Year: 2023

Geophysical and Hydrogeological Bayesian Inversion by Adaptive Sequential Monte Carlo

Amaya Macarena

Amaya Macarena, 2023, Geophysical and Hydrogeological Bayesian Inversion by Adaptive Sequential Monte Carlo

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive.
http://serval.unil.ch

UNIL | Université de Lausanne

Faculté des géosciences et de l'environnement
Institut des sciences de la Terre

Geophysical and Hydrogeological Bayesian Inversion by Adaptive Sequential Monte Carlo

Thèse de doctorat

Présentée à la

Faculté des géosciences et de l'environnement

pour l'obtention du grade de

Docteure en sciences de la Terre

par

Macarena Amaya

Diplôme en Géophysique
Facultad de Ciencias Astronómicas y Geofísicas
Universidad Nacional de La Plata, Argentina

Jury

Prof. Dr. Marie-Elodie Perga, Présidente
Prof. Dr. Niklas Linde, Directeur de these
Prof. Dr. James Irving, Expert interne
Dr. Anneli Guthke, Experte externe
Prof. Dr. Philippe Renard, Expert externe

Lausanne, 2023

UNIL | Université de Lausanne
Faculté des géosciences et de l'environnement
bâtiment Géopolis bureau 4631

IMPRIMATUR

Vu le rapport présenté par le jury d'examen, composé de

Présidente de la séance publique :	Mme la Professeure Marie-Elodie Perga
Présidente du colloque :	Mme la Professeure Marie-Elodie Perga
Directeur de thèse :	M. le Professeur Niklas Linde
Expert interne :	M. le Professeur James Irving
Experte externe :	Mme la Docteure Anneli Guthke
Expert externe :	M. le Professeur Philippe Renard

Le Doyen de la Faculté des géosciences et de l'environnement autorise l'impression de la thèse de

Madame Macarena AMAYA

Titulaire d'un
Master of Geophysics
de l'Université de la Plata, Argentine

intitulée

GEOPHYSICAL AND HYDROGEOLOGICAL BAYESIAN INVERSION BY ADAPTIVE SEQUENTIAL MONTE CARLO

Lausanne, le 23 octobre 2023

Pour le Doyen de la Faculté des géosciences et de l'environnement

Professeure Marie-Elodie Perga

Contents

Abbreviations

1-D, 2-D, 3-D ... One, two, and three Dimensional

AIC ..... Akaike information criteria

AIS ..... Annealed importance sampling

ASMC ..... Adaptive sequential Monte Carlo

BFMC ..... Brute force Monte Carlo

BIC ..... Bayesian information criteria

CESS ..... Conditional effective sample size

CM ..... Conceptual model

CRIM ..... Complex refractive index method

DREAM$_{(ZS)}$ ..... Differential evolution adaptive Metropolis ZS

ESS ..... Effective sample size

FDTD ..... Finite-difference time-domain

GAN ..... Generative adversarial network

GPR ..... Ground penetrating radar

HF ..... High-fidelity

IS ..... Importance sampling

KDE ..... Kernel density estimation

KIC ..... Kashyap information criteria

LP ..... Laplace-Metropolis

MAP ..... Maximum a posteriori

MCMC ..... Markov chain Monte Carlo

MPS ..... Multiple-point statistics

PCA ..... Principal component analysis

PCE ..... Polinomial chaos expansion

PDF ..... Probability density function

PT ..... Parallel tempering

RMSE ..... Root mean square error

SA ..... Simulated annealing

SGAN ..... Spatial generative adversarial network

SGR ..... Sequential geostatistical resampling

SMC ..... Sequential Monte Carlo

SSIM ..... Structural similarity index measure

ST ..... Simulated tempering

TI ..... Training image

List of Figures

List of Tables

Abstract

Geophysics and hydrogeology are scientific disciplines that play important roles in understanding the subsurface environment. By proposing and implementing new methods to explore and characterize the subsurface, it is possible to make predictions, manage resources and mitigate hazards in better ways. The data acquired in geophysical and hydrogeological surveys are used to infer relevant properties of the subsurface by solving inverse problems. A probabilistic framework to solve these problems that accounts for different uncertainty sources is provided by Bayesian inversion. Unfortunately, the estimation of the posterior probability density function (PDF) provided by widely used standard Markov chain Monte Carlo (MCMC) methods can be limited when inverse problems are non-linear and high-dimensional. In practice, they may fail to properly sample the posterior PDF in allowed computational times by getting trapped in local minimas. Alternative methods relying on tempering overcome these difficulties by enhancing the freedom of exploration using a temperature variable reducing the influence of the likelihood function in Bayes' theorem. Sequential Monte Carlo (SMC) is a group of methods that build a sequence of importance sampling steps between distributions with gradually decreasing temperatures, going from the prior to the posterior PDF. These particle approaches use a set of evolving particles and their associated weights to approximate the posterior PDF. An appealing property of SMC methods is that they also offer an estimation of the evidence (marginal likelihood), a key quantity for comparing and ranking competitive conceptual models. The challenge of SMC and other tempering methods of pre-defining a suitable sequence of temperatures is overcome by adaptive SMC (ASMC) that adaptively finds the most suitable temperature decrement. In this thesis, we implement the ASMC method in challenging synthetic geophysical and hydrogeological test cases where the prior information is encoded either by deep generative neural networks or training images explored by multiple-point statistics methods to handle geologically-realistic prior PDFs. We find that ASMC is considerably more efficient to locate and sample the posterior PDF than a state-of-the-art adaptive MCMC method and parallel tempering, a popular MCMC method relying on tempering. We demonstrate that the ASMC evidence estimations are robust regarding proposal schemes and inversion parameters. Still, ASMC is computationally costly when considering expensive forward solvers in geophysics and hydrology. To seek important computational gains while targeting the original posterior PDF and evidence, we propose a multifidelity ASMC approach relying on surrogate solvers that are updated as the inversion progresses. For a multivariate Gaussian test example, we show that this method accelerates the inversion while reaching similar accuracy as methods relying on high-fidelity solvers only. We find that ASMC is very powerful to solve complex geophysical and hydrogeological inverse problems, especially when non-linearity and multi-modality challenge standard MCMC approaches.

Key words : Bayesian inversion, Bayesian model selection, sequential Monte Carlo, geostatistics, geophysics, hydrogeology, multifidelity methods.

Résumé

La géophysique et l'hydrogéologie sont des disciplines scientifiques qui jouent un rôle important dans la compréhension de l'environnement souterrain. En proposant et en mettant en œuvre de nouvelles méthodes pour explorer et caractériser le sous-sol, il est possible de faire des prévisions, de gérer les ressources et d'atténuer les risques de manière plus efficace. Les données acquises lors des études géophysiques et hydrogéologiques sont utilisées pour déduire les propriétés pertinentes du sous-sol en résolvant des problèmes inverses. L'inversion bayésienne offre un cadre probabiliste pour résoudre ces problèmes en tenant compte des différentes sources d'incertitude. Malheureusement, l'estimation de la fonction de densité de probabilité a posteriori fournie par les méthodes standard de Markov chain Monte Carlo (MCMC) largement utilisées peut être restreinte lorsque les problèmes inverses sont non linéaires et de haute dimension. Dans la pratique, elles risquent de ne pas parvenir à échantillonner correctement la fonction de densité de probabilité a posteriori dans les temps de calcul autorisés en se retrouvant piégées dans des minima locaux. D'autres méthodes reposant sur le tempering permettent de surmonter ces difficultés en augmentant la liberté d'exploration à l'aide d'une variable de température réduisant l'influence de la fonction de vraisemblance dans le théorème de Bayes. La méthode sequential Monte Carlo (SMC) est un groupe de méthodes qui construisent une séquence d'étapes d'échantillonnage préférentiel entre des distributions dont les températures diminuent progressivement, en passant de la fonction de densité de probabilité a priori à la posteriori. Ces approches particulaires utilisent un ensemble de particules évolutives et leurs poids associés pour approximer la fonction de densité de probabilité a posteriori. Une propriété intéressante des méthodes SMC est qu'elles offrent également une estimation de l'évidence (vraisemblance marginale), une quantité clé pour comparer et classer des modèles conceptuels compétitifs. Le défi du SMC et d'autres méthodes de tempering consistant à prédéfinir une séquence appropriée de températures est surmonté par le SMC adaptatif (ASMC) qui trouve de manière adaptative la décrémentation de température la plus appropriée. Dans cette thèse, nous mettons en œuvre la méthode ASMC dans des cas d'essai géophysiques et hydrogéologiques synthétiques difficiles où les informations préalables sont encodées soit par des réseaux génératifs de neurones profond, soit par des images d'entraînement explorées par des méthodes statistiques à points multiples pour traiter des priors géologiquement réalistes. Nous constatons que la méthode ASMC est considérablement plus efficace pour localiser et échantillonner la fonction de densité de probabilité a posteriori qu'une méthode MCMC adaptative et que le parallel tempering, une méthode MCMC populaire reposant sur le tempering. Nous démontrons que les estimations de l'évidence par ASMC sont robustes en ce qui concerne les schémas de proposition et les paramètres d'inversion. Cependant, la méthode ASMC est coûteuse en termes de calcul si l'on considère les solveurs utilisés en géophysique et en hydrologie. Pour obtenir des gains de calcul importants tout en ciblant la fonction de densité de probabilité a posteriori et l'évidence originale, nous proposons une approche ASMC multifidélité reposant sur des surrogate solveurs qui sont mis à jour au fur et à mesure que l'inversion progresse. Pour un exemple de test gaussien multivarié, nous montrons que cette méthode accélère l'inversion tout en atteignant une exactitude similaire à celle des méthodes reposant uniquement sur des solveurs de haute fidélité. Nous constatons que la méthode ASMC est très puissante pour résoudre des problèmes inverses géophysiques et hydrogéologiques complexes, en particulier lorsque la non-linéarité et la multi-modalité défient les approches MCMC standard.

Mots clefs : Inversion bayésienne, sélection de modèles bayésiens, Monte Carlo séquentiel, géostatistique, géophysique, hydrogéologie, méthodes multifidélité.

Résumé pour le large public

Pour étudier notre planète et son environnement, nous élaborons des idéalisations et des simplifications des systèmes naturels sous la forme de modèles conceptuels. La géophysique et l'hydrogéologie sont des disciplines scientifiques des géosciences qui étudient les propriétés et la dynamique du sous-sol. Dans ces domaines, les scientifiques ont mis au point des techniques de mesure sur le terrain qui fournissent des informations indirectes sur le sous-sol. Pour transformer les données acquises en quantités utiles caractérisant le sous-sol, nous devons effectuer des processus mathématiques et informatiques connus sous le nom d'inversions. L'approche bayésienne offre un cadre pour y parvenir ; au lieu de rechercher une solution optimale, elle décrit la probabilité de chaque réalisation possible du modèle. Dans cette thèse, nous explorons une méthode bayésienne particulière qui n'a pas encore été popularisée dans les domaines de la géophysique et de l'hydrogéologie. Cette méthode, appelée adaptive sequential Monte Carlo (ASMC), incorpore des stratégies qui améliorent les chemins empruntés pour explorer les solutions possibles du problème, ce qui la rend potentiellement plus efficace et plus puissante que les méthodes standard. Elle présente également des avantages liés à sa nature adaptative qui permettent de réduire le temps de test et l'effort de l'utilisateur. Nous évaluons les performances de l'ASMC à l'aide de différents cas tests géophysiques et hydrogéologiques complexes. Nous comparons les résultats avec des approches de pointe largement utilisées en géophysique et en hydrogéologie, et nous constatons que l'ASMC surpasse souvent ces méthodes lorsque les problèmes sont difficiles. Nous proposons en outre des stratégies pour prendre en compte des modèles conceptuels plus réalistes sur le plan géologique, pour différencier les modèles conceptuels alternatifs, ainsi que pour accélérer l'inversion qui peut être très coûteuse en termes de calcul. Nous encourageons l'utilisation de l'ASMC en géophysique et en hydrogéologie, en particulier pour les problèmes non linéaires impliquant plusieurs modèles conceptuels géologiquement réalistes.

Chapter 1

Introduction

1.1 The inverse problem in geophysics and hydrogeology

Geophysics and hydrogeology are interconnected scientific disciplines within the broader field of geoscience. While geophysics focuses on understanding the Earth's structure, composition, and dynamics (*Lowrie and Fichtner*, 2020), hydrogeology specifically deals with the distribution, movement, and reactions of water and its solutes in the subsurface (*Domenico and Schwartz*, 1997). The study of these disciplines is essential for managing resources, mitigating natural hazards, protecting the environment, and ensuring sustainable development (*Rubin and Hubbard*, 2006; *Reynolds*, 2011). They provide valuable insights into the Earth's processes and help inform decision-making for various sectors, including environmental conservation, water management, and infrastructure development.

Inverse problems can be thought of as puzzles in which we are given observational data about a system we are interested in studying, and we need to reconstruct some of its properties. They are widely studied in diverse fields such as medical imaging (*Arridge*, 1999), economics (*Horowitz*, 2014), weather forecasting (*Cotter et al.*, 2009), astronomy (*Craig and Brown*, 1986), and signal processing (*Pike et al.*, 1984), among others. They play an essential role in geophysics and hydrogeology as they enable us to derive meaningful information about subsurface structures and processes that are otherwise difficult (or impossible) to directly observe. The inversion process consists of inferring the unknown parameters of the subsurface based on data measured at or near the Earth's surface. In geophysics, the data are acquired through methods such as seismic, electromagnetic, gravity and magnetic surveys, or well logging (*Everett*, 2013). These techniques are used to infer information about the physical properties of the Earth, such as elastic properties and electrical resistivity. Similarly, in hydrogeology, the objective is to estimate parameters related to groundwater flow, aquifer properties, or contaminant transport based on measurements collected from monitoring wells, pumping tests, tracer studies, or other hydrogeological measurements. The targeted properties of the hydrogeological system are often those that govern the movement and storage of water or contaminants in the subsurface, for example, hydraulic conductivity, porosity, or dispersivity (*Fetter*, 2018).

The opposite of inverse problems is *forward problems*, in which the response of a system (simulated measurements or observations) is predicted given a set of input parameters.

Contrary to the forward problem, the inverse problem is challenging due to its inherent ill-posed nature (to be well-posed, the solution needs to exist, be unique and stable) (*Mueller and Siltanen*, 2012). There are two main approaches to tackle the inverse problem. Deterministic inversion focuses on finding a single solution for the unknowns by minimizing an objective function subject to constraints (*Parker*, 1994). On the other hand, probabilistic inversion treats the unknown parameters as random variables and targets a range of possible solutions along with their associated probabilities (*Tarantola*, 2005). Bayesian inversion is a specific approach within probabilistic inversion that follows the principles of Bayesian statistics. It is widely used in geophysics and hydrogeology as it offers a comprehensive and flexible framework for solving practical inverse problems. By incorporating prior knowledge and accounting for uncertainty sources, it addresses the challenges of ill-posedness, leading to improved understanding and decision-making in these fields (*Stuart*, 2010).

1.2 Bayesian inversion

Solving inverse problems requires a *parametrization* of the system under study in order to represent it using a finite number of parameters that capture its essential characteristics. The chosen parametrization can significantly impact the success and efficiency of inverse problem-solving as it defines the unknown parameters to be inferred in the process. The number of parameters to estimate is referred to as the *dimension* of the inverse problem. A well-chosen parametrization should strike a balance between simplicity and the ability to capture the relevant features and variability of the system. For geophysical and hydrogeological studies, it usually implies a spatial discretization of the domain under study.

Bayesian inversion enables the incorporation of prior assumptions regarding the system based on previous studies, expert knowledge, physical principles, or other sources of information. The *prior* probability density function (PDF) describes the expected uncertainty and variability of the parameter values before observing the data. It represents the initial belief about the parameters to infer and it can be built relying on empirical data, theoretical considerations, but also subjective judgments (*Scales and Tenorio*, 2001). In this thesis, a *conceptual model* refers to a parametrization of the subsurface together with an associated prior PDF (*Brunetti*, 2018).

Bayesian inversion targets the *posterior* PDF of the system parameters after considering the observed data. If we consider a conceptual model M_k with parameters $\boldsymbol{\theta}$, and a given a set of data observations $\mathbf{y}$, Bayes' theorem is expressed as (*Bayes*, 1763; *Calvetti and Somersalo*, 2007):

$$\pi(\boldsymbol{\theta}|\mathbf{y}, M_k) = \frac{\pi(\boldsymbol{\theta}|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}, M_k)}{\pi(\mathbf{y}|M_k)}, \tag{1.1}$$

where $\pi(\boldsymbol{\theta}|\mathbf{y}, M_k)$ is the posterior PDF and $\pi(\boldsymbol{\theta}|M_k)$ is the prior PDF. The *likelihood function* $p(\mathbf{y}|\boldsymbol{\theta}, M_k)$ assesses the probability that a particular model realization, under a prescribed

error model, has generated the observed data $\mathbf{y}$. Hence, it quantifies the agreement or fit between the model predictions and the actual measurements, taking into account the uncertainties in the observations. When the errors or uncertainties in the observed data are assumed to follow a Gaussian (normal) distribution, the likelihood is given by:

$$p(\mathbf{y}|\boldsymbol{\theta}, M_k) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\boldsymbol{C}|^{-\frac{1}{2}} exp\left(-\frac{1}{2}(F(\boldsymbol{\theta}) - \mathbf{y})^T \boldsymbol{C}^{-1}(F(\boldsymbol{\theta}) - \mathbf{y})\right), \tag{1.2}$$

where n is the number of data points, $\boldsymbol{C}$ the data covariance matrix describing the estimated data uncertainty and correlations and $|C|$ its determinant. The *forward operator* $F(\boldsymbol{\theta})$ is a mathematical model of the physical system that provides the simulated response for a particular set of values assigned to the parameters $\boldsymbol{\theta}$.

The normalizing constant in Bayes' theorem $\pi(\mathbf{y}|M_k)$ is known as the *evidence* or marginal likelihood. The evidence can be challenging to calculate for high-dimensional problems as it is a multi-dimensional integral (as many dimensions as the number of model parameters) over the prior:

$$\pi(\mathbf{y}|M_k) = \int \pi(\boldsymbol{\theta}|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}, M_k)\, d\boldsymbol{\theta}. \tag{1.3}$$

For many applications, the computation of the evidence is not required. The posterior PDF in *Bayesian inference* can then be estimated up to a constant as the following proportionality holds:

$$\pi(\boldsymbol{\theta}|\mathbf{y}, M_k) \propto \pi(\boldsymbol{\theta}|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}, M_k). \tag{1.4}$$

The evidence can be ignored as long as the model parametrization is constant. This is not the case in some Bayesian inversion approaches, for example, transdimensional Markov chain Monte Carlo (*Green*, 1995; *Sisson*, 2005; *Bodin and Sambridge*, 2009), in which the number of parameters to infer is treated as an unknown. When the focus of the study is the uncertainty in the conceptual model, the evidence is the key quantity to estimate. If we are considering different and competitive conceptual models, the evidence can be used to compare and rank them, a process known as *Bayesian model selection* (*Knuth et al.*, 2015).

1.2.1 Monte Carlo-based Bayesian inference

The posterior PDF can only be recovered analytically for very specific scenarios. For more general cases, one needs to explore the parameter space to generate representative posterior samples (*Tarantola*, 2005). Random or pseudo-random explorations can be performed by *Monte Carlo* methods (*Metropolis and Ulam*, 1949), which have been widely used for inverse problems in Earth sciences (*Sambridge and Mosegaard*, 2002). This group of computational algorithms relies on repeated generation of samples from a specific distribution over a given sampling space. Rejection-sampling is a popular Monte Carlo method that generates random samples from a simpler density, and accepts or rejects them such that, provided enough samples, the accepted ones approximate the target distribution (*Von Neumann et al.*, 1951). This method is useful to approximate the posterior distribution in lower dimensions, however, the low probability of accepting samples in higher dimensions makes the process very inefficient.

