the physical domain $\mathcal{F}(g(\boldsymbol{\theta}))$. The forward solver used to compute the responses is a high-fidelity solver with modelling error assumed to be zero. In this paper, we rely on PCE surrogates trained using the Matlab-based package UQLab (Marelli & Sudret 2014); details on the implementation of PCE can be found in Marelli *et al.* (2022). The marginal distributions of the input parameters are approximated by Kernel density estimation (Torre *et al.* 2019).

A Bayesian inversion that incorporates dimensionality-reduction and surrogate modelling needs to consider the errors induced by these simplifications. If we assume the various error sources to be normally distributed, the corresponding likelihood function (*cf.*, eq. 1) can be expressed as (Tarantola 2005):

$$p(\mathbf{y}|\boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\hat{\mathcal{M}}(\boldsymbol{\theta}) - \mathbf{d_d} - \mathbf{y})^T \mathbf{C}^{-1} (\hat{\mathcal{M}}(\boldsymbol{\theta}) - \mathbf{d_d} - \mathbf{y})\right), \tag{A4}$$

where $n$ is the number of data points, $\mathbf{y}$ the observed data, $\mathbf{C}$ the covariance matrix and $|\mathbf{C}|$ its determinant. The $\mathbf{d_d}$ variable accounts for the modelling bias (Hansen *et al.* 2014). The covariance matrix can be designed to account not only for the data errors, but for the errors associated with the dimensionality reduction and surrogate modelling as well:

$$\mathbf{C} = \mathbf{C}_y + \mathbf{C}_{\mathrm{PCA}} + \mathbf{C}_{\mathrm{PCE}}, \tag{A5}$$

where $\mathbf{C}_y$ is the data covariance matrix, $\mathbf{C}_{\mathrm{PCA}}$ the covariance of the PCA projection error and $\mathbf{C}_{\mathrm{PCE}}$ the covariance of the surrogate modelling error. Both $\mathbf{C}_{\mathrm{PCA}}$ and $\mathbf{C}_{\mathrm{PCE}}$ can be computed as:

$$\mathbf{C}_{\mathrm{PCA/PCE}} = \frac{1}{N}\mathbf{D}_m\mathbf{D}_m^T, \tag{A6}$$

where $\mathbf{D}_m$ is a misfit matrix. In the case of $\mathbf{C}_{\mathrm{PCA}}$, $\mathbf{D}_m$ contains the difference between the high-fidelity solver output for the full models and the models resulting from the PCA projection computed on a set of samples. For $\mathbf{C}_{\mathrm{PCE}}$, it is convenient in practice to derive $\mathbf{D}_m$ from the PCE training set using cross-validation, to avoid having to compute additional expensive high-fidelity responses. The $j$th $\mathbf{D}_m$ column can be estimated with the output of a surrogate that is trained on all the input parameters excepted the $j$th (Blatman 2009), we refer to this $\mathbf{C}_{\mathrm{PCE}}$ as the leave-one-out ($L_{\mathrm{OO}}$) covariance.

## APPENDIX B: FULL ASMC–SURR–HF ALGORITHM

Algorithm 1 describes our proposed method for estimating the posterior PDF and the evidence. To update the PCE surrogate, the high-fidelity solver responses for the $N$ particles states are computed after every $T_{\mathrm{fp}}$ inverse temperatures, and the models together with the responses are saved. The PCE surrogate is retrained at regular intervals of $T_{\mathrm{up}}$ inverse temperatures. During an interval between surrogate updates, a total amount of $\tau_0 = N\lfloor T_{\mathrm{up}}/T_{\mathrm{fp}}\rfloor$ are collected. The initial surrogate $s_0$ is trained using $\tau_0$ prior samples and their high-fidelity solver responses. The training set for the surrogate sequence is cumulative, that is, to train the $s_j$ surrogate, the training set size will be $\tau_j = (j+1)\tau$ (includes $\tau_0$). In this way, the algorithm takes advantage of all the available high-fidelity evaluations, while the new samples enlarge the training set by focusing progressively on parameter regions of significant posterior probability mass. The leave-one-out modelling error covariance matrix $\mathbf{C}_{\mathrm{PCE}}$ is obtained from the calculated coefficients and the training set, and is updated in the inversion together with the surrogate. After the transition to the high-fidelity solver, $\mathbf{C}_{\mathrm{PCE}}$ is set to zero, as no modelling errors are assumed.

## APPENDIX C: RERUNS WITH FIXED INVERSE TEMPERATURE SEQUENCE

As the adaptivity of the temperature sequence can not ensure unbiadness on the evidence estimations, we provide a set of reruns (Table C1) in which we keep fixed the adaptively-determined sequence of $\alpha_s$ of the eikonal runs of Table 1. Comparing the metrics in the two tables, there are no significant differences in the performance metrics suggesting that bias due to the adaptivity is small compared to the inherent variability in the results due to the stochastic nature of the algorithm.