More efficient explorations of the parameter space are possible by using *Markov chain Monte Carlo* (MCMC) methods (see for example *Gilks et al.* (1995)). The basic idea behind these methods is to build sequences of evolving states of the values taken by the model parameters $\boldsymbol{\theta}$. In these sequences, known as Markov chains, each state depends only on the immediately previous state. In MCMC methods, the Markov chains gradually converge to a unique and stationary distribution that does not depend on their initial state and approximates the posterior PDF, given that the detailed balance condition is fulfilled (*Mosegaard and Sambridge*, 2002):

$$P(\boldsymbol{\theta}_i|\boldsymbol{\theta}_j)h(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j d\boldsymbol{\theta}_i = P(\boldsymbol{\theta}_j|\boldsymbol{\theta}_i)h(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i d\boldsymbol{\theta}_j, \tag{1.5}$$

where $h(\boldsymbol{\theta})$ is the desired equilibrium distribution, and $P(\boldsymbol{\theta}_i|\boldsymbol{\theta}_j)$ is the transition probability distribution that gives the probability of visiting $\boldsymbol{\theta}_i$ when the algorithm is at $\boldsymbol{\theta}_j$.

A proposal distribution is needed to suggest moves within the chains, that is, to generate a candidate set of parameters for the next state. One of the first MCMC approaches is the random walk Metropolis (*Metropolis et al.*, 1953), in which the chains evolve by iteratively proposing moves using a symmetric proposal distribution. These moves are accepted according to the Metropolis rule, that expresses the probability that the chain moves from the current state $\boldsymbol{\theta}_c$ to a proposed state $\boldsymbol{\theta}_p$:

$$\Gamma(\boldsymbol{\theta}_p, \boldsymbol{\theta}_c) = min\left(1, \frac{\pi(\boldsymbol{\theta}_p)p(\mathbf{y}|\boldsymbol{\theta}_p)}{\pi(\boldsymbol{\theta}_c)p(\mathbf{y}|\boldsymbol{\theta}_c)}\right). \tag{1.6}$$

The proposal distribution and its scale play an important role as it directly impacts the acceptance rate. This rate is defined as the ratio of the number of accepted moves to the total number of proposed moves. If the proposal scale is too small, the candidate set of

parameters will be similar to each other. The resulting acceptance rate will be high but the Markov chain becomes slow at exploring different parts of the posterior distribution. On the other hand, if the proposal scale is too high, the acceptance rate will be low and a significant amount of computational resources will be wasted in simulating the forward response of candidate states that are unlikely to be accepted. It is not trivial to define an efficient proposal distribution but it is crucial, especially in higher dimensions when sampling efficiency is computationally needed. Multiple variations of the Metropolis-Hastings algorithm have been proposed that aim at improving the sampling through more complex proposal distributions schemes (*Haario et al.*, 2006; *Bai et al.*, 2011; *Vrugt*, 2016).

Tempering-based methods

There is a need to broaden the scope of inverse problems that can be tackled in terms of complexity, by employing more effective and robust inversion algorithms (*Sambridge*, 2014). Even if MCMC methods have been widely applied in diverse fields, and much effort has been dedicated to improving the schemes and the proposal distributions, the challenge of thoroughly exploring complex posterior PDFs in realistic computational times remains. Relevant factors that contribute to the complexity of an inverse problem are (i) the dimension of the inverse problem and spatial correlation among parameters, (ii) the non-linearity of the associated forward problem, and (iii) the existence of more than one mode in the posterior PDF, known as multimodality. What is observed in practice is that the Markov chains become trapped in regions of local minima for extended periods and are unable to transition between regions of high posterior probability.

To overcome such challenges, a group of methods relying on *tempering* have been proposed. By introducing a variable, known as the temperature, they modify the prominence of the local and global minimas by re-scaling the objective function. *Kirkpatrick et al.* (1983) relied on this idea when developing the popular method of simulated annealing (SA), where a schedule of conditional distributions with decreasing temperatures are sampled to find a global minimum of the objective function. Tempering was first introduced in the geosciences by *Rothman* (1985, 1986) for global optimization of the residual statics in reflection seismology.

Tempering in Bayesian inversion involves altering the influence of the likelihood function. The temperature $\mathscr{T} \geq 1$, or its inverse $\alpha = 1/\mathscr{T}$ $0 \leq \alpha \leq 1$, is introduced to change the likelihood weight in equation 1.4:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\, p(\mathbf{y}|\boldsymbol{\theta})^{\alpha}, \tag{1.7}$$

where we have assumed a fixed conceptual model and suppressed M_k for simplicity. An $\alpha = 0$ means that we sample the prior PDF and an $\alpha = 1$ means that we sample proportional to the posterior PDF. Decreasing α (or increasing the temperature) flattens the likelihood function, such that the considered data have less influence and the sampling gets closer to the prior distribution (see Figure 1.1). In this way, tempering enhances the freedom of exploration of probabilistic samplers and reduces the risk of getting trapped in local minimas. The tempered distributions in equation 1.7 are known as unnormalized *power posteriors*.

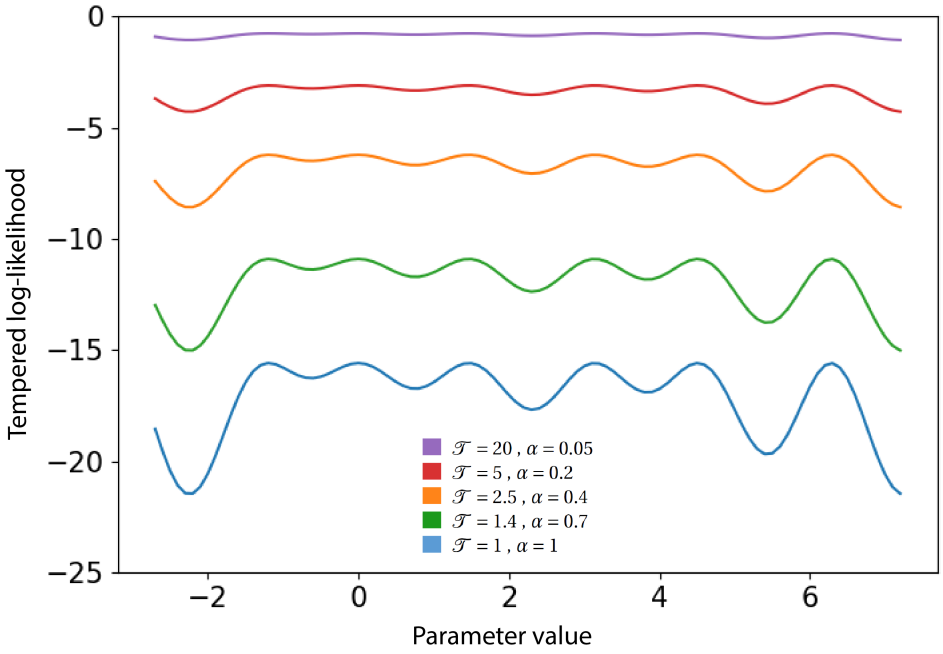

Figure 1.1: Effect of tempering in a 1-D multi-modal Gaussian likelihood. If the temperature increases, α decreases, and the function becomes flatter with less pronounced maximas of the likelihood function (minimas for a corresponding misfit function).

Simulated tempering (ST) (*Marinari and Parisi*, 1992; *Geyer and Thompson*, 1995) is a probabilistic sampling method that relies on a temperature ladder that can either decrease or increase, where the proposed temperature changes are accepted or rejected in a stochastic way. Compared to SA, the detailed balance conditions are ensured when transitioning between temperatures, however, the implementation is difficult as it requires computation of the tempered distributions' normalizing constants for the acceptance rule (*Sambridge*, 2014). A similar method that overcomes this issue is parallel tempering (PT) (*Geyer*, 1991; *Falcioni and Deem*, 1999; *Earl and Deem*, 2005), which relies on MCMC chains that explore individually different power posteriors (each chain has a fixed temperature), instead of a single walker that decreases or increases its temperature. The higher temperature chains are freer to explore the whole space, and by proposing eventual swaps between the states of the chains, the information is transferred to the low temperature chains that can explore the individual modes. Examples of PT applications in the geosciences can be found in *Mohamed et al.* (2012), *Carter and White* (2013), *Sambridge* (2014), *Laloy et al.* (2016), *Wang et al.* (2019), *Reuschen et al.* (2020).

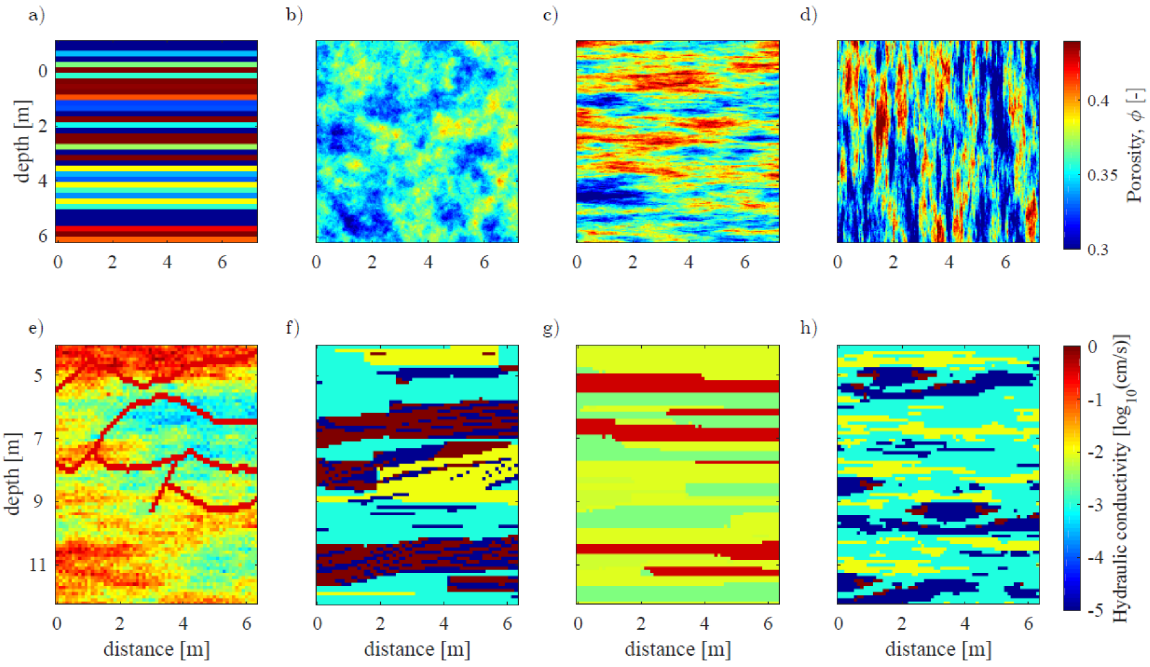

Figure 1.2: Taken from *Brunetti* (2018), examples of alternative conceptual models in a hydrogeological study: (a) horizontal layers, (b-d) multi-Gaussian models with different correlation structures, and (e-h) models obtained through multiple-point statistics approaches.

1.2.2 Bayesian model selection

We rely on models to study the world around us, but we need to keep in mind that *"All models are wrong, but some are useful"* (*Box*, 1979). Indeed, models are idealizations whose purpose is to represent the system we are interested in. It is not possible to validate conceptual models in the geosciences, implying that we should instead try to falsify them (*Popper*, 2005; *Oreskes et al.*, 1994) or compare plausible alternative models in relative terms. Inversion results and associated uncertainty quantification are only meaningful as long as the impact of the underlying assumptions is small. The choice of our conceptual model, sometimes our primary and major assumption when performing inversion, brings with it a great amount of subjectivity when addressing Earth science problems (*Bond et al.*, 2007). A more transparent approach then is to consider competitive conceptual models and quantitatively compare them through Bayesian model selection (see Figure 1.2 from *Brunetti* (2018)).

The evidence (equation 1.3) is the key quantity in Bayesian model selection as it quantifies the consistency of a conceptual model with the observed data (*Kass and Raftery*, 1995). An outstanding property of the evidence is that it honors the principle of parsimony or Occam's Razor, in the sense that it implicitly embodies a trade-off between minimizing model complexity and goodness of fit (*Angluin and Smith*, 1983; *Jefferys and Berger*, 1992; *MacKay*, 2003).

Model selection can be performed by relying on mathematical approximations or numerical approaches. A group of mathematical approximations known as information criteria (IC), for

example Kashyap (KIC, *Kass and Raftery* (1995)), Akaike (AIC, *Poeter and Anderson* (2005)) and Bayesian (BIC, *Gideon* (1978)), have shown to perform inconsistently, leading to contradicting model ranking (*Lu et al.*, 2011; *Schöniger et al.*, 2014). The Laplace-Metropolis method (*Lewis and Raftery*, 1997) evaluated at the maximum a posteriori (MAP) point is a mathematical approximation of the evidence that has shown to perform better than the information criteria (*Schöniger et al.*, 2014). On the other hand, simulation-based approaches can be more reliable but they become computationally demanding in high-dimensions (*Evans and Swartz*, 1995). Examples of simulation-based approaches include brute force Monte Carlo (BFMC) or arithmetic mean approach (*Hammersley*, 1960), Monte Carlo integration with importance sampling (*Hammersley and Handscomb*, 1964) (e.g., Gaussian mixture importance sampling by *Volpi et al.* (2017)), Monte Carlo integration with posterior sampling or harmonic mean approach (*Newton and Raftery*, 1994), nested sampling (*Skilling*, 2006), thermodynamic integration (*Gelman and Meng*, 1998; *Friel and Pettitt*, 2008), stepping stone sampling (*Xie et al.*, 2011), and sequential Monte Carlo (*Zhou et al.*, 2016).

1.3 Adaptive sequential Monte Carlo

1.3.1 Sequential Monte Carlo

A group of algorithms to perform Bayesian inversion is the family of sequential Monte Carlo (SMC) methods (*Chopin*, 2002; *Del Moral et al.*, 2006). The original application of SMC methods was as *particle filters* for data assimilation problems with observations that become available with time. In these problems, the inference is performed on-the-go by updating the posterior of the states using the new data; examples are radar measurements employed to monitor an aircraft's position or stock market data to estimate the volatility of financial instruments (*Doucet et al.*, 2001). Nearly all particle filtering methods can be interpreted as instances of a generic SMC algorithm (*Doucet and Johansen*, 2011). In this thesis, we use the SMC method in the context of tempered exploration for Bayesian inversion, by relying on a sequence of *importance sampling* steps between power posteriors at gradually decreasing temperatures.

1.3.2 Importance sampling

Importance sampling (IS) is a Monte Carlo approach approximating the mathematical expectation of a target distribution by computing a weighted average of random samples drawn from a different distribution known as the importance distribution (*Hammersley and Handscomb*, 1964). While sampling the target distribution is impractical or even impossible, the importance distribution should be easy to sample from. The concept underlying IS is to weight samples obtained from the importance distribution such that they emulate samples derived from the target distribution. If the importance distribution is $q(\boldsymbol{\theta})$ and the target distribution is $l(\boldsymbol{\theta})$, the importance weight w of a sample θ_s is given by:

$$w = \frac{l(\boldsymbol{\theta_s})}{q(\boldsymbol{\theta_s})}. \tag{1.8}$$

The effectiveness of importance sampling relies on the choice of a suitable importance distribution. An ideal importance distribution is slightly wider than the target distribution but with the same mean (*Geweke*, 1989), which ensures that all possible values of the target distribution can be sampled and the variance of the importance weights is not too large.

1.3.3 Sequence of power posteriors

When the importance and target distribution are not sufficiently close, IS may fail to properly estimate the target distribution, a frequent situation when using the prior PDF as the importance distribution to target the posterior PDF (*Agapiou et al.*, 2017). To address this, *Neal* (2001) proposed the annealed importance sampling (AIS) method in which the prior and the posterior PDF are linked through a sequence of intermediate importance sampling steps involving power posteriors at different temperatures. The inverse temperature schedule $\alpha_t \in [0,1]$ defines the sequence of power posteriors $\{\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})\}_{t=0}^{T}$ (equation 1.7). The method is initialized by sampling from the prior $\alpha_{t=0} = 0$, and then gradually considers increases in α, thereby, giving more influence to the likelihood function, until $\alpha_{t=T} = 1$. Each intermediate power posterior is approximated using a number of K Markov steps (iterations of a Markov chain targeting the corresponding power posterior, section 1.2.1), and to transition from one power posterior to the following one, an intermediate importance sampling step is performed.

Both AIS and SMC are particle methods enabling to perform Bayesian inversion. In these approaches, a number of N particles with evolving states run in parallel. Each of the particles targets sequentially the sequence of power posteriors and has an associated weight resulting from the intermediate importance sampling steps.

1.3.4 Incremental weights

If we consider a transition from α_{t-1} to α_t, the corresponding intermediate importance sampling step is performed using the approximation of the power posterior $\gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})$ as the importance distribution for estimating $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$. For a particle i, the *incremental weight* w_t^i resulting from this step is:

$$w_t^i = \frac{\gamma_t(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}{\gamma_{t-1}(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}. \tag{1.9}$$

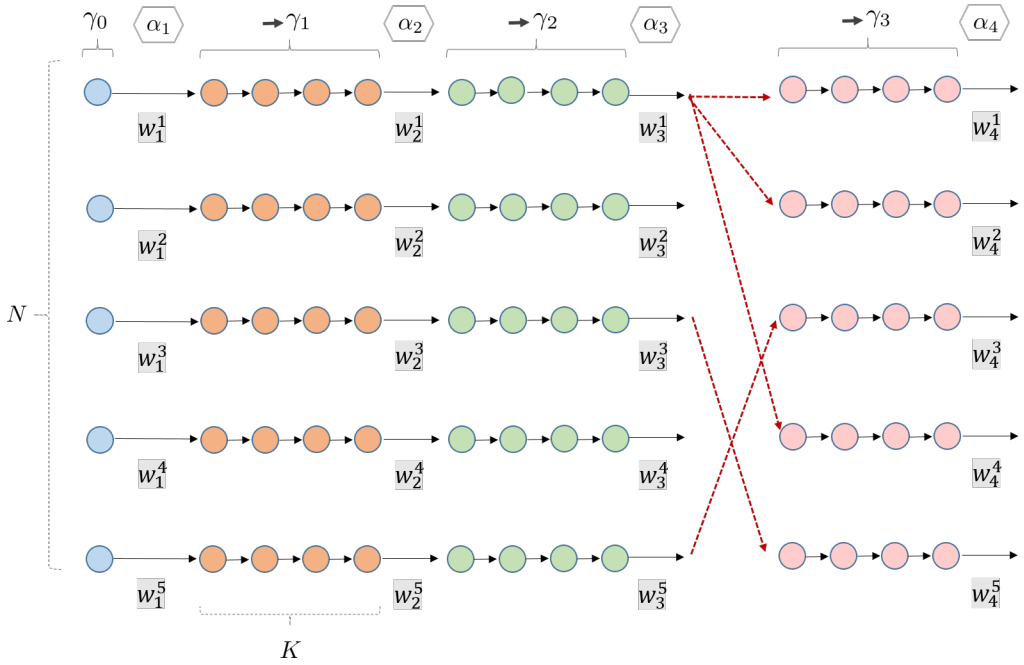

Figure 1.3: Schematic representation of sequential Monte Carlo (SMC) using $N = 5$ particles evolving in parallel. After the initial sampling from the prior PDF (blue circles), $K = 4$ Markov steps are performed to approximate one of the power posteriors γ_t. In these power posteriors, the likelihood is raised to an inverse temperature α_t that increases gradually. At the end of each approximation, an importance sampling step is performed to calculate an incremental weight w_t. A resampling step is indicated by the red dashed lines.

As the sequence of power posterior progresses, the incremental weights are multiplied in the form of *normalized particle weights*:

$$W_t^i = \frac{W_{t-1}^i w_t{}^i}{\sum_{j=1}^N W_{t-1}^j w_t^j}. \tag{1.10}$$

The final normalized particle weights W_T^i are the ones determining the relative weights of each of the particles final states approximating the posterior PDF. Figure 1.3 shows a schematic simplified representation of SMC.

1.3.5 Resampling

The variance of the particle weights is an indicator of the quality of the particle approximation, as a high variance indicates that only a fraction of the particles contribute in a significant way to the approximation. A complication of AIS is that, due to stochastic convergence behaviors of the intermediate MCMC steps, the particle weights tend to diverge as the algorithm progresses. The various SMC methods overcome this issue by incorporating *resampling* steps (Figure 1.3). In these steps, the particles are re-initialized by replicating their states according to their normalized weights, such that they all have the same normalized weight after the resampling step. That is, the states of particles with higher weights are more likely to be replicated than states of particles with very low weights. A comparison of different resampling schemes can be found in *Douc and Cappé* (2005).

In this thesis, we rely on systematic resampling, due to its simple and efficient implementation. Figure 1.4 illustrates the procedure, in which the following steps are followed (*Bolić et al.*, 2004):

1. Draw a random value U from the uniform distribution $\mathscr{U}\left[0, \frac{1}{N}\right]$ and define the points $U^i = U + \frac{(i-1)}{N}$, $i = 1, ..., N$ (red dots in Figure 1.4a).

2. Build the normalized weights W_t^i cumulative function (blue ladder in Figure 1.4a).

3. The number of replications of a particle state in the resampling step is equal to the number of points that fall into the vertical jump of the cumulative function corresponding to that particle.

For example, the state of particle number number 3 is replicated three times, but the state of particle number 4 is lost as no points fall into the corresponding jump (Figure 1.4b).

1.3.6 Posterior estimation

Contrary to MCMC methods, the posterior distribution in particle methods as SMC is not approximated by a cumulative set of samples gathered as the chain progresses with equal weights. In SMC methods, if N particles are used, the initial normalized weights are all equal to $1/N$. These weights evolve and, at the end of the SMC algorithm, there is a set of N particle states, each one with a corresponding normalized final weight W_T^i. The W_T^i give the relative weights to the final particle states to approximate the posterior distribution. Figure 1.5 illustrates the difference between how SMC and MCMC results are used to estimate the posterior PDF. Evidently, the number of particles used in particle approaches is a key variable that impacts the posterior estimation quality.

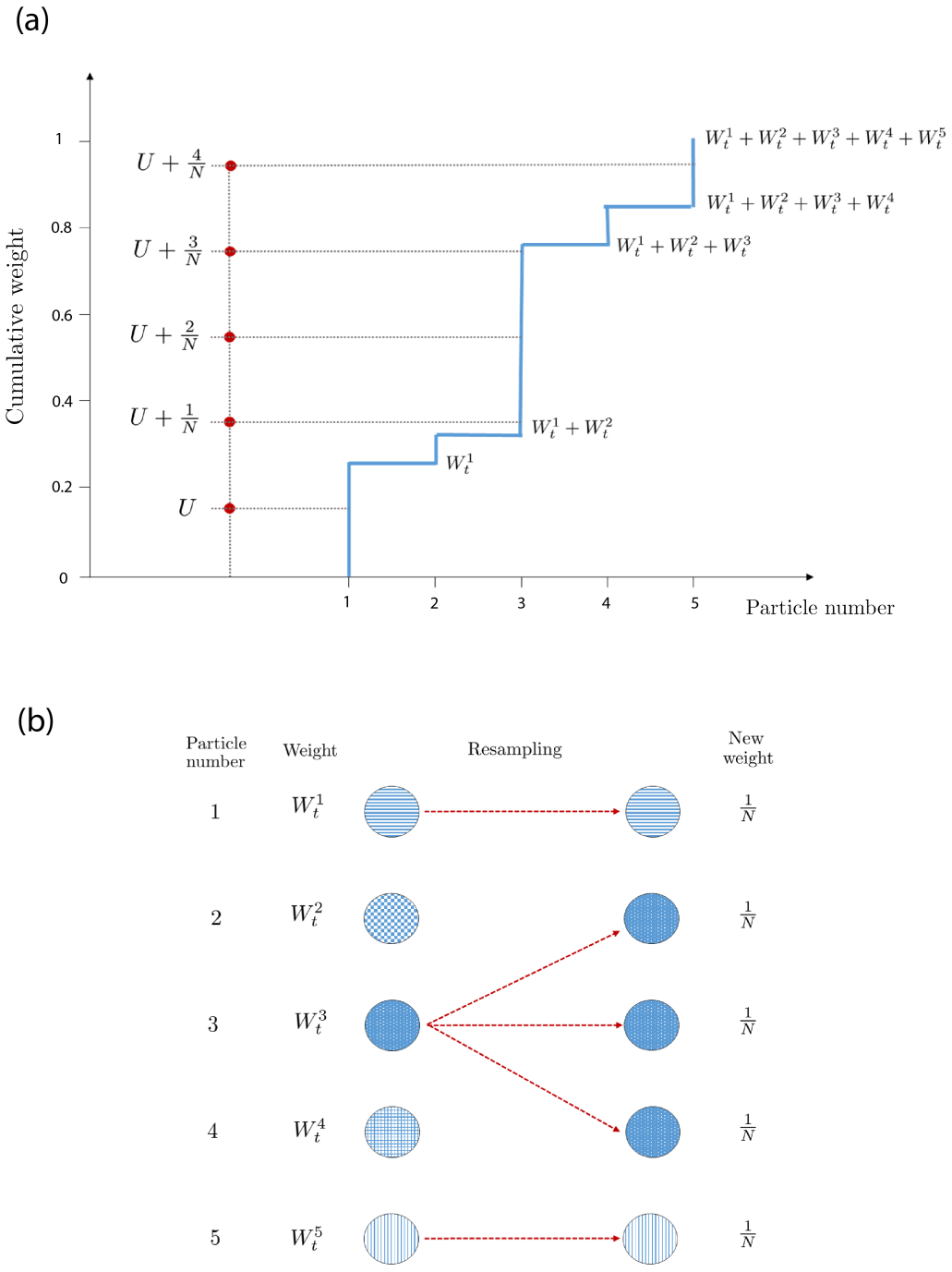

Figure 1.4: Simplified example of a systematic resampling step using $N = 5$ particles: (a) a random number U is drawn from $\mathcal{U}\left[0, \frac{1}{N}\right]$ to define the points $U^i = U + \frac{(i-1)}{N}$ (red dots), the number of replications is equal to the number of points falling in the jump corresponding to that particle in the normalized weights cumulative function (blue ladder), and (b) the particles are replicated and the normalized weights are reinitialized to $\frac{1}{N}$.

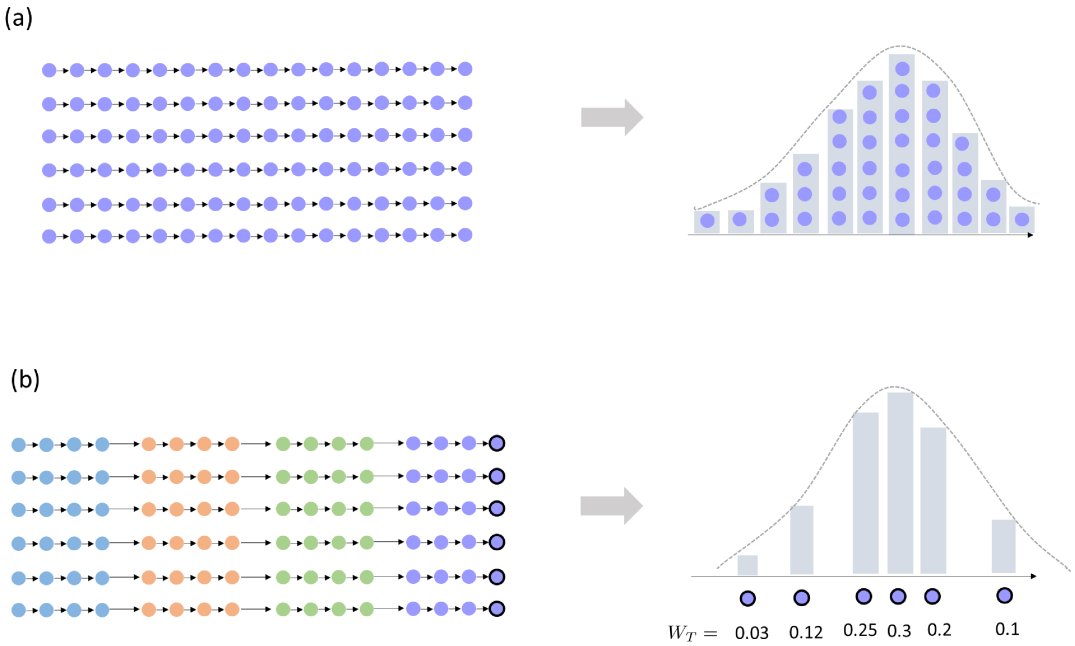

Figure 1.5: Diagram representing the difference between (a) MCMC and (b) SMC posterior approximations. The circles represent the chains/particle states and the color denotes the temperature. In MCMC, all chains target the posterior (equivalent to an inverse temperature $\alpha = 1$, purple circles). After an initial burn-in period that needs to be discarded, every sample (regardless of the chain and the location within the chain) is used to approximate the posterior with the same weight. MCMC samples are often highly correlated and generally long chains are needed to obtain a sufficient number of independent samples (*Kass et al.*, 1998). On the other hand, SMC methods target sequentially power posteriors at gradually-increasing α_t, and when $\alpha = 1$, the final normalized weight of a given particle describes its weight in the particle approximation. For illustrative purposes, we only used six particles, although usually more particles are used in practice.

1.3.7 Evidence estimation

An advantage of SMC over MCMC algorithms is that they also provide an estimation of the evidence. This means that with (almost) the same computational cost, we obtain both a numerical approximation of the posterior PDF and the evidence, making it a flexible method to perform Bayesian inference, Bayesian model selection or both.

The normalizing constant of an unnormalized power posterior $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$ is given by $Z_t = \int \gamma_t(\boldsymbol{\theta}_t|\mathbf{y})\, d\boldsymbol{\theta}_t$. Following *Del Moral et al.* (2006), the ratio of the normalizing constants of two consecutive power posteriors in SMC can be approximated as:

$$\frac{Z_t}{Z_{t-1}} \approx \sum_{i=1}^{N} W_{t-1}^{i}\, w_t^i. \tag{1.11}$$

Given that the prior PDF integrates to 1 ($Z_0 = 1$), the evidence (equation 1.3) can be estimated as:

$$\pi(\mathbf{y}) = Z_T = \frac{Z_T}{Z_0} = \prod_{t=1}^{T} \frac{Z_t}{Z_{t-1}} \approx \prod_{t=1}^{T} \sum_{i=1}^{N} W_{t-1}^{i}\, {w_t}^i. \tag{1.12}$$

In practice, the evidence is updated using the evolving particle incremental and normalized weights as the inverse temperature increases. Compared to SMC, alternative MCMC tempering-based methods may require a post-processing stage to calculate the evidence, for example, using parallel tempering results to approximate the evidence by thermodynamic integration (*Bailer-Jones*, 2015).

1.3.8 Adaptive scheme

In this thesis, we refer to *adaptive* sequential Monte Carlo (ASMC) following *Zhou et al.* (2016), as the SMC method that combines automatic tuning of both the proposal scale and the sequence of temperatures, with decisions on when to perform resampling steps. Note that ASMC is used in the literature to denote any SMC algorithm in which adaptivity is included.

Resampling criterion

The divergence of the weights, also known as *sample impoverishment* or *weight degeneracy*, can lead to many of the particles having extremely low normalized importance weights and, therefore, not making a substantial contribution to approximating the target distribution (*Cappé et al.*, 2005). In this thesis, we adopt a scheme in which decisions to perform resampling are based on the effective sample size (ESS):

$$ESS_t = \frac{(\sum_{i=1}^{N} W_{t-1}^i w_t^i)^2}{\sum_{j=1}^{N} (W_{t-1}^j)^2 (w_t^j)^2}. \tag{1.13}$$

The ESS indicates the number of samples that effectively contribute to the particle approximation. The ESS is initially N and decreases as the power posterior sequence progresses and the weights diverge; resampling steps are performed when the ESS falls below a pre-defined fixed threshold.

Temperature ladder

One common challenge of tempering-based methods is the definition of the temperature ladder. Determining an optimal sequence of temperatures to efficiently sample the space may require substantial user effort. In SMC, the inverse temperature increments can be determined on-the-go by relying on the conditional effective sample size (CESS) describing the quality of the importance sampling distribution:

$$CESS_t = N \frac{(\sum_{i=1}^{N} W_{t-1}^i w_t^i)^2}{\sum_{j=1}^{N} W_{t-1}^j (w_t^j)^2}. \tag{1.14}$$

While the ESS (equation 1.13) considers the accumulated discrepancy between the importance and target distributions, the CESS focuses on the following IS step (*Zhou et al.*, 2016). In practice, one can easily rely on an algorithm that searches between possible inverse temperature increments $\Delta\alpha_t = \alpha_t - \alpha_{t-1}$, to obtain the CESS that is the closest to a pre-defined value.

Proposal scale

Changing the scale of the proposal distribution in MCMC algorithms violates the detailed balance condition (equation 1.5). In SMC methods, however, tuning the proposal scale for each power posterior using the results from previously considered temperatures does not, as it is only used to perform the K Markov steps to approximate the importance distribution for the next intermediate importance sampling step (*Jasra et al.*, 2011; *Del Moral et al.*, 2012). In ASMC, we consider an adaptive tuning where the proposal scale can be modified by a certain percentage. The proposal scale used to approximate a power posterior is tuned depending on the acceptance rate of the K Markov steps approximating the previous one: if the acceptance rate reaches some above (below) pre-defined threshold, the proposal scale is increased (reduced).

1.4 Geologically-realistic priors

Bayesian inversion is strongly influenced by the choice of the conceptual model. Ideally, a prior PDF should incorporate any relevant knowledge about the system. Nonetheless, reliable prior information is not always available, and it might be challenging to convert the information into a PDF (*Linde et al.*, 2015). If available, including reliable information within the prior should be encouraged as it can strongly reduce the complexity of solving the inverse problem and the results will be more useful (*Hansen et al.*, 2012). When limited prior information is available, it might be tempting to perform the inversion with few prior assumptions. However, the resulting realizations will often have no resemblance to geological formations and the results will, hence, be of limited value. One solution is to acknowledge that the prior is uncertain and that it is rather to be considered as one of an infinite number of alternative conceptual models. By then comparing the evidence for the different priors, one can then determine the most suitable conceptual model using Bayesian model selection (*Linde*, 2014; *Brunetti et al.*, 2017, 2019).

The choice of a suitable and useful prior is not trivial and highly problem-dependent. Modeling the spatial statistics of subsurface heterogeneity has traditionally been approached using geostatistical models relying on two-point statistics, for example, multi-Gaussian models (*Kitanidis*, 1997). One advantage of this type of priors is that they are parametric, meaning that they can be represented with mathematical expressions allowing to evaluate the prior probability of any set of parameter values. Unfortunately, multi-Gaussian models are generally insufficient to represent complex geological structures and account for connectivity patterns (*Gómez-Hernández and Wen*, 1998; *Zinn and Harvey*, 2003; *Renard and Allard*, 2013). This limitation can lead to inaccurate predictions and underestimation of the uncertainty.

Alternative approaches have been proposed to integrate geological knowledge and account for enhanced complexity in the prior representation (*Linde et al.*, 2015). One option is to rely on typically two-dimensional or three-dimensional conceptualizations of the subsurface known as *training images* (TI). These continuous or categorical images are obtained from diverse sources like remote sensing data, geological maps, physics-based modeling of geological processes, observed outcrops, or other available information. Their objective is to capture the main subsurface spatial structures and heterogeneity patterns. To generate realizations that honor the higher-order statistics and patterns of the TI, there is a group of algorithms that rely on multiple-point statistics (MPS)(*Mariethoz and Caers*, 2014). Even if MPS-based prior models conceptualized by TIs are generally non-parametric or implicit, they have been successfully incorporated in Bayesian inversion by sequential resampling schemes related to Gibbs sampling (*Fu and Gómez-Hernández*, 2009; *Mariethoz et al.*, 2010a; *Hansen et al.*, 2012; *Ruggeri et al.*, 2015) or patch-based approaches influenced by texture synthesis (*Zahner et al.*, 2016).

On the other hand, artificial intelligence has been growing and revolutionizing most research fields including the geosciences (*Karpatne et al.*, 2018; *Bergen et al.*, 2019). Deep learning is a subset of machine learning that focuses on the development and training of artificial neural networks with multiple layers (*LeCun et al.*, 2015). Deep generative neural networks trained on TIs learn a transform between a low-dimensional latent space with a prescribed prior

and the original image space (*Laloy et al.*, 2017, 2018; *Mosser et al.*, 2017, 2020). However, even if there is a substantial dimensionality reduction as the inversion is performed over the latent parameters, there is an increase in the non-linearity inherited from the neural network transformation.

Despite current challenges, both traditional MPS algorithms and generative neural networks enable the use of more realistic prior models and, consequently, lead to more representative estimations and predictions resulting from the inversion process.

1.5 Surrogate forward modeling

Both SMC and MCMC inversions require repeated evaluations of the forward operator. Forward solvers have been developed to mimic as accurately as possible the system response, but they are generally rather computationally expensive. In this thesis, we refer to this type of solvers as *high-fidelity* solvers. Alternative solvers that aim to emulate the high-fidelity ones at lower computational times are known as *surrogate*, proxy or meta solvers. The objective of surrogate solvers is to accelerate the response computation while maintaining reasonable accuracy. Surrogate models are generally constructed based on a limited set of input-output data points generated from the high-fidelity solver. Although widespread utilization of surrogate modeling in geophysical and hydrogeological inversion has started only recently, they enable the opportunity to significantly lower the computational demand of Bayesian inversion schemes (*Linde et al.*, 2017).

Replacing a high-fidelity solver with a surrogate solver in a Bayesian inversion usually introduces some degree of approximation error. The accuracy of the surrogate model depends on the quality and representativeness of the training data, the problem considered and the surrogate modeling technique used. Validation and verification procedures are typically performed to assess the accuracy and reliability of the surrogate model before deploying it for further analysis. In Bayesian inversion, ignoring this model error can lead to inaccurate estimations and it is recommended to account for it within the likelihood function (*Kaipio and Somersalo*, 2007; *Nissinen et al.*, 2007; *Hansen et al.*, 2014). Careful combination of high- and low-fidelity simulations in Bayesian inversion within specialized algorithms is an attractive strategy that allows accelerating the computation while ensuring accuracy and convergence (*Peherstorfer et al.*, 2018). We refer to these hybrid schemes as *multifidelity* approaches. An example is two-stage Markov chain Monte Carlo (MCMC), in which high-fidelity simulations are computed on samples that have been pre-accepted by the low-fidelity solver (*Christen and Fox*, 2005).

There are different approaches to create surrogate models. Kriging-based approaches involve modeling the spatial correlation between data points to predict values at unobserved locations (e.g, Gaussian Process Regression, *Santner et al.* (2003), *Rasmussen* (2004)). These are localized surrogates, as they rely on local information to make accurate predictions near the points of the training set. Localized techniques are better for interpolation, whereas global surrogates are better for extrapolation, giving them higher global accuracy (*Marelli et al.*, 2021). Examples of global surrogates are those obtained with deep-learning, which have been

used, for example, to model reservoir subsurface flow in geophysics (*Jin et al.*, 2020). On the other hand, polynomial chaos expansion (PCE) (*Xiu and Karniadakis*, 2002; *Blatman and Sudret*, 2011) is a mathematical approach that approximates the high-fidelity solver through its spectral representation in an orthonormal polynomial basis. The polynomial expansion can be truncated at a certain order to balance accuracy and computational efficiency. Combinations of localized and global approaches have been proposed to train surrogates with good extrapolation and local accuracy, for example, PCE-Kriging (*Schobi et al.*, 2015; *Kersaudy et al.*, 2015). In this thesis, we use regression-based sparse (meaning that most of the coefficients are zero) PCE surrogate models because they have shown strong extrapolation capabilities and are robust with respect to noise (*Blatman and Sudret*, 2011; *Lüthen et al.*, 2021; *Marelli et al.*, 2022; *Meles et al.*, 2022). PCE have been successfully implemented for surrogate models within various Bayesian inversion frameworks (*Marzouk et al.*, 2007; *Marzouk and Xiu*, 2009; *Wagner et al.*, 2020; *Meles et al.*, 2022).