**Algorithm 1: ASMC–SURR**

The part concerning the ASMC section is adapted from Zhou *et al.* (2016) (their algorithm 4).

```
Variables to pre-define:
  Number of particles ($N$), optimal CESS ($CESS_{\mathrm{op}}$), ESS threshold ($ESS^*$),
  number of MCMC steps for each intermediate distribution ($K$), minimal acceptance rate ($AR_{\min}$),
  percentage of change ($f$) of the proposal scale $\phi$,
  period for evaluating and saving high-fidelity forward simulations for surrogate update ($T_{\mathrm{fp}}$),
  period of surrogate retraining ($T_{\mathrm{up}}$).
Initialization: Set $t = 0$
  Train initial surrogate $s_j = s_0$
  Set $\alpha = 0$
  Sample $\boldsymbol{\theta}_0$ from the prior $\pi(\boldsymbol{\theta})$ $N$ times
  Set the $N$-dimensional vector of normalized weights $\mathbf{W}_0 = [\frac{1}{N}; \frac{1}{N}; ...; \frac{1}{N}]$
  Set evidence $\pi(\mathbf{y}) = 1$
  Set $t_{\mathrm{fp}} = 0$
  Set $t_{\mathrm{up}} = 0$
Iteration : Set $t = t + 1$
    $t_{\mathrm{fp}} = t_{\mathrm{fp}} + 1$,
    $t_{\mathrm{up}} = t_{\mathrm{up}} + 1$.
  Do K MCMC steps for each of the N particles (chains):
    Propose moves $\boldsymbol{\theta}_p$ and accept or reject based on acceptance criterion (eq. 12).
  Tune proposal scale
    If acceptance rate $AR < AR_{\min}$ then decrease proposal scale factor: $\phi = \phi \times (1 - \frac{f}{100})$
    If $\phi < \phi_{\min}$ then $\phi = \phi_{\min}$.
  Search for next power posterior
    If $t_{\mathrm{up}} < T_{\mathrm{up}}$ do binary search for the increment $\Delta\alpha$ that gives the CESS (eq. 11) that is the closest to $CESS_{\mathrm{op}}$
    Update $\alpha = \min(1, \alpha + \Delta\alpha)$ and define the following power posterior.
    Perform the IS step: compute the weight increments $w_t$ (eq. 6),
    update and save the normalized weights $W_t^i$ (eq. 7) and the evidence $\pi(\mathbf{y})$ (eq. 10)
  Evaluate high-fidelity - Update surrogate
    If $t_{\mathrm{fp}} = T_{\mathrm{fp}}$ compute the $N$ current high-fidelity forward responses and save them for training, set $t_{\mathrm{fp}} = 0$.
    If $t_{\mathrm{up}} = T_{\mathrm{up}}$, retrain surrogate and replace, update covariance matrix (eqs A5 and A6), set $t_{\mathrm{up}} = 0$.
      Search for the $\alpha$ that yields the highest CESS.
      Perform IS step to transition to the updated surrogate solver, compute the weight increments $w[s_j, s_{j+1}]_t^i$ (eq. 13),
      update and save the normalized weights $W_t^i$ (eq. 7) and the evidence $\pi(\mathbf{y})$ (equation10).
  Resampling
    Calculate ESS (eq. 8), if $ESS < ESS^*$ do resampling: re-organize the particle states and update $\mathbf{W}_t = [\frac{1}{N}; \frac{1}{N}; ...; \frac{1}{N}]$
Repeat until $\alpha = 1$
```

**Algorithm 1 extension: ASMC–SURR–HF**

```
  Transition to the high-fidelity solver
    Search for the $\alpha$ that yields the highest CESS.
    Perform IS step to transition to the high-fidelity forward solver, compute weight increments $w[s_j, HF]_t^i$ (eq. 16 ).
    Update and save the normalized weights $W_t^i$ (eq. 7) and the evidence $\pi(\mathbf{y})$ (eq. 10).
    Update covariance matrix.
  Continue-Iteration-loop using the high-fidelity solver without Evaluate high-fidelity - Update surrogate step, until $\alpha = 1$.
```

**Table C1.** Resulting length of the ASMC runs and performance metrics of the reruns using the eikonal solver as high-fidelity. In these runs the sequence of temperatures is not calculated adaptively, but fixed from the first round of runs (Table 1).

| | ASMC$_{\text{eikonal}}$ | ASMC$_{\text{eikonal}}$-SURR$_{\text{prior}}$ | ASMC$_{\text{eikonal}}$-SURR$_{\text{prior}}$–HF | ASMC$_{\text{eikonal}}$-SURR$_{\text{updates}}$ | ASMC$_{\text{eikonal}}$-SURR$_{\text{updates}}$–HF |
|---|---|---|---|---|---|
| *L* | 192 | 173 | 251 | 180 | 203 |
| *HF sim*[$\times 10^3$] | 4800 | 1 | 1950 | 1 | 576 |
| *HF sim relative* | 100 per cent | 0.02 per cent | 40.63 per cent | 0.02 per cent | 12 per cent |
| *PCE sim*[$\times 10^3$] | 0 | 4325 | 4325 | 4500 | 4500 |
| *Resampling* | 1 | 1 | 3 | 3 | 4 |
| *SSIM* | 1 | 0.84 | 0.93 | 0.89 | 0.91 |
| *Mean* $\sigma$ | 0.83 | 0.83 | 0.80 | 0.84 | 0.80 |
| *Mean log-score* | 1.43 | 1.76 | 1.52 | 1.58 | 1.45 |
| *RMSE output* | 0.50 ns | 0.84 ns | 0.49 ns | 0.65 ns | 0.49 ns |
| *Log-evidence* | −111.2 | −121.6 | −121.6 | −113.4 | −114.5 |