1.6 Objectives

Bayesian inversion provides a comprehensive framework to solve complex inverse problems where different sources of uncertainty can be accounted for. For this reason, it has become an essential tool in the field of geophysics and hydrogeology, allowing to infer subsurface properties, make predictions on the studied systems, quantify the estimation uncertainties, and compare competitive conceptual models. Current challenges when relying on the Bayesian framework to solve inverse problems include the difficulty of properly exploring the different regions of the posterior PDF, especially when the problems are high-dimensional, multi-modal, and non-linear. Even if Bayesian approaches are powerful tools, the counterpart of most algorithms is the high computational demand associated with the large number of forward evaluations needed. There is, therefore, a necessity of improving existing methods, testing approaches developed in other contexts and developing new strategies to enhance the capacity and efficiency of exploration. On the other hand, geoscientists have a growing interest in building realistic conceptual models that resemble natural structures and behavior. The definition and extent of the prior PDF that can be considered in inversion studies is expanding due to developments in MPS and deep learning. Accounting for geological realism while defining a prior that can be properly and efficiently incorporated in a Bayesian inversion is a current challenge.

In this thesis, we explore the ASMC method, to address the following objectives:

- evaluate the ability of ASMC to estimate the posterior PDF and the evidence in challenging synthetic hydrogeological and geophysical test cases,
- compare the performance with MCMC: standard MCMC, adaptive MCMC and parallel tempering,
- account for geological realism in conceptual models by multiple-point statistics and deep-learning based priors,

- seek ASMC speed-ups by proposing a multifidelity framework in which we combine high-fidelity and low-fidelity solvers aiming for both a fast and accurate inversion.

The use of classical SMC and particle filter approaches in geophysics and hydrology (*Van Leeuwen*, 2009; *Werner et al.*, 2011; *Vrugt et al.*, 2013; *Khaki et al.*, 2019; *Veettil and Clark*, 2020; *Miltenberger et al.*, 2023), as well as adaptive SMC for Bayesian inversion purposes (*Jeremiah et al.*, 2011, 2012; *Zhu et al.*, 2018; *Lee et al.*, 2020; *Davies et al.*, 2023) is, to the best of our knowledge, still very limited. Through this thesis we aim to present the advantages, potential and limitations of ASMC through geophysical and hydrogeological test cases, hoping to motivate and encourage its application and its testing for inversion and model selection in complex real-case scenarios throughout the geosciences.

1.7 Outline

To address the presented objectives, this thesis includes two published articles (Chapters 2 and 3) and a third to be submitted (Chapter 4). Below is a brief description of its structure:

- In **Chapter 2**, we implement the ASMC method in a synthetic GPR crosshole tomography problem to estimate the posterior PDF and the evidence. We compare the performance with MCMC using two alternative schemes for model proposals within the inversion: a basic uncorrelated Gaussian and an adaptive DREAM$_{(ZS)}$ algorithm. Two different conceptual models with priors encoded by spatial generative adversarial neural networks (SGANs) are considered and the evidences are evaluated for different scenarios. The uncertainty of the evidence estimations is assessed by repetitive runs and a modified single-run estimator that is proposed for the first time.

- In **Chapter 3**, we consider a challenging hydrogeological groundwater transport inverse problem, designed to have a bimodal posterior distribution. The results are compared with results obtained using parallel tempering for a similar computational budget. We rely on a MPS-based proposal scheme named sequential geostatistical resampling. We investigate the relative sampling of the two modes as the temperature progresses and discuss how ASMC intermediate results can be interpreted as corresponding to assumptions of higher data errors.

- In **Chapter 4**, we propose speed-ups in the ASMC inversion considering surrogate modeling and two types of high-fidelity solvers. We compare the performance when using surrogates trained with prior samples against an algorithm in which the surrogate is updated on-the-go. We combine tempering steps with surrogate updates and propose a final transition to the high-fidelity solver. Through this multifidelity framework, we seek computational time reductions without compromising accuracy.

- In **Chapter 5:** we summarize the main conclusions of the thesis, discuss the main implications, highlight the advantages and challenges of implementing ASMC in geophysical and hydrological inverse problems, and propose an outlook.

Contribution statement: I had the main responsibility for the development of the projects (chapters 2, 3 and 4) including conceptualization, methodology, software implementation and writing the manuscript drafts.

Code availability: The codes and test examples of the projects (chapters 2, 3 and 4) are available in the following GitHub repository: https://github.com/amaya-macarena.

Chapter 2

Adaptive sequential Monte Carlo for posterior inference and model selection among complex geological priors

---

Macarena Amaya, Niklas Linde and Eric Laloy.

Published[1] in *Geophysical Journal International* and herein slightly adapted to fit the theme of this thesis.

[1]Amaya et al. (2021). Adaptive sequential Monte Carlo for posterior inference and model selection among complex geological priors. *Geophysical Journal International*, **226**(2) 1220–1238, doi:10.1093/gji/ggab170

Abstract

Bayesian model selection enables comparison and ranking of conceptual subsurface models described by spatial prior models, according to the support provided by available geophysical data. Deep generative neural networks can efficiently encode such complex spatial priors, thereby, allowing for a strong model dimensionality reduction that comes at the price of enhanced non-linearity. In this setting, we explore a recent adaptive sequential Monte Carlo (ASMC) approach that builds on annealed importance sampling (AIS); a method that provides both the posterior probability density function (PDF) and the evidence (a central quantity for Bayesian model selection) through a particle approximation. Both techniques are well suited to parallel computation and rely on importance sampling over a sequence of intermediate distributions, linking the prior and the posterior PDF. Each subsequent distribution is approximated by updating the particle weights and states, compared with the previous approximation, using a small pre-defined number of Markov chain Monte Carlo (MCMC) proposal steps. Compared with AIS, the ASMC method adaptively tunes the tempering between neighboring distributions and performs resampling of particles when the variance of the particle weights becomes too large. We evaluate ASMC using two different conceptual models and associated synthetic cross-hole ground penetrating radar (GPR) tomography data. For the most challenging test case, we find that the ASMC method is faster and more reliable in locating the posterior PDF than state-of-the-art adaptive MCMC. The evidence estimates are found to be robust with respect to the choice of ASMC algorithmic variables and much less sensitive to the model proposal type than MCMC. The variance of the evidence estimates are best estimated by replication of ASMC runs, while approximations based on single runs provide comparable estimates when using a sufficient number of proposal steps in approximating each intermediate distribution.

2.1 Introduction

Markov chain Monte Carlo (MCMC) methods are, for strongly non-linear inverse problems and a limited computational budget, not always able to locate the posterior probability density function (PDF) of interest or to explore it sufficiently. Parallel tempering (*Earl and Deem*, 2005) is a well-known approach to circumvent such issues and it was popularized in geophysics by *Sambridge* (2014). Parallel tempering runs multiple interacting chains targeting a sequence of power posteriors including faster moving chains at higher temperatures (i.e., corresponding to less weight being given to the likelihood function). Such chains help to locate significant modes of the posterior distribution that can, through a swapping mechanism, be explored by the chain targeting the posterior distribution of interest for which the temperature is 1. The resulting increase in the ability to bypass local minima and explore multimodal distributions is offset by the need for many parallel chains and a carefully-tailored temperature sequence to ensure efficient mixing among chains.

Neal (2001) introduced the annealed importance sampling (AIS) method, which is also well suited to derive information about the posterior PDF of interest when confronted with highly non-linear or multi-modal inverse problems. AIS is a particle method in which many

particles (the evolution of each particle is represented by an individual chain) are evolving in parallel. Particle methods rely on the states and weights of a collection of evolving particles to approximate distributions of interest. This is in contrast to MCMC methods in which all states have the same weight and the distribution of interest is approximated by proposal and acceptance mechanisms ensuring that sampling is proportional to the posterior probability density. In developing AIS, *Neal* (2001) demonstrates how intermediate results obtained by simulated annealing (*Kirkpatrick et al.*, 1983), typically used for global optimisation, can be re-interpreted as a sequence of importance sampling steps from approximations of intermediate posterior PDFs at gradually decreasing temperatures (i.e., annealing), thereby, creating a succession of approximations of intermediate distributions between the prior to the posterior distribution of interest. This method has several attractive properties: (1) it inherits from simulated annealing the ability to bypass problems with local minima by initially allowing large steps and efficient exploration before focusing on a more detailed search in areas of high posterior probability; (2) it is well suited for parallelization; (3) the final states and their associated importance weights approximate the posterior distribution; and (4) it offers directly an approximation of the evidence, the central quantity in Bayesian model selection.

Even if AIS is still widely used, it suffers from two main deficiencies: (1) it is very challenging to pre-define an appropriate annealing sequence (i.e., the sequence of inverse temperatures to which the likelihood function is raised) and (2) the populations of importance weights have increasingly higher variances as the AIS run progresses, thereby, increasing the risk of obtaining poor estimates of the posterior PDF and the evidence. Sequential Monte Carlo (SMC) (*Doucet and Johansen*, 2011) represents a family of particle methods that are widely used in science and engineering, particularly for data assimilation tasks, but their use in geophysics has been limited to date (see review by *Linde et al.* (2017)). At the most basic level, SMC relies on importance sampling combined with resampling steps which ensures that the particle approximation of the high-dimensional posterior PDF is of sufficient quality. The resampling step tends to reinitialize particles of low probability by states of higher probability, thereby avoiding that computational time is wasted in regions of low posterior density. *Zhou et al.* (2016) proposed an adaptive SMC algorithm (referred to hereafter as ASMC) that addresses the limitations of AIS stated above by adaptively tuning the progression between intermediate distributions and by resampling when the variance of the particle weights becomes too large.

The prior PDF has a strong impact on Bayesian geophysical inversion results (*Hansen et al.*, 2012) and should reflect the existing geological knowledge at a site (see review by *Linde et al.* (2015)). One effective way of encoding prior knowledge in a low-dimensional latent vector of uncorrelated parameters is offered by deep generative neural networks (*Goodfellow et al.*, 2014). *Laloy et al.* (2017) and *Laloy et al.* (2018) demonstrated using variational autoencoders (*Kingma and Welling*, 2013) and generative adversarial networks (GAN) (*Goodfellow et al.*, 2014), respectively, that the generated realizations of such networks are of high quality and that inversion can be successfully performed on this latent space. The challenge when working with deep generative neural networks is the highly non-linear transform linking the latent variables to the image representation (i.e., the typically gridded model of physical properties). This non-linearity often leads to poor and unreliable convergence when applying

gradient-based optimization methods (*Laloy et al.*, 2019) and inversion on such latent spaces may challenge state-of-the-art MCMC algorithms (*Laloy et al.*, 2018).

Here, we explore the performance of the ASMC method (*Zhou et al.*, 2016) when used together with deep generative networks to approximate evidences and posterior distributions using geophysical data. As examples, we consider crosshole geophysical ground-penetrating radar (GPR) data and GAN-based priors, which implies highly non-linear and challenging inverse problems. In ASMC, the approximations of intermediate posterior distributions is achieved by successively, at each temperature, performing a small number of Markov steps. As model proposals, we consider both an elaborate proposal scheme influenced by evolutionary algorithms and a basic uncorrelated Gaussian proposal. Through these examples, we demonstrate that the ASMC method is: (1) easy to implement in existing MCMC algorithms; (2) well-suited for parallelization; (3) robust to parameter settings and model proposal schemes; (4) providing posterior approximations that can be superior to those offered by state-of-the-art MCMC; and (5) deriving accurate evidence estimations without strong distributional assumptions.

2.2 Method

In our method description below, we rely largely on the notation of *Zhou et al.* (2016) who introduced the ASMC algorithm.

2.2.1 Bayesian inference and model comparison

Bayes' theorem expresses the posterior PDF of a conceptual model M_k with parameters $\boldsymbol{\theta}$, given a set of observations $\mathbf{y}$ as:

$$\pi(\boldsymbol{\theta}|\mathbf{y}, M_k) = \frac{\pi(\boldsymbol{\theta}|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}, M_k)}{\pi(\mathbf{y}|M_k)}. \tag{2.1}$$

All the knowledge about the model parameters that is available before considering the data is encapsulated in the prior PDF $\pi(\boldsymbol{\theta}|M_k)$. The likelihood function $p(\mathbf{y}|\boldsymbol{\theta}, M_k)$ quantifies how likely it is that a specific model realization gave rise to the observations when considering a prescribed error model. The normalizing constant $\pi(\mathbf{y}|M_k)$ is referred to as the evidence or the marginal likelihood, and it is a multidimensional integral over the parameter space:

$$\pi(\mathbf{y}|M_k) = \int \pi(\boldsymbol{\theta}|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}, M_k)\, d\boldsymbol{\theta}. \tag{2.2}$$

The evidence quantifies the support provided by the data to the conceptual model under consideration, as formalized by the prior PDF, and can be used to rank different conceptual models. *Schöniger et al.* (2014) describe and compare different methods to estimate the evidence and found that numerical approaches generate more reliable estimates than mathematical approximations of equation 2.2 that yield analytical expressions. Recent studies comparing state-of-the-art approaches to evidence estimation in geophysical and hydrogeological contexts include *Brunetti et al.* (2017) and *Brunetti et al.* (2019).

2.2.2 Adaptive sequential Monte Carlo

Importance sampling

Brute force Monte Carlo (BFMC), also known as the arithmetic mean approach, evaluates many realizations drawn from the prior and the corresponding evidence estimate is their average likelihood. Unfortunately, BFMC suffers from the curse of dimensionality (*Curtis and Lomax*, 2001) in that most draws from the prior, when considering a handful or more unknown model parameters and high-quality data, have negligible likelihoods. Consequently, high likelihood regions contributing strongly to the mean are poorly sampled, leading to high-variance evidence estimates and frequent underestimation of evidence values as demonstrated by *Brunetti et al.* (2017). Throughout this manuscript, a high-variance estimate refers to that obtained by estimators of a mean quantity (e.g., the mean of the sampled likelihoods) for which repeated estimations lead to widely different estimates.

Compared to BFMC, importance sampling offers lower-variance estimates, whereby Monte Carlo samples are drawn proportionally to a so-called importance distribution $q(\boldsymbol{\theta}, M_k)$ (*Hammersley and Handscomb*, 1964). In order to sample regions with a high contribution to the mean, this distribution is chosen to be as close as possible to the target distribution; in this case the posterior PDF. To account for the biased sampling procedure, every sample θ^i drawn from $q(\boldsymbol{\theta}, M_k)$ is associated with an importance weight defined as

$$w^i = \frac{\pi(\boldsymbol{\theta}^i|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}^i, M_k)}{q(\boldsymbol{\theta}^i, M_k)}, \tag{2.3}$$

that determines the corresponding weight in the mean estimation. Assuming that $q(\boldsymbol{\theta}, M_k) \neq 0$ whenever $\pi(\boldsymbol{\theta}|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}, M_k) \neq 0$, and if the number of draws $N \to \infty$, then the following approximation holds (*Neal*, 2001):

$$\frac{\sum_{i=1}^{N} w^i}{N} \approx \frac{\int \pi(\boldsymbol{\theta}|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}, M_k)\, d\boldsymbol{\theta}}{\int q(\boldsymbol{\theta}, M_k)\, d\boldsymbol{\theta}}. \tag{2.4}$$

In the particular case of using the prior as the importance distribution (equivalent to BFMC) and noting that its evidence is equal to one (the integral of the prior PDF is 1), the evidence of M_k is approximated by the mean of the N weights:

$$\pi(\mathbf{y}|M_k) = \frac{\int \pi(\boldsymbol{\theta}|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}, M_k)\, d\boldsymbol{\theta}}{\int \pi(\boldsymbol{\theta}, M_k)\, d\boldsymbol{\theta}} \approx \frac{\sum_{i=1}^{N} w^i}{N} = \frac{\sum_{i=1}^{N} \frac{\pi(\boldsymbol{\theta}^i|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}^i, M_k)}{\pi(\boldsymbol{\theta}^i|M_k)}}{N} = \frac{\sum_{i=1}^{N} p(\mathbf{y}|\boldsymbol{\theta}^i, M_k)}{N}, \tag{2.5}$$

which reduces to the average of the sampled likelihood as discussed above. The importance distribution strongly influences the accuracy of importance sampling and unreliable high-variance estimates are obtained when the importance distribution is far from the target distribution. Therefore, if the prior PDF is markedly different from the posterior PDF, then the quality of the evidence estimate in equation 2.5 is low. Below, we explain how to obtain low-variance estimates of evidences by relying on a succession of importance sampling steps with importance distributions that are close to intermediate target distributions known as power posteriors.

Annealed importance sampling (AIS)

Simulated annealing (*Kirkpatrick et al.*, 1983) is a well-known global optimizer that bypasses local minima by gradually reducing the parameter space exploration using a sequence of intermediate target distributions (i.e., power posteriors characterized by an annealing scheme of successively decreasing temperatures). In developing AIS, *Neal* (2001) took advantage of this sequence of transitional target distributions starting at the prior PDF (infinite temperature) and ending at the posterior PDF (temperature of 1). The algorithm runs in parallel with each chain being interpreted as a particle with an evolving weight and state. From the resulting sequence of intermediate importance weights and states, it is possible to estimate both the posterior PDF and the evidence. AIS shares all the exploratory advantages of simulated annealing and allows for, potentially, high-quality posterior PDF and evidence estimations by creating a smooth path between the prior and the posterior PDF. A schematic visualization of AIS is given in Figure 2.1a.

In the following, we consider a given conceptual model M_k and suppress the corresponding subindex k for simplicity. The unnormalized power posterior PDFs $\left\{\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})\right\}_{t=0}^{T}$ are:

$$\gamma_t(\boldsymbol{\theta}_t|\mathbf{y}) \equiv \pi(\boldsymbol{\theta}_t)\, p(\mathbf{y}|\boldsymbol{\theta}_t)^{\alpha_t}, \tag{2.6}$$

where $\pi(\boldsymbol{\theta}_t)$ is the prior probability density function and $p(\mathbf{y}|\boldsymbol{\theta}_t)$ the likelihood. The annealing schedule $\alpha_t \in [0,1]$ of inverse temperatures defines these power posteriors, where $\alpha_{t=0} = 0$ gives the prior and $\alpha_{t=T} = 1$ the posterior PDF. At small α_t, the contribution of the likelihood is small and the corresponding power posterior is close to the prior PDF. As α_t grows, the influence of the likelihood on the power posterior increases. We denote Z_t as the normalizing constant of the corresponding power posterior, implying that the normalized power PDF is:

$$\pi_t(\boldsymbol{\theta}_t|\mathbf{y}) = \frac{\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})}{Z_t}. \tag{2.7}$$

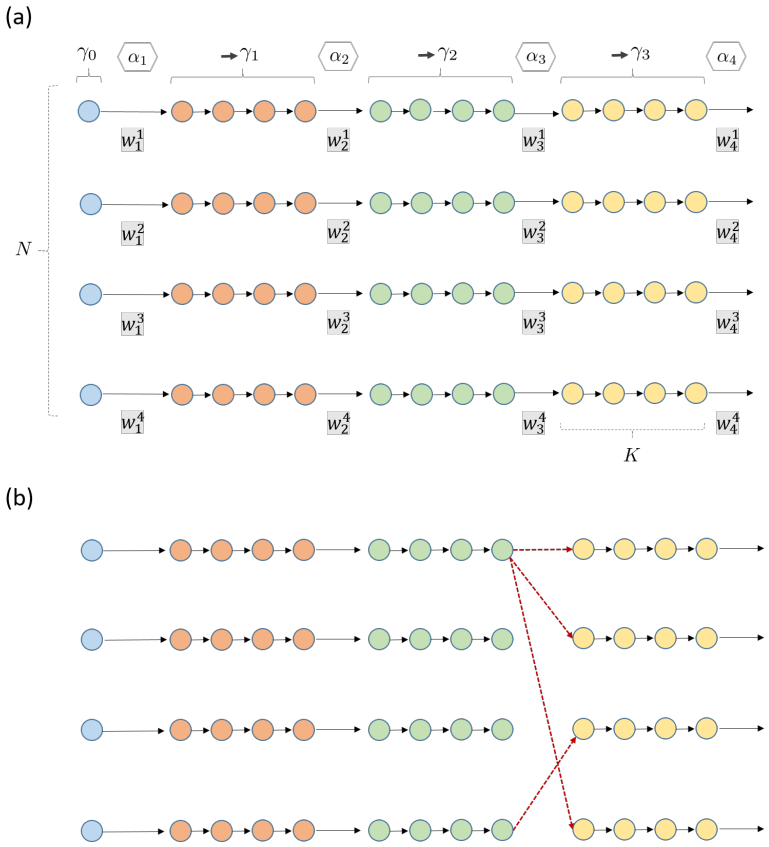

Figure 2.1: (a) Schematic representation of annealed importance sampling (AIS) using $N = 4$ particles evolving in parallel. Except for the initialization step, each color represents $K = 4$ Markov steps in which the particle system moves from approximating a previous unnormalized power-posterior to a new one. After each $K = 4$ Markov steps, the sampled states are used in an importance sampling step to determine the incremental weights w_t associated with the change in the intermediate posterior PDF. (b) In adaptive sequential Monte Carlo (ASMC), one main difference compared with AIS is that the α-sequence determining the intermediate posterior distributions is no longer fixed but determined adaptively. Furthermore, resampling occurs when the variance of the weights are too large. Such a resampling step is here visualized with dashed red lines.

By using $\gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})$ as an importance distribution for $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$, we define the unnormalized incremental weights w_t for particle i at state $\boldsymbol{\theta}^i_{t-1}$ as:

$$w^i_t = \frac{\gamma_t(\boldsymbol{\theta}^i_{t-1}|\mathbf{y})}{\gamma_{t-1}(\boldsymbol{\theta}^i_{t-1}|\mathbf{y})}. \tag{2.8}$$

Except for the initialization step, the corresponding importance distributions $\gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})$ are approximated by updating N particles using K Markov steps targeting $\gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})$ starting at a previous estimation of $\gamma_{t-2}(\boldsymbol{\theta}_{t-2}|\mathbf{y})$. Without these Markov steps, the AIS algorithm would reduce to BFMC. This process is schematized in Figure 2.1a for $N = 4$ and $K = 4$.

It is customary to work with normalized weights defined as:

$$W^i_t = \frac{W^i_{t-1} {w_t}^i}{\sum_{j=1}^N W^j_{t-1} w^j_t}, \tag{2.9}$$

where W_{t-1} are the previously defined normalized weights, that is, $\sum_{i=1}^N W^i_{t-1} = 1$. The final normalized weights W^i_T determine the relative probabilities of each of the final N states, thereby, approximating the posterior distribution through a particle approximation.

Evidence estimation

One major advantage of AIS and ASMC in the context of Bayesian model selection is that the evidence is readily obtained. The ratio of the normalizing constants of two consecutive intermediate distributions $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$ and $\gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})$ is:

$$\frac{Z_t}{Z_{t-1}} = \frac{\int \gamma_t(\boldsymbol{\theta}_t|\mathbf{y}) d\boldsymbol{\theta}_t}{\int \gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y}) d\boldsymbol{\theta}_{t-1}}, t \tag{2.10}$$

and it can be approximated as (*Del Moral et al.*, 2006):

$$\frac{Z_t}{Z_{t-1}} \approx \sum_{i=1}^N W^i_{t-1} w^i_t. \tag{2.11}$$

The posterior PDF of interest is the last distribution of the sequence ($\alpha_{t=T} = 1$), therefore, its normalizing constant is the evidence, $Z_T = \pi(\mathbf{y})$. Since the normalizing constant of the prior PDF, Z_0, is equal to one, the evidence can be estimated as the product of the intermediate ratios:

$$\pi(\mathbf{y}) = Z_T = \frac{Z_T}{Z_0} = \prod_{t=1}^{T} \frac{Z_t}{Z_{t-1}} \approx \prod_{t=1}^{T} \sum_{i=1}^{N} W_{t-1}^{i} {w_t}^{i}. \tag{2.12}$$

Adaptive sequence of intermediate distributions

Zhou et al. (2016) introduce several adaptations to AIS leading to the more robust ASMC algorithm that requires much less tuning. The choice of the annealing schedule in equation 2.6 has a strong impact on performance and it is generally difficult to assign a proper α-sequence in advance. *Zhou et al.* (2016) solve this by introducing an adaptive procedure relying on the conditional effective sample size (CESS):

$$CESS = N \frac{(\sum_{i=1}^{N} W_{t-1}^{i} {w_t}^{i})^2}{\sum_{j=1}^{N} W_{t-1}^{j} ({w_t}^{j})^2}. \tag{2.13}$$

The CESS measures the quality of the current intermediate distribution as an importance distribution to calculate expectations of the following one. To define the next distribution in the sequence (Figure 2.1a), a binary search is performed for the α-increment for which the CESS is the closest to a pre-defined target value. The larger this target value is, the better the approximation, but the slower is the algorithm as the L number of intermediate distributions grows.

Resampling

The variance of the importance weights provides an indicator of the quality of the importance estimator. The importance weights invariably diverge over time leading to high variances, for example, because of poor convergence of some particles. To circumvent this, SMC methods rely on resampling (*Del Moral et al.*, 2006; *Doucet and Johansen*, 2011). Resampling consists of reinitializing the states of each particle by replicating them according to a probability that is proportional to their current normalized weights. After resampling, the new states are assigned equal weights of $1/N$. Figure 2.1b illustrates a resampling step. The purpose of this operation is to limit the variance of the weights by excluding states with lower weights and replicating those with higher weights. Since high-dimensional posterior distributions are estimated using N particles only, it is essential that all samples contribute meaningfully to this approximation by avoiding regions of very low probability. We rely herein on systematic resampling, which is easy to implement and performs well with respect to alternative resampling schemes (*Doucet and Johansen*, 2011). The resampling step impacts the variance of estimates (*Douc and Cappé*, 2005) and it is often beneficial to only perform resampling occasionally. To decide when to apply resampling, we follow standard practice by relying on a quantity that considers the history of the weight variance evolution, namely the effective sample size (ESS) (*Kong et al.*, 1994):

$$ESS_t = \frac{(\sum_{i=1}^{N} W_{t-1}^i w_t^i)^2}{\sum_{j=1}^{N} (W_{t-1}^j)^2 (w_t^j)^2}. \tag{2.14}$$

The ESS can be interpreted as reflecting the number of effective samples in the particle approximation and resampling is applied when the ESS is lower than a pre-defined threshold.

Evidence uncertainty estimation

The most reliable approach to assess uncertainty on evidence estimates is to perform multiple ASMC runs and calculate the resulting variance of the estimates. This is the approach used by *Zhou et al.* (2016) when introducing ASMC. Even if such Monte Carlo replication is easily parallelized, it implies a significant computational overhead as the total computational effort grows linearly with the number of replicates. In recent years, progress has been made in obtaining evidence variance estimates from single SMC runs. The first consistent estimator was proposed by *Chan and Lai* (2013) and a refined estimator was later introduced by *Lee and Whiteley* (2018). We consider a modified form of this latter estimator in *Doucet and Lee* (2018) that we adopted to account for occasional resampling. The resulting expression should be interpreted as a relative variance contribution of the evidence estimate contribution since the last resampling time:

$$\frac{V_t^N}{\left(\eta_t^N\right)^2} = \frac{1}{\left(\eta_t^N\right)^2} \left(\frac{N}{N-1}\right)^n \frac{1}{N(N-1)} \sum_{i=1}^{N} \left[\sum_{j:E_t^j=i} (N W_{t-1}^j w_t^j - \eta_t^N) \right]^2, \tag{2.15}$$

where $\eta_t^N = \sum_{i=1}^{N} N W_{t-1}^i w_t^i$ and n is the cumulative number of resampling steps that has been performed until t. The index E_t^j is the so-called Eve index of particle j at time t, which traces the origin of the particles. If no resampling is done, the Eve indices stay constant and are equal to $1:N$. After resampling, the states of the particles are reorganized and the Eve indices change, denoting the original particle that moved to that position. A graphical illustration of this process is given by *Lee and Whiteley* (2018). The number of remaining unique Eve indices along the run can be interpreted as a conservative estimate of the number of independent particles.

We compute the estimator in equation 2.15 before each resampling step and at the last step of the ASMC algorithm. We then sum the resulting contributions:

$$\sigma_r = \sqrt{\sum_{h=0}^{R} \frac{V_h^N}{\left(\eta_h^N\right)^2}}, \tag{2.16}$$

where R is the total number of resampling times. This equation is valid under the assumption that the individual contributions in the sum are independent (*Brown and Neal*, 1991). Hence, we assume here that the particles decorrelate from each other between resampling steps.

Markov proposals and acceptance criteria

We implemented ASMC within the popular differential evolution adaptive Metropolis ZS (DREAM$_{(ZS)}$) algorithm (*Laloy and Vrugt*, 2012). In this MCMC algorithm, model proposal updates with respect to the present state are drawn proportionally to random differences of past states, thus, helping to better explore the target distribution by automatically determining the scale and direction of the model proposals. If we consider $\mathbf{J}$ as a $m \times d$ dimensional matrix that contains m past states of the chains, where d is the number of parameters, the jump vector for the i-th chain is given by (*Vrugt*, 2016):

$$d\boldsymbol{\theta}_A^i = \zeta_{d^*} + (1_{d^*} + \lambda_{d*})\psi(\delta, d^*) \sum_{j=1}^{\delta} (\mathbf{J}_A^{a_j} - \mathbf{J}_A^{b_j}). \tag{2.17}$$

If the current state is $\boldsymbol{\theta}^i$, then the candidate point for particle i is $\boldsymbol{\theta}_{prop}^i = \boldsymbol{\theta}^i + d\boldsymbol{\theta}^i$. The number of pairs used to generate the jump is given by δ, and $\mathbf{a}$ and $\mathbf{b}$ are vectors of integers drawn without replacement from $\{1, .., m\}$. The parameters ζ and λ are sampled independently from pre-defined uniform and normal distributions, respectively. This algorithm implements subspace sampling, which implies that only a random subset A of d^*-dimensions from the original parameter space is updated at each proposal step. The difference between past states is multiplied by a fixed proposal scale referred to as jump rate $\psi(\delta, d^*) = \frac{2.38}{\sqrt{2\delta d^*}}\epsilon$, where ϵ is an user-defined factor that we introduce to further control the size of the jumps. In contrast to MCMC, ASMC allows straightforward adaptation of the $\epsilon-$factor on-the-go without violating detailed balance condition. This tuning of ϵ is achieved by using the acceptance rate (AR) of the last K Markov steps to target an acceptance rate above AR_{min}. To implement this, ϵ is initialized to a comparatively large value and a percentage decrease of its value f is made when the acceptance rate falls below AR_{min}. For comparison purposes, we also consider standard model proposals given by uncorrelated Gaussian draws centered on the previous state. For this case, the jump vector for the i-th chain is given by:

$$d\boldsymbol{\theta}_A^i \overset{i.i.d.}{\sim} \mathcal{N}_A(0, \epsilon^2). \tag{2.18}$$

Our considered model proposals are symmetric and the prior PDF is uniform. Consequently, with proper boundary handling, the proposed moves are accepted according to the likelihood ratio (*Mosegaard and Tarantola*, 1995). The probability to accept each candidate model during the K Markov steps used to approximate $\gamma_t(\boldsymbol{\theta}_t|y)$ is:

Algorithm 1: ASCM algorithm adopted from *Zhou et al.* (2016); their algorithm 4.

Assignment of user-defined variables:
  Define number of particles (N), optimal CESS ($CESS_{op}$), ESS threshold (ESS^*),
  number of MCMC iterations at each intermediate distribution (K), minimal acceptance rate (AR_{min}),
  initial proposal scale factor (ϵ) and its percentage decrease (f).
Initialization: Set $t = 0$
  Set $\alpha = 0$
  Sample $\boldsymbol{\theta}_0$ from the prior $\pi(\boldsymbol{\theta}_t|M_k)$ N times
  Set the N-dimensional vector of normalized weights $\mathbf{W}_0 = [\frac{1}{N}; \frac{1}{N}; ...; \frac{1}{N}]$
  Set evidence $\pi(\mathbf{y}|M_k) = 1$
Iteration : Set $t = t + 1$
  Search for incremental distribution
    Do binary search for the increment $\Delta\alpha$ that gives the CESS (equation 2.13) that is the closest to $CESS_{op}$.
    Update $\alpha = min(1, \alpha + \Delta\alpha)$ and define the intermediate distribution $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y}) = \pi(\boldsymbol{\theta}_t|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}_t)^{\alpha}$.
    Compute the weight increments w_t^i (equation 2.8), update and save the normalized weights W_t^i (equation 2.9)
    and the evidence $\pi(\mathbf{y}|M_k) = \pi(\mathbf{y}|M_k) \sum_{i=1}^{N} W_{t-1}^i w_t^i$ (equation 2.12).
  Resampling
    Calculate ESS (equation 2.14), if $ESS < ESS^*$ do resampling: re-organize $\boldsymbol{\theta}_t$ states and update $\mathbf{W}_t = [\frac{1}{N}; \frac{1}{N}; ...; \frac{1}{N}]$
  Do K MCMC iterations for each of the N particles (chains):
    Propose moves $\boldsymbol{\theta}_{prop}$ (equation 2.17 and 2.18) and accept or reject based on acceptance criterion (equation 2.19)
    using $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$.
    Save the N $\boldsymbol{\theta}$ and their likelihoods.
    Set last state as $\boldsymbol{\theta}_{t+1}$
  Tune proposal scale
    If acceptance rate $AR < AR_{min}$ then decrease proposal scale factor: $\epsilon = \epsilon * (1 - \frac{f}{100})$
Repeat until α=1

$$P = min\left\{1, \frac{p(\mathbf{y}|\boldsymbol{\theta}_{prop})^{\alpha_t}}{p(\mathbf{y}|\boldsymbol{\theta})^{\alpha_t}}\right\}. \tag{2.19}$$

2.2.3 Full ASMC algorithm

The full algorithm is given in Algorithm 1, for which the total number of iterations per considered particle (chain) is equal to L (number of intermediate distributions) $\times$ K (MCMC steps per distribution).

This algorithm has several important strengths: (i) it requires a rather small number of user-defined parameters; (ii) the posterior PDF and the evidence are estimated; (iii) the variance of the weights are used to assess accuracy, (iv) the adaptation of classical MCMC algorithms into ASMC is straightforward, and (v) the acceptance rate is controlled throughout the inversion.

2.2.4 The Laplace-Metropolis method

MCMC algorithms provide an approximation of the posterior distribution, however, they need to be combined with an additional estimation procedure to provide evidence estimates. For later comparison purposes with ASMC, we mention here the Laplace-Metropolis estimator (*Lewis and Raftery*, 1997), a mathematical approximation of the evidence using a Taylor expansion around the maximum a posteriori (MAP) estimate. Assuming that the posterior PDF is well approximated by a normal distribution, the resulting evidence estimate is:

$$\pi(\mathbf{y}|M_k) = (2\pi)^{\frac{d}{2}} |\mathbf{H}(\boldsymbol{\theta}^*)|^{\frac{1}{2}} \pi(\boldsymbol{\theta}^*|M_k)\, p(\mathbf{y}|\boldsymbol{\theta}^*, M_k), \tag{2.20}$$

where $\boldsymbol{\theta}^*$ is the MAP estimate, d is the number of parameters and $|\mathbf{H}(\boldsymbol{\theta}^*)|$ is the determinant of minus the inverse Hessian matrix evaluated at the MAP, which is approximated from the MCMC-based samples from the posterior.

2.2.5 From implicit to prescribed geostatistical priors

Multiple-point statistics (MPS) (*Mariethoz and Caers*, 2014) is a sub-field of geostatistics aiming at producing conditional geostatistical model realizations of high geological realism, thereby, capturing more meaningful connectivity patterns than those offered, for instance, by classical multivariate Gaussian priors (*Renard and Allard*, 2013). MPS algorithms produce model realizations that are in agreement with the spatial patterns found in a so-called training image (TI). A TI is a gridded representation of the targeted spatial field obtained from geological information such as outcrops or process-based simulation methods (*Koltermann and Gorelick*, 1996). Performing inversion (*Mariethoz et al.*, 2010a; *Hansen et al.*, 2012; *Linde et al.*, 2015) and model selection (*Brunetti et al.*, 2019) based on one or more TIs commonly requires inversion algorithms that work with so-called implicit priors. That is, the MPS algorithm provides model realizations that are drawn proportionally to the prior, but the prior density of a given realization is unknown. Two main issues arise with this approach: (1) the generation of conditional prior realizations may be computationally expensive in MCMC settings when a large number of model proposals are needed, and (2) the implicit prior model precludes the calculation of prior probability densities as needed in many state-of-the-art inversion and model selection methods.

Deep learning (*LeCun et al.*, 2015) applied to geoscientific problems has been growing rapidly in recent years (*Bergen et al.*, 2019; *Karpatne et al.*, 2018). In particular, deep generative neural networks offer an attractive approach to build an explicit prior PDF from training images (*Laloy et al.*, 2017, 2018; *Mosser et al.*, 2017, 2020), that is, a prior for which the prior density of any realization is easily calculated. This is achieved by learning a non-linear transform between a low-dimensional latent space with a prescribed prior (typically an uncorrelated standard normal or bounded uniform prior) and the image space (on which the forward simulations are performed). To do this, the neural network is trained repeatedly with pieces of a large TI or MPS realizations. Such tailor-made model parametrizations achieve significant dimensionality reduction by leveraging spatial patterns in the TI. Inversion is

then performed on the latent space and the resulting posterior is mapped, using the trained transform, into a posterior on the original image space (a so-called push-forward operation). We rely on a spatial generative adversarial neural network (SGAN) (*Jetchev et al.*, 2016), where each dimension of the latent space influences a given region of the generated image space. The network's weights are learned by adversarial training (*Goodfellow et al.*, 2014). The latter consists of a competition between a so-called discriminator and a generator: the discriminator aims to distinguish fake (i.e., realizations by the generator) and real (i.e., training samples) images, while the generator tries to fool it by generating realizations similar to the training samples. This is mathematically translated in a minimization-maximization problem (see the book by *Goodfellow et al.* (2016), for details). The main computational effort is related to training and once trained, the computational cost to draw model proposals in the latent space and to map them into the image space (for further forward computations) is very low. The motivation of evaluating ASMC using a deep-learning based parameterization is two-fold: (1) the SGAN parameterization implies strong non-linearity which makes it difficult for MCMC algorithms to converge when performing inversion on the SGAN latent space (*Laloy et al.*, 2018), thus providing challenging test examples for which the added value of ASMC for posterior inference can be demonstrated and (2) to build on recent work (*Brunetti et al.*, 2019) on MPS-based Bayesian model selection to highlight the value of prescribed priors when performing model selection among MPS-based prior models.

2.3 Results

2.3.1 Test examples

Two conceptual 2-D models represented by TIs were used to assess ASMC for inversion and model selection purposes. These TIs are used to train SGANs that generate realizations honoring the multiple-point statistics of the TIs (*Laloy et al.*, 2018). The first conceptual model (Figure 2.2a) is represented by a binary channelized training image (CM1) (*Zahner et al.*, 2016) and the second one (Figure 2.2b) is represented by a tri-categorical training image characterizing braided-river aquifer deposits (CM2) (*Pirot et al.*, 2015). The SGAN generators are assigned uniform priors on the latent space: the CM1-realizations and the more complex CM2-realizations have 15 and 45 latent variables, respectively. All realizations correspond to an image dimension of 129×65 cells that is cropped to 125×60, with a discretization of 0.1 m $\times$ 0.1 m (Figure 2.3).

Our synthetic data correspond to simulated crosshole ground-penetrating radar (GPR) first-arrival travel times with a geometry consisting of two boreholes that are 5.8 m apart. A total of 24 sources and 24 receivers are placed equidistantly every 0.5 meters in depth. First-arrival times were calculated using the *time2d* algorithm by *Podvin and Lecomte* (1991). Following common practice, the data were filtered according to a maximum angle between sources and receivers of 45 degrees (*Peterson*, 2001), resulting in 444 travel times. In order to assign velocities to each facies, the corresponding dielectric constants were approximated using the complex refractive index method (CRIM) (*Roth et al.*, 1990). Representative porosities for CM2 were taken from *Pirot et al.* (2019) and adjusted to CM1 to have the same mean

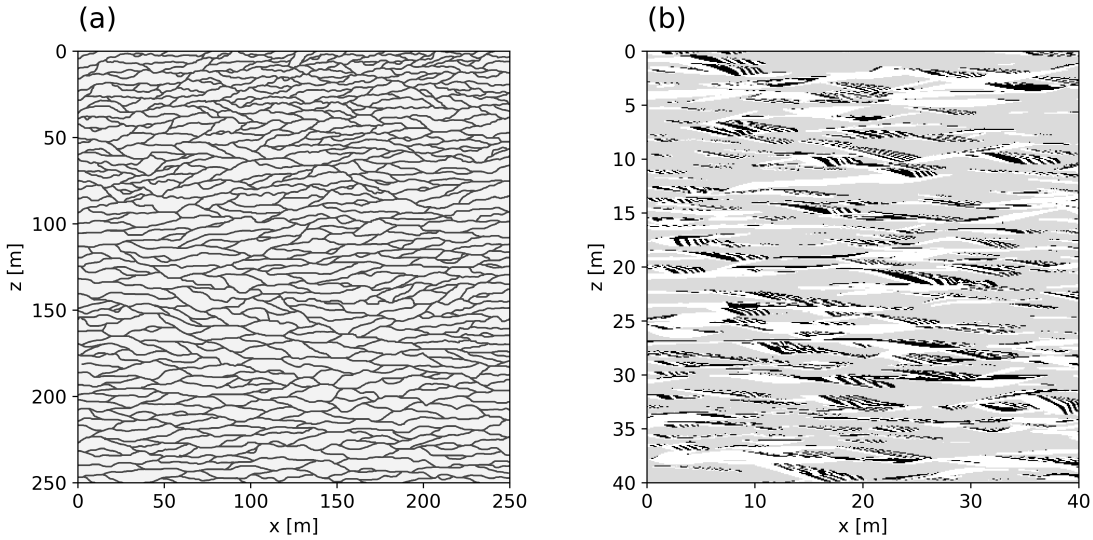

Figure 2.2: Training images: (a) 2500 × 2500 binary channelized training image (CM1) (*Zahner et al.*, 2016) and (b) 400 × 400 tri-categorical training image representing a braided aquifer (CM2) (*Pirot et al.*, 2015). The discretization of the cells is 0.1 m × 0.1 m.

and variance. The two reference models used to produce the synthetic data are shown in Figure 2.3. They were obtained as a randomly chosen realization from the respective SGAN generators. Uncorrelated Gaussian random noise with standard deviation $\sigma = 1$ ns was added to the resulting travel times simulated from these models.

2.3.2 ASMC performance

We first present the parameter settings and the performance of the ASMC algorithm (section 2.2.3) using DREAM$_{(ZS)}$ proposals (ASMC-DREAM) with $N = 40$ particles. To tune the proposal scale, we apply a 20% decrease ($f = 0.2$) with $AR_{min} = 0.25$. The starting large proposal scale ϵ is gradually decreased as the annealing progresses (i.e., the inverse temperature α increases towards 1). We implemented adaptive selection of the α-sequence, using a binary search defined on a range of α-increments from 10^{-5} to 10^{-2}, to find the increments with the $CESS$ that is the closest to the target $CESS_{op}$. The $CESS_{op}/N$ ratio is in practice chosen close to 1. The closer it is to 1, the higher the number of intermediate distributions and the larger is the quality of estimates. Resampling is applied whenever ESS/N falls below 0.5. Table 2.1 contains the user-defined parameters and the resulting sequence lengths. The total number of forward simulations of each ASMC run is $N \times K \times L$.

Figures 2.4a-b show the evolution of the likelihood raised to the power of the corresponding α in the natural log-scale for CM1 and CM2, respectively. This type of plotting is consistent with the target distribution $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$ at each step (equation 2.6). The black dashed line indicates the target log-likelihood calculated with the random noise realization used to noise-contaminate the forward response of the reference model, raised to the power of the corresponding α. Figures 2.4c-d present correspondingly the acceptance rate evolution. As α grows, the acceptance rate for a given jump rate decreases as the targeted posterior distribution gives

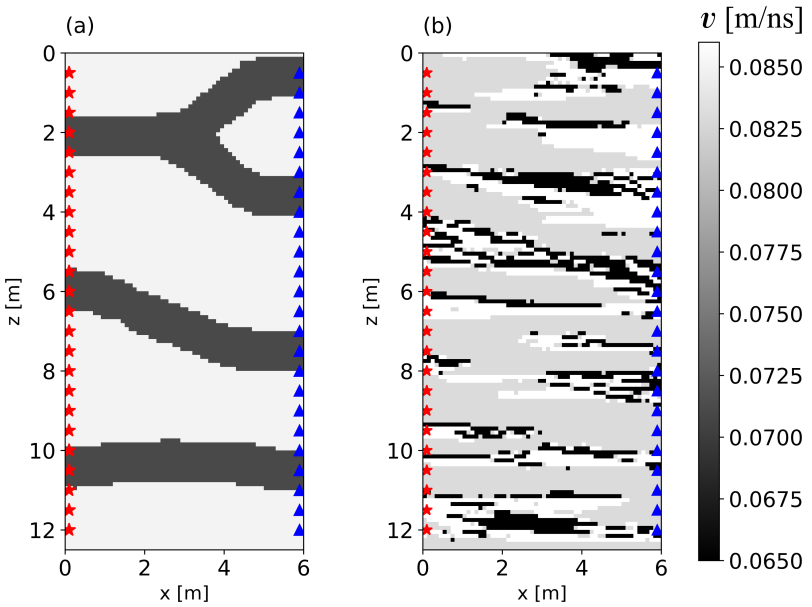

Figure 2.3: Reference models with associated velocities. (a) CM1: channel velocity v=0.071 m/ns and matrix velocity v=0.085 m/ns. (b) From *Pirot et al.* (2015) CM2: gray gravel (gray) v=0.083 m/ns, open framework (black) v=0.065 m/ns and bimodal (white) v=0.086 m/ns. Red stars and blue triangles represent GPR sources and receivers, respectively.

larger weights to the likelihood. When the acceptance rate falls below $AR_{min} = 0.25$, the proposal scale is reduced causing a small increase, after which the acceptance rate starts decreasing again until another reduction of the proposal scale is required, thereby, keeping the acceptance rate in a range between 25% and 40%. Figures 2.4e-f show the optimized sequence of α-values defining the intermediate posterior distributions, obtained through a binary search of the α-increments. In Figures 2.4g-h, the logarithm of the normalized weight of each particle is plotted against the α-index. Finally, Figures 2.4i-j shows the evolution of the natural logarithm of the evidence vs. α.

To ensure convergence with the more complex test case CM2, we had to choose a higher $CESS_{op}$ and K than for CM1, which resulted in an approximately 4.7 times longer run. Despite these adaptations, more resampling steps were needed compared to CM1 (see Table 2.1), which reinforces the impression that it is a more challenging scenario. The increasing complexity of CM2 is also indicated by the fact that the intermediate target distributions are well-approximated for CM1 (Figure 2.4a) for which the sampled likelihoods fall close to the dashed line, while this is less the case for CM2 (Figure 2.4b). However, both test cases reached the target log-likelihood and the resampling fulfills its role of limiting the variance of the weights.

Algorithm 1 is applicable to other model proposals than DREAM$_{(ZS)}$. This is demonstrated using standard (vanilla) MCMC model proposals based on uncorrelated random Gaussian perturbations (ASMC-Gauss). In this case, the algorithm starts with a high standard deviation of the centered Gaussian model proposal and it is subsequently decreased when the acceptance rate falls below 25%. The user-defined parameters were chosen to be the same as

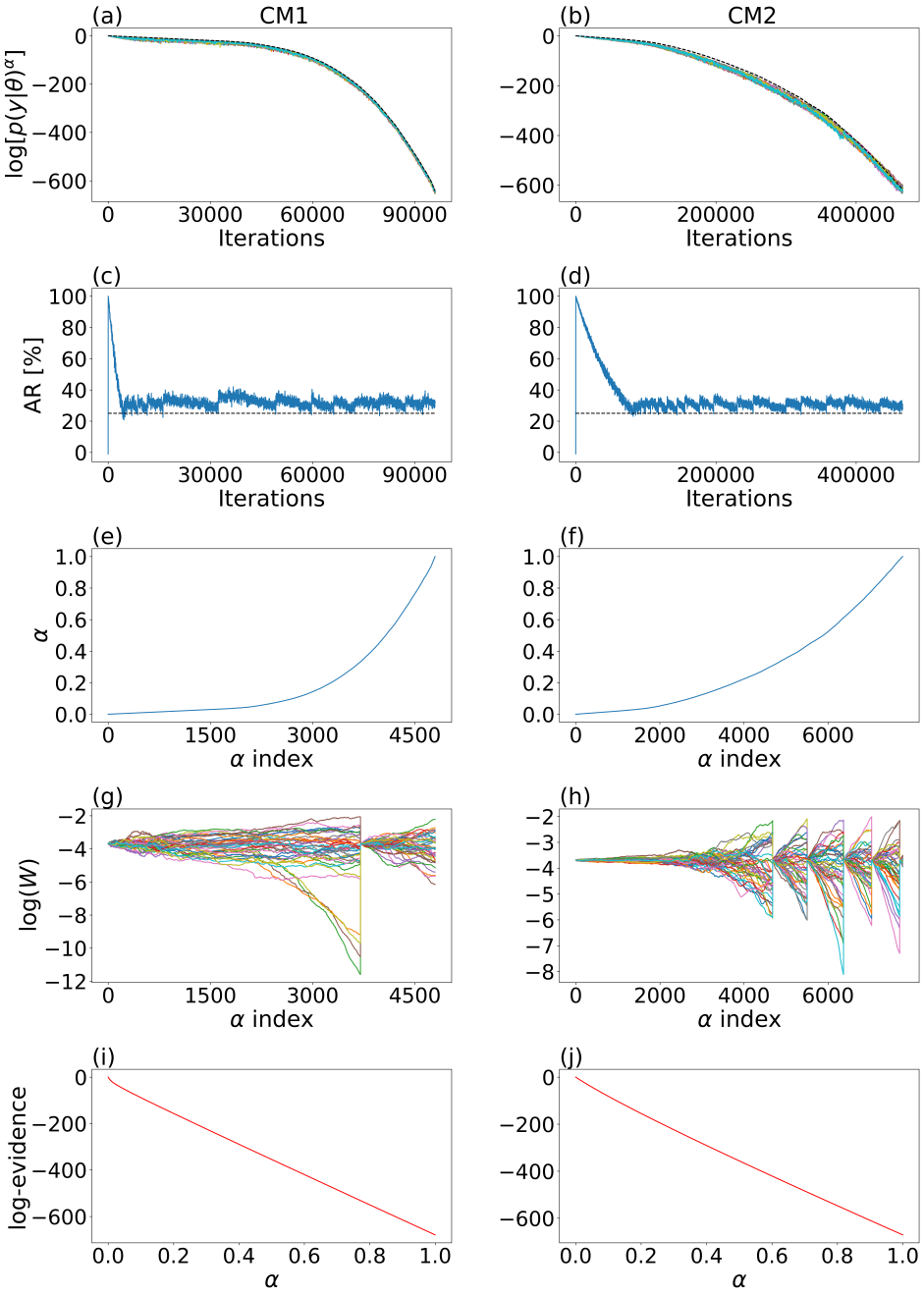

Figure 2.4: Results of ASMC with DREAM$_{(ZS)}$ model proposals (ASMC-DREAM) for conceptual models CM1 (left column) and CM2 (right column): (a) and (b) natural logarithm of the likelihood to the power of α vs. iterations per particle. Each color represents a different particle and the black dashed line indicates the logarithm of the likelihood to the power of α calculated using the random noise realization used to noise-contaminate the forward-simulated true model; (c) and (d) acceptance rate vs. iterations per particle, the dashed line indicates a 25% threshold; (e) and (f) α-sequence vs. α index; (g) and (h) natural log-normalized weights vs. α index where each color represents a different particle; (i) and (j) natural log-evidence evolution vs. α.

Table 2.1: ASMC user-defined parameters and resulting sequence length.

	ASMC-DREAM CM1	ASMC-DREAM CM2	ASMC-Gauss CM1	ASMC-Gauss CM2
Particles (N)	40	40	40	40
$CESS_{op}/N$	0.999993	0.999996	0.999993	0.999996
ESS^*/N	0.5	0.5	0.5	0.5
AR_{min}	25%	25%	25%	25%
K iterations	20	60	20	60
L intermediate distributions	4798	7775	4871	7673
Iterations per particle	95960	466500	97420	460380
Resampling times	1	5	2	3
Total numerical demand [$\times 10^5$]	38.384	186.600	38.968	184.152

for the ASMC-DREAM tests detailed in Table 2.1, leading to a similar sequence length as for ASMC-DREAM. The corresponding results are shown in Figure 2.5. For CM1, ASMC-Gauss needed one more resampling time (Fig. 2.5c) compared to ASMC-DREAM due to a faster increase in the variance of the weights. Otherwise, the performance of ASMC-DREAM (Figure 2.4) and ASMC-Gauss (Figure 2.5) are very similar.

2.3.3 MCMC performance

For comparative purposes, we also perform MCMC inversions (no ASMC) using 40 chains and a similar number of forward simulations. Again, we consider two tests: one using DREAM$_{(ZS)}$ (MCMC-DREAM) and one with random Gaussian perturbations (MCMC-Gauss). Extensive manual tuning of the inversion parameters was needed to achieve satisfactory results. Figure 2.6 shows the results obtained for conceptual models CM1 and CM2. The log-likelihood evolution is shown in Figures 2.6a-d and the acceptance rate in Figures 2.6e-h. In order to assess convergence, the potential scale reduction factor $\hat{R}$ is calculated (*Gelman and Rubin*, 1992) and plotted in Figures 2.6i-l, with convergence declared when $\hat{R}$ is below 1.2 for all model parameters.

The only MCMC run reaching convergence is MCMC-DREAM for CM1 at around 10,000 iterations. For this conceptual model, the results obtained with MCMC-Gauss are unsatisfactory with only a few of the chains approaching the target likelihood, while the others are trapped in local minima, thereby, demonstrating a vastly superior performance of MCMC-DREAM compared with MCMC-Gauss. For CM2, none of the MCMC inversions converge within the allotted computational time, as $\hat{R}$ does not fall below 1.2. This is also reflected in the likelihood evolution: the majority of sampled likelihoods remains below the target likelihood along the run. To summarize, we find for a similar computational budget that the ASMC algorithm reaches the target likelihood for both conceptual models and model proposal types, while the MCMC runs only approximate the target likelihood for CM1 using MCMC-DREAM.

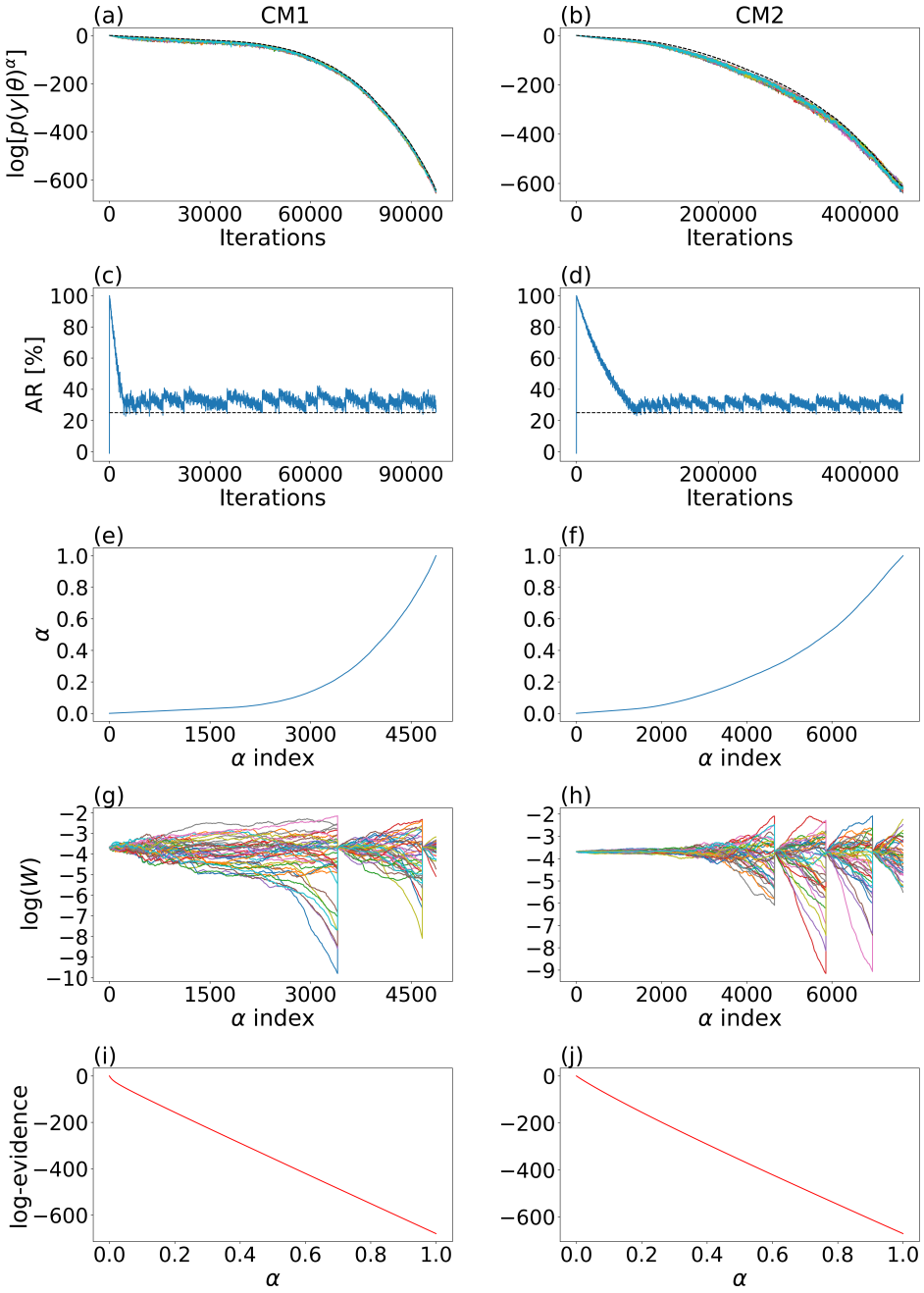

Figure 2.5: Results of ASMC with standard MCMC (ASMC-Gauss) for conceptual models CM1 (left column) and CM2 (right column): (a) and (b) natural logarithm of the likelihood to the power of α vs. iterations per particle. Each color represents a different particle and the black dashed line indicates the logarithm of the likelihood to the power of α calculated using the random noise realization used to noise-contaminate the forward-simulated true model; (c) and (d) acceptance rate vs. iterations per particle, the dashed line indicates a 25% threshold; (e) and (f) α-sequence vs. α index; (g) and (h) natural log-normalized weights vs. α index, each color represents a different particle; (i) and (j) natural log-evidence evolution vs. α.

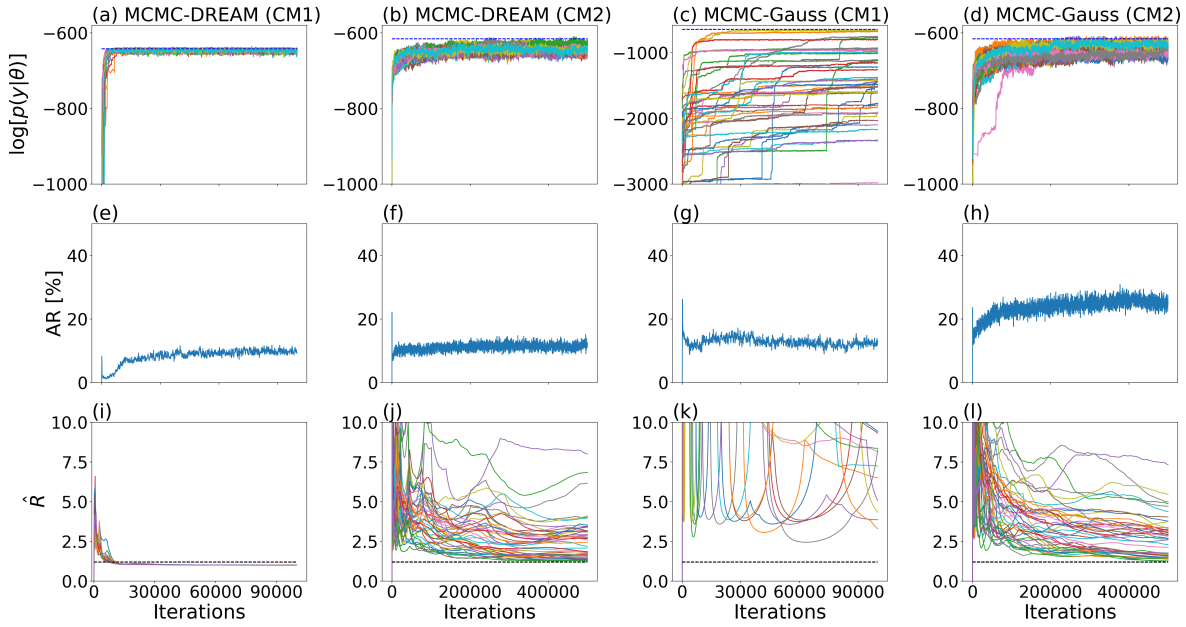

Figure 2.6: MCMC inversion results from DREAM$_{(ZS)}$ (MCMC-DREAM) and standard MCMC with Gaussian model proposals (MCMC-Gauss) for conceptual models CM1 and CM2. (a)-(d) the natural logarithm of the likelihood vs. iterations, where each color represents a different particle and the black dashed line indicates the log-likelihood calculated using the random noise realization, (e)-(h) the acceptance rate evolution, and (i)-(l) the evolution of the potential scale reduction factor $\hat{R}$ with each color representing a different parameter and the black dashed lines indicating the value below which convergence is declared ($\hat{R} = 1.2$).

2.3.4 Posterior distributions

We focus now on the posterior approximations obtained with ASMC-DREAM and MCMC-DREAM. For MCMC-DREAM, the posterior is obtained by first removing the so-called burn-in period, that is, the number of iterations needed to reach the target likelihood, from which it starts to sample from the posterior PDF. The remaining samples contribute equally to the posterior estimations. This is not the case for ASMC, for which the posterior PDF is approximated by the last states and weights of the particles (chains).

For a smoother representation of the posterior PDF approximated by ASMC, we applied kernel density estimation (KDE) (*Scott*, 2015). Figure 2.7 compares the estimated posteriors for CM1. The KDE bandwidth impacts on the level of smoothing, that we chose to kept fixed for the 15 parameter posteriors. Nevertheless, the estimated posteriors are overall very similar, which suggests that ASMC provides a good estimation of the posterior. No comparison is provided for CM2 as the MCMC-DREAM algorithm did not converge, neither in terms of reaching the target likelihood nor in terms of exploration of the posterior PDF.

We now consider the posterior means and variances in the image space by translating the posterior realizations in the latent space using the SGAN generator. For ASMC-DREAM, the mean and standard deviation images correspond to the last states of the chains weighted by their weights. For MCMC-DREAM, the mean and standard deviation images are obtained using the equally weighted states in the second half of the chains. The means and standard deviations for CM1 are very similar for ASMC-DREAM (Figure 2.8b-c) and MCMC-DREAM (Figure 2.8d-e) that both approximate the true model very well (Figure 2.8a). For CM2, we see a much better defined mean model and smaller standard deviations for ASMC-DREAM (Figure 2.8g-h). The poorer approximations by MCMC-DREAM 2.8i-h) is a direct consequence of the fact that this run did not converge. Table 2.2 shows the log-likelihood range for the different inversions. For MCMC-DREAM, the second halves of the chains are considered for the range, while only the last states of the particles are considered for ASMC-DREAM.

2.3.5 Evidence estimation

Even if the theoretical basis of the ASMC method for evidence estimation is well-established (*Zhou et al.*, 2016), we start this section by considering a simple example that allows for comparison with BFMC (see section 2.2.2). We consider CM1 in a high-noise setting using uncorrelated Gaussian random noise with standard deviation $\sigma = 15$ ns. This is certainly an unrealistically high noise level, but it allows us to obtain reliable evidence estimates through BFMC using 2 million prior samples. The resulting log-evidence obtained by BFMC is -1798.92, while the corresponding ASMC-DREAM run using $K = 5$ and $CESS_{op}/N = 0.9999$ (resulting in 1100 iterations per particle) led to a log-evidence estimate of -1798.86, which is practically identical to the BFMC estimate.

After having established that our ASMC implementation provides accurate evidence estimation by comparison with BFMC, we now return to the original low-noise $\sigma = 1$ ns setting. For the test examples considered in the previous sections, the evidence estimates obtained with ASMC-DREAM and ASMC-Gauss given in Table 2.2 (i.e., the last computed values

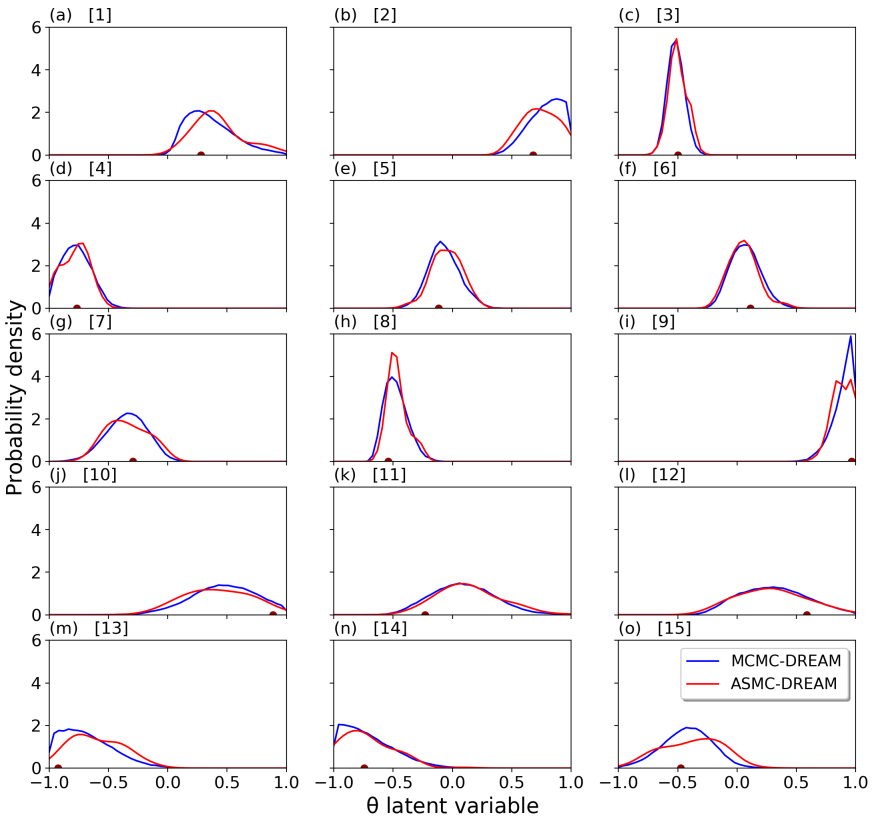

Figure 2.7: Estimated marginal posterior distributions for CM1 using ASMC with DREAM$_{(ZS)}$-proposal (ASMC-DREAM) and regular DREAM$_{(ZS)}$ (MCMC-DREAM) with a comparable number of forward computations. Results are shown for all latent model parameters that have bounded uniform priors between -1 and 1.

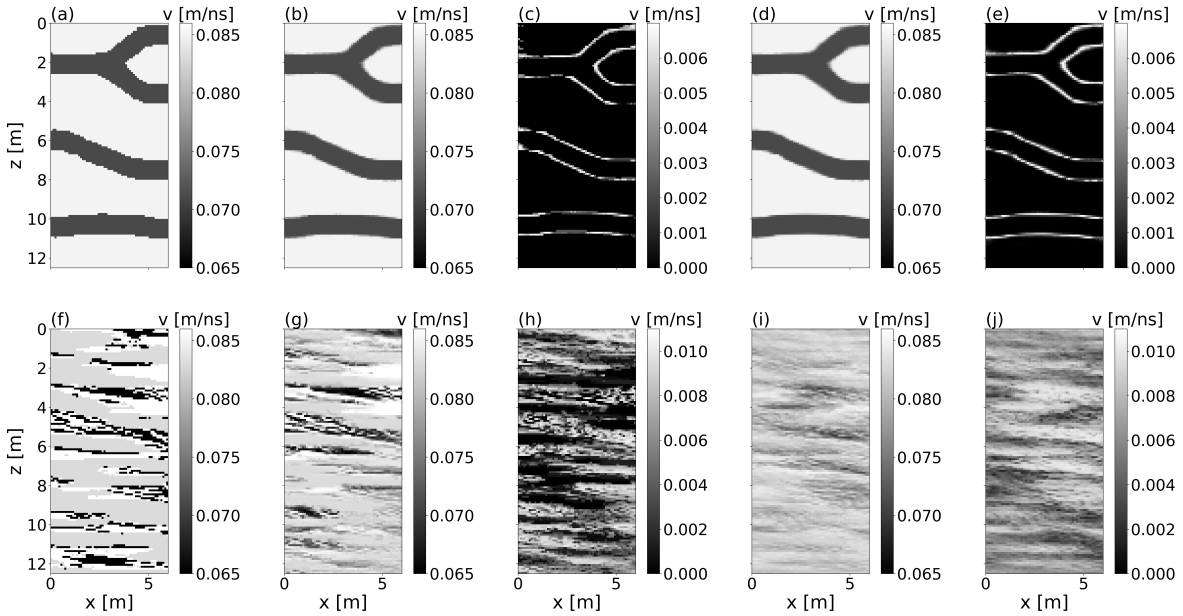

Figure 2.8: Reference model for (a) CM1 and (f) CM2; mean of the weighted final states from ASMC-DREAM for (b) CM1 and (g) CM2; standard deviations of the corresponding weighted final states for (c) CM1 and (h) CM2; mean of the second half of the MCMC chains obtained with MCMC-DREAM for (d) CM1 and (i) CM2 (not converged); corresponding standard deviations for (e) CM1 and (j) CM2 (not converged).

shown in Figures 2.4i-j and 2.5i-j) are very close to each other. For comparison purposes, we also calculate the Laplace-Metropolis evidence estimator (LM) using the MCMC-DREAM inversion results (equation 2.20). This is done for CM1 only as MCMC-DREAM did not converge for CM2. The Laplace-Metropolis estimate (Table 2.2) is only slightly lower than the ASMC-DREAM and ASMC-Gauss estimates. The close agreement between ASMC-DREAM and ASMC-Gauss, and the close agreement considering the simplifying assumptions of the Laplace-Metropolis method, suggest again that the results obtained with ASMC are accurate.

Until now, we have considered that the right conceptual (prior) model was used in the inversions. That is, the noise-contaminated data were generated with a realization of the assumed prior PDF. We now consider how the evidence changes if we make the wrong assumption, that is, use the noise-contaminated data generated from a prior draw of another conceptual model. In Figure 2.9 we display the evidence evolution for two such incorrect scenarios using ASMC-DREAM with combinations of CM1 and CM2 in the data generation and inversion process. The resulting log-evidence estimates (Table 2.2) are many hundreds of times smaller than the estimations obtained by making the right assumption, suggesting in these simple scenarios that the true conceptual model can easily be inferred if it is in the set of considered conceptual models.

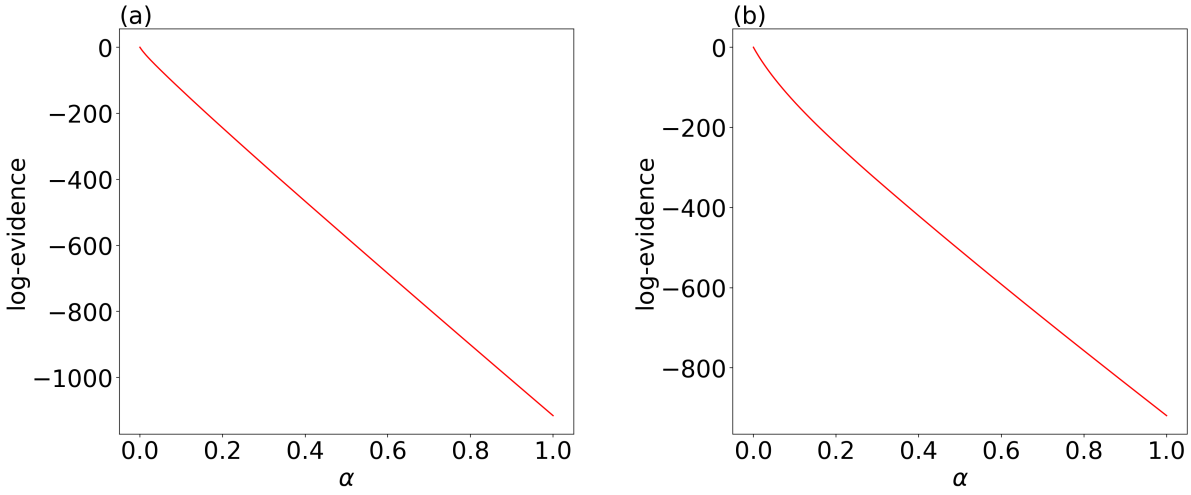

Figure 2.9: ASMC-DREAM evidence evolution with respect to the α-sequence evolution when making incorrect assumptions about the underlying conceptual model: (a) CM1-based prior in the inversion using data generated from a prior realization from CM2, and (b) CM2-based prior in the inversion using data generated from a prior realization from CM1.

Table 2.2: Natural log-likelihood range, natural log-evidence estimation and number of resampling steps for the different inversion cases. The log-likelihoods of the reference models are -642.34 (CM1) and −616.00 (CM2).

	Log-likelihood range	Log-evidence estimation	Resampling times
CM1 inv - CM1 data/ ASMC-DREAM	[-652.03; -641.02]	-679.48	1
CM1 inv - CM1 data/ MCMC-DREAM	[-666.07; -636.71]	-678.39*(LM)*	-
CM1 inv - CM1 data/ ASMC-Gauss	[-654.79; -640.65]	-679.80	2
CM2 inv - CM2 data/ ASMC-DREAM	[-628.60; -603.91]	-671.18	5
CM2 inv - CM2 data/ MCMC-DREAM	[-682.90; -612.23]	-	-
CM2 inv - CM2 data/ ASMC-Gauss	[-638.64; -611.15]	-671.49	3
CM1 inv - CM2 data/ ASMC-DREAM	[-1086.42;-1063.34]	-1115.76	5
CM2 inv - CM1 data/ ASMC-DREAM	[-831.70; -795.19]	-919.17	9

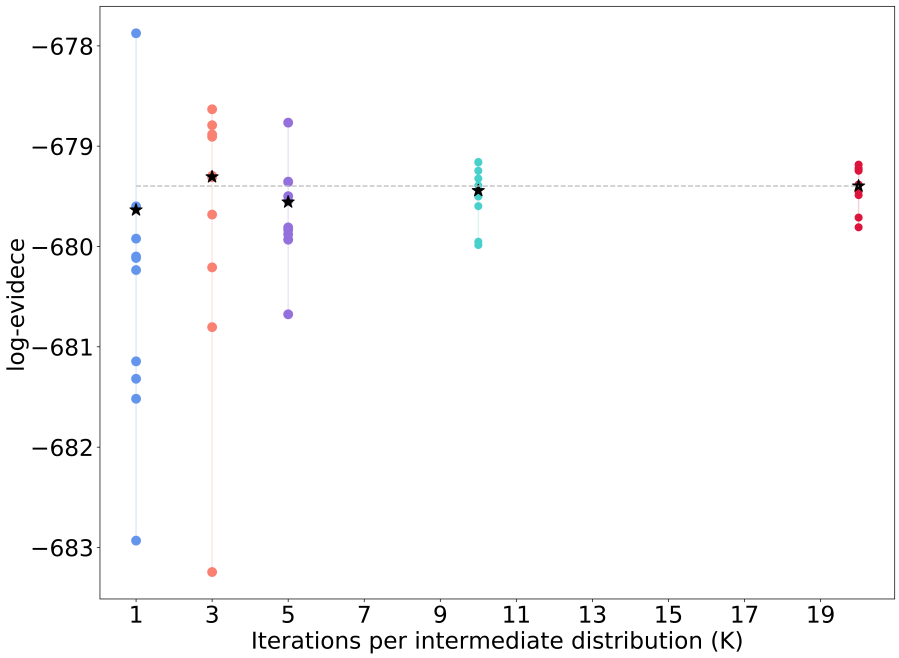

Figure 2.10: Natural log-evidence estimations for ten replications of the ASMC-DREAM algorithm applied to CM1 using $K = 1, 3, 5, 10, 20$ iterations per intermediate distribution, where each colored point denotes a given replication. The gray dashed line represents the mean of the $K = 20$ replications and the black stars the corresponding mean for each K.

2.3.6 Evidence uncertainty quantification

We first assess the uncertainty of the evidence estimations by performing Monte Carlo replication. For the low noise ASMC-DREAM tests shown in section 2.3.2, we performed ten separate runs of ASMC-DREAM for CM1 and five for CM2. We varied K and kept all other parameters fixed. Figure 2.10 shows the corresponding evidence estimations for CM1 and their means in logarithmic units. Table 2.3 shows the relative standard deviation for both conceptual models. For CM1, it decreases almost by a factor of 10 when moving from $K = 1$ to $K = 20$. For this case, even $K = 1$ leads to rather high-quality estimates with a relative standard deviation of 1.72. The decrease is less abrupt for CM2 when increasing $K = 5$ to $K = 60$.

From a computational standpoint, it is beneficial if high-quality uncertainty estimates of the evidences would be obtained from one ASMC run only. Hence, we assess how the predictions of equations 2.15 and 2.16 compare with the estimates based on Monte Carlo replications. For smaller K, resampling compensates for the faster increasing variance of the weights, but this is at the expense of strong correlations between the particles. The impact of resampling on the variance estimation in equation 2.15 is primarily embodied in the sum involving the Eve indices. For smaller K, more resampling is needed and the number of remaining Eve indices are smaller. Figure 2.11 illustrates the evolution of the Eve indices E_t^i for $K = 1$ and $K = 5$ as the CM1 α-sequence progresses. Of the original 40 Eve indices, there are at the end only 3 and 8 Eve indices surviving for $K = 1$ and $K = 5$, respectively. For $K = 20$, there are 15 surviving Eve indeces. The larger the number of surviving Eve indices, the less is the risk of mode collapse in which the ASMC algorithm only explore a small part of the

posterior distribution. This basically implies that the higher-quality estimates are obtained by using larger K or $CESS_{op}$, but this comes at the cost of an increasing number of forward simulations. Table 2.3 shows the relative standard deviation obtained with Monte Carlo replication and the single ASMC run estimates. For CM1, the relative standard deviations calculated with both estimators are similar for $K = 10$ and $K = 20$ suggesting that equations 2.15 and 2.16 may provide high-quality uncertainty estimates for long-enough ASMC runs. For small K, we observe significant underestimation of the relative standard deviations. For $K = 1$, the single ASMC estimation is three times smaller than those obtained by Monte Carlo replication. Why does the single-run ASMC uncertainty estimation work well for large K, but not for small ones? To shed some light on this question, we present in Figure 2.12 the evolution of the difference between the weighted mean of the 40 particles' likelihoods $\hat{p}(\mathbf{y}|\boldsymbol{\theta})$ and the target log-likelihood calculated with the noise realization $p_n(\mathbf{y}|\boldsymbol{\theta})$, both raised to the power of the corresponding α with the differences expressed in logarithmic units, that is, $\Delta log[p(\mathbf{y}|\boldsymbol{\theta})^{\alpha}] = log[\hat{p}(\mathbf{y}|\boldsymbol{\theta})^{\alpha}] - log[p_n(\mathbf{y}|\boldsymbol{\theta})^{\alpha}]$. This difference is shown for the ten replications and for the different K-values considered. In addition, Table 2.3 shows the variance and the root-mean-square error (RMSE) for the last states ($\alpha = 1$) $\Delta log[p(\mathbf{y}|\boldsymbol{\theta}])$ that decrease with increasing K. We observe in Figure 2.12 that when K decreases, the trajectories becomes more separate and show more auto-correlation. At $K = 20$ and $K = 10$ for which the single-ASMC estimates worked well, we observe that the trajectories overlap and cross each other, thereby, suggesting that the information content of one individual ASMC run is not so much different than another. In contrast, for $K = 1$ (Figure 2.12a) the mean trajectories tend to be more separated from each other suggesting that they sample slightly different posteriors. The Monte Carlo replications account for these differences between individual ASMC runs, while this is impossible when considering estimates from a single ASMC run. This suggests then that the single-run evidence estimator should only be trusted when performing a sufficient number of K iterations, thereby, ensuring that the approximations of the intermediate distributions for different ASMC runs are small. In practice, this suggests that it is useful to run at least two ASMC runs and to ensure that the weighted mean-likelihoods of their particles are similar and tend to cross multiple times during the ASMC runs. If this is not the case, our results suggest that the uncertainty estimation of the evidence obtained from one ASMC run is too small.

This finding is also supported by the CM2 estimations in Table 2.3. This is clearly a more challenging conceptual model, where the K used for the ASMC runs was three times higher than for CM1. Even if the single-run uncertainty estimations decrease consistently when increasing K, the values are too low compared to those of Monte Carlo replication. This suggests that K was not large enough to trust the single-run estimator. This is also reflected in the higher variance and the RMSE of the likelihood difference compared to CM1. This suggests that either Monte Carlo replications are needed to obtain an accurate error estimator or K should be increased to improve the reliability of the single-run estimator.

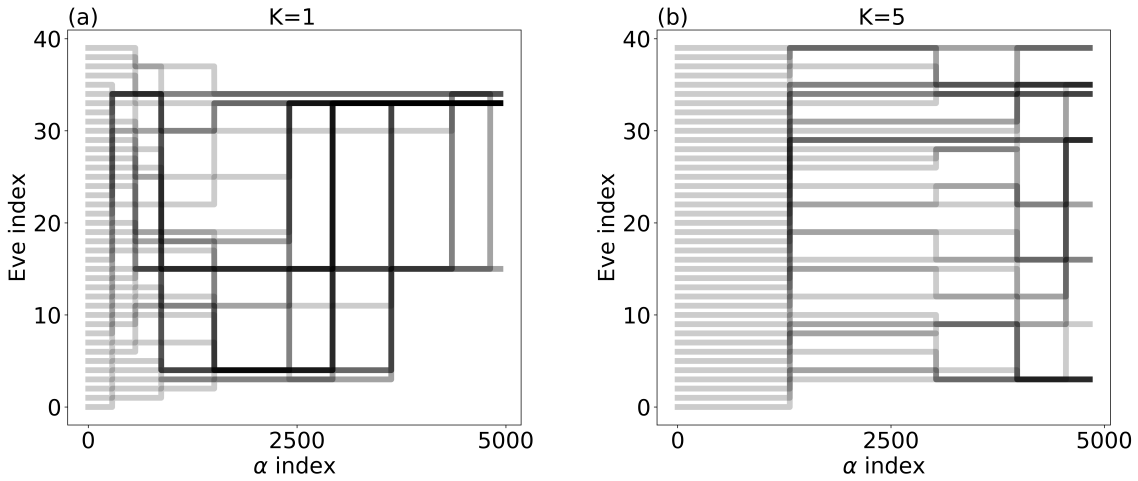

Figure 2.11: Eve index evolution vs. α-sequence evolution for (a) $K = 1$ and (b) $K = 5$. The increasing opacity indicates superposition, that is, replication of specific Eve indices for different particles.

Table 2.3: Relative standard deviation of evidence estimations obtained with ASMC-DREAM using different K iterations per intermediate distribution. Results are shown for estimates based on a single run (equations 2.16 and 2.15) and by ten replications for CM1 and five replications for CM2 of the ASMC algorithm. Variance and root-mean-square error (RMSE) of the difference between the average log-likelihoods and the target (noise) log-likelihood are shown for the replications.

K	σ_r [single run]	σ_r [replications]	$\sigma^2(\Delta log[p(y\|\boldsymbol{\theta})])$	$RMSE(\Delta log[p(y\|\boldsymbol{\theta})])$
		CM1		
1	0.62	1.72	1.70	2.99
3	0.42	0.66	1.42	1.91
5	0.35	0.50	0.62	1.84
10	0.29	0.27	0.67	1.14
20	0.21	0.20	0.69	1.47
		CM2		
5	0.47	1.92	8.45	43.34
10	0.40	1.56	4.16	21.59
20	0.38	1.02	3.66	13.89
40	0.36	1.52	5.06	7.70
60	0.33	1.22	6.36	2.46

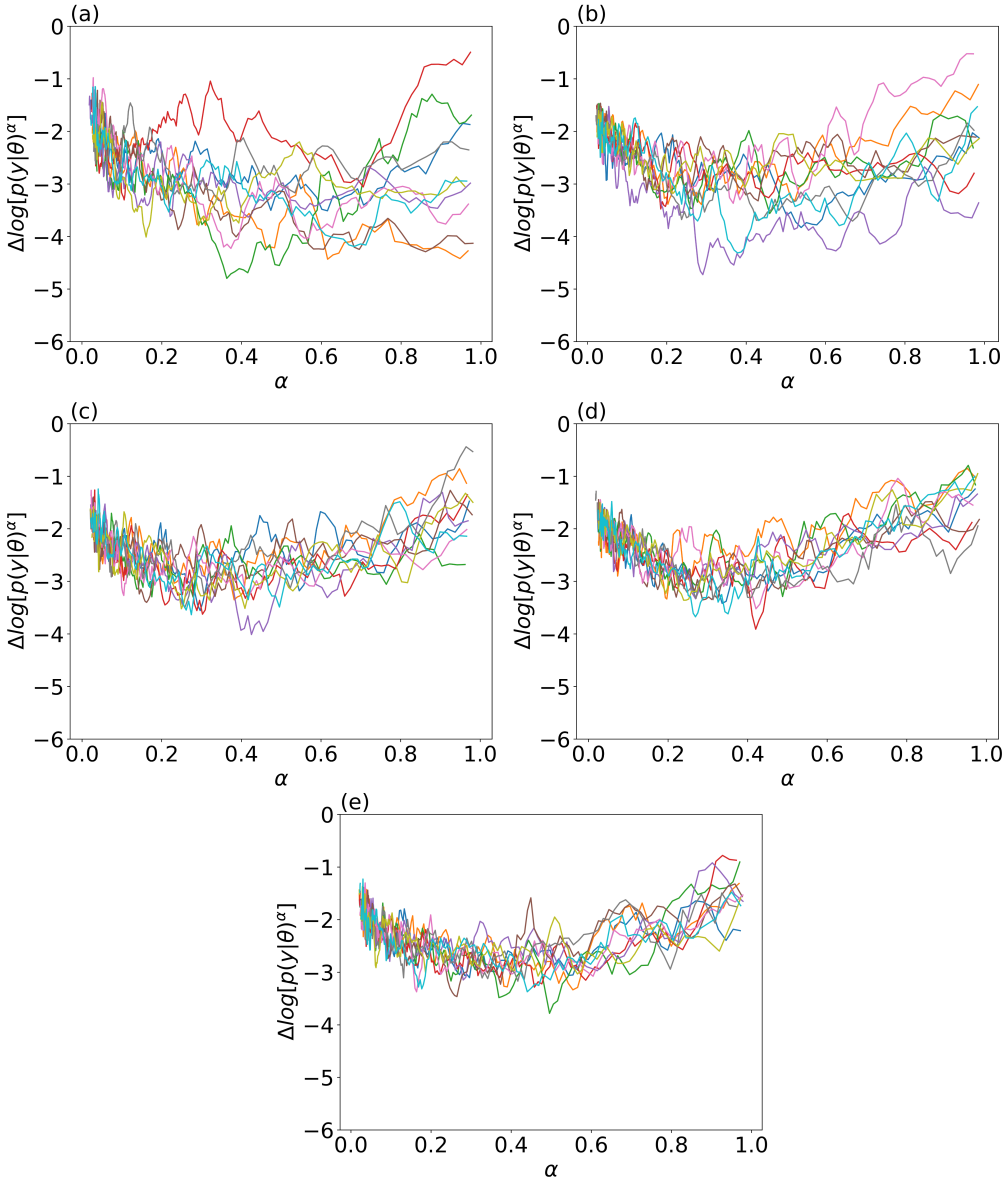

Figure 2.12: Evolution of the difference between the weighted mean log-likelihood $\hat{p}(\mathbf{y}|\boldsymbol{\theta})$ and the target log-likelihood calculated with the noise realization $p_n(\mathbf{y}|\boldsymbol{\theta})$ raised to α, where each color represents one replication, for (a) $K = 1$, (b) $K = 3$, (c) $K = 5$ (d) $K = 10$, and (e) $K = 20$

2.4 Discussion

Our results suggest that ASMC can provide accurate approximations of posterior PDFs for challenging inverse problems for which state-of-the-art adaptive MCMC fails to converge when considering a similar number of forward simulations (Figure 2.8). Furthermore, ASMC is very well suited for parallel computation, which is less the case for most MCMC methods. A general recommendation for practical applications is that the algorithmic variables K and $CESS_{op}$ in Algorithm 1 are chosen sufficiently large to ensure that the weighted-mean likelihood of the particles is close to the target likelihood during the ASMC run (Figure 2.12). Clearly, if the total number of forward simulations are insufficient, the ASMC algorithm fails in sampling posterior realizations of high likelihood for most particles. This leads to an impoverished particle approximation of the posterior PDF as evidenced by few surviving Eve indices (Figure 2.11) and mode collapse.

A similar argument holds for the evidence estimation. ASMC provides an unbiased estimation, as shown for the high-noise setting example (section 2.3.5). However, the evidence estimation procedure will only be reliable if the particles approximate the target power posteriors well enough. In addition, too low K and $CESS_{op}$ lead to frequent resampling that increases the estimation variance. Our results also suggest that error approximations based on single ASMC runs (eqs. 2.15 and 2.16) are too optimistic in such settings, but reliable for sufficiently long ASMC runs (Table 2.3). We also note that the relative standard deviations of the evidence estimates (Figure 2.10) are several orders of magnitude smaller than the evidences obtained for the consistent and inconsistent prior models (Table 2.2).

Providing practical recommendations for parameter settings away from easily-recognizable degenerate conditions is challenging. Of course, the larger the N the better, as the particle approximation of the parameter space will be improved. Our choice of $N = 40$ was dictated by the number of forward runs we could perform in parallel on one compute node, while much larger values are possible on modern computational architectures. An important point is how well the posterior can be described by a weighted average of N particles. The complexity of the posterior distribution depends on several factors like the dimension of the parameter space, the physics, the number and type of data, and the experimental design. Consequently, a much larger number of particles might be needed in challenging high-dimensional settings with strong parameter correlations or for problems with multi-modal posterior PDFs. In agreement with *Neal* (2001), we recommend distributing the total number of forward runs for each ASMC particle by favouring a large number of intermediate distributions over larger K. In practice, we typically first choose a suitably large $CESS_{op}$ and then vary K. In contrast to K, the influence of $CESS_{op}$ on the total number of forward simulations is non-linear and difficult to predict before running the algorithm. The trial tests in this study suggest that $CESS_{op}$ needs to be larger than $0.99N$, for our considered ranges of K, in order to reach the target misfit and build a smooth α-sequence. After fixing $CESS_{op}$, one can then first run the ASMC with an initially small K before re-running it with a twice as large value. If the difference between the resulting evidence estimates for these two choices of K are much smaller than the computed evidences for competing conceptual models, and if the inferred posteriors are similar, then this choice of K is probably sufficient. If important differences are observed between the ASMC runs obtained for the different K, then one needs to further

double K, and so on. Finally, the proposal scale ϵ needs to be initialized with a high enough value such that the initial acceptance rate is above AR_{min}. After this, the automatic rescaling of this parameter ensures high-quality estimates regardless of the model proposal scheme.

The observed relative insensitivity of the ASMC results to the model proposal type (Figures 2.4 and 2.5) is noteworthy, as the MCMC results (Figure 2.6) are highly sensitive to this choice. CM1 and CM2 present different levels of complexity. For CM1, MCMC-DREAM achieves convergence without difficulty (Fig. 2.6i), while this is far from being the case for MCMC-Gauss (Figure 2.6j). For CM2, both MCMC approaches fail (Figures 2.6k and l), while ASMC-DREAM and ASMC-Gauss perform similarly well for both CM1 and CM2 (Figures 2.4 and 2.5). The underlying reason for the success of ASMC and its insensitivity to the proposal mechanism is likely found due to the following factors. On the one hand, the adaptive scaling of the proposals (e.g., Figure 2.4c) and the tempering (e.g., Figure 2.4d) allow the particles to more easily move away from local minima, while resampling, on the other hand, gives priority to the high-likelihood regions (e.g., Figure 2.4h). Clearly, no such tuning of the proposal scale is possible when using MCMC as it violates detailed balance conditions. We stress that the comparisons made herein are with MCMC algorithms running at a unitary temperature, while parallel tempering-based MCMC methods might not have these problems (*Sambridge*, 2014).

The presented ASMC method share similarities to other approaches for evidence estimation. Nested Sampling (*Skilling*, 2004) reduces the evidence multidimensional integral to sampling of a one-dimensional integral over prior mass elements, using an increasing constraint on the log-likelihood lower bound. Other methods rely on MCMC sampling using power posteriors. For instance, thermodynamic integration (TIE) (*Gelman and Meng*, 1998), also called path sampling, reduces the evidence computation to a one-dimensional integral of the expectation of the likelihood over α. *Zeng et al.* (2018) shows that TIE performs better than nested sampling in terms of accuracy and stability. Stepping Stone Sampling (SS) (*Xie et al.*, 2011) also rely on power-posteriors but improves in accuracy compared with TIE by formulating the evidence estimation by the product of ratios of intermediate normalizing constants, that is, similarly to AIS and ASMC. An important practical difference is that SS is often performed in parallel by running multiple MCMC runs targeting different power posteriors (*Brunetti et al.*, 2019). Since each chain starts from the prior, the total computational cost is high, and perhaps more importantly, there is no solution to deal with MCMC chains for α close to one that do not converge (as in our MCMC trials with both MCMC-Gauss and MCMC-DREAM for CM2). This latter problem can be circumvented by running the SS algorithm sequentially using a similar tempering sequence as for ASMC. However, the α-sequence needs to be pre-defined, while ASMC allows for adaptive tuning. Even if not presented here, we stress that the improvements offered by ASMC over AIS are drastic. Despite extensive testing and tuning of AIS parameters, we were unable to match the performance of ASMC.

2.5 Conclusions

This study demonstrates that adaptive sequential Monte Carlo (ASMC) is a powerful method to approximate the posterior PDF and estimate the evidence in non-linear geophysical inverse problems. Crosshole GPR examples in which complex geological priors are parameterized through deep generative networks are used for demonstration purposes, but the method is of wide applicability. ASMC is robust with respect to the type of model proposals used and to algorithmic settings, implying a comparatively low user effort required for tuning the algorithm for a given application. ASMC is particularly useful for moderately to strongly non-linear inverse problems and for multi-modal distributions, where targeting the posterior distribution with MCMC algorithms may result in poor convergence. For the considered examples, ASMC outperforms state-of-the-art adaptive MCMC in estimating posterior PDFs. The major advantage of ASMC compared with MCMC in a Bayesian model selection context is that it provides straightforward computation of the evidence. Reliable uncertainty estimation of evidence estimates is possible from single ASMC runs, provided that they are long enough. We hope that this study will stimulate further adaptations of sequential Monte Carlo in a geophysical context, and more specifically, lead researchers to the adaptation of ASMC when confronted with challenging inference problems and model selection tasks.

Chapter 3

Hydrogeological multiple-point statistics inversion by adaptive sequential Monte Carlo

Macarena Amaya, Niklas Linde and Eric Laloy.

Published[1] in *Advances in Water Resources*, herein slightly adapted to fit the theme of this thesis with an appendix added after publication.

[1]Amaya et al. (2022). Hydrogeological multiple-point statistics inversion by adaptive sequential Monte Carlo. *Advances in Water Resources*, **166** 104252, doi:10.1016/j.advwatres.2022.104252

Abstract

For strongly non-linear and high-dimensional inverse problems, Markov chain Monte Carlo (MCMC) methods may fail to properly explore the posterior probability density function (PDF) given a realistic computational budget and they are generally poorly amenable to parallelization. Particle methods approximate the posterior PDF using the states and weights of a population of evolving particles and they are very well suited to parallelization. We focus on adaptive sequential Monte Carlo (ASMC), an extension of annealed importance sampling (AIS). In AIS and ASMC, importance sampling is performed over a sequence of intermediate distributions, known as power posteriors, linking the prior to the posterior PDF. They also provide estimates of the evidence (marginal likelihood) as needed for Bayesian model selection, at basically no additional cost. ASMC performs better than AIS as it adaptively tunes the tempering schedule and performs resampling of particles when the variance of the particle weights becomes too large. We consider a challenging synthetic groundwater transport inverse problem with a categorical channelized 2-D hydraulic conductivity field defined such that the posterior facies distribution includes two distinct modes. The model proposals are obtained by iteratively re-simulating a fraction of the current model using conditional multiple-point statistics (MPS) simulations. We examine how ASMC explores the posterior PDF and compare with results obtained with parallel tempering (PT), a state-of-the-art MCMC inversion approach that runs multiple interacting chains targeting different power posteriors. For a similar computational budget, ASMC outperforms PT as the ASMC-derived models fit the data better and recovers the reference likelihood. Moreover, we show that ASMC partly retrieves both posterior modes, while none of them is recovered by PT. Lastly, we demonstrate how the power posteriors obtained by ASMC can be used to assess the influence of the assumed data errors on the posterior means and variances, as well as on the evidence. We suggest that ASMC can advantageously replace MCMC for solving many challenging inverse problems arising in the field of water resources.

3.1 Introduction

Markov chain Monte Carlo (MCMC) methods are widely used to tackle probabilistic inverse problems arising in hydrology. As the dimensionality of the parameter space and the non-linearity of the forward problem increases, standard MCMC methods often fail to explore the posterior probability density function (PDF) given realistic computational constraints. This happens as the Markov chains may be trapped in local minima for long times or they may be unable to move between modes of high posterior probability. To circumvent such issues, methods exploring a series of so-called power posteriors have been developed. In a power posterior, less weight is given to the likelihood function as it is raised to the inverse of a temperature that is larger than one, something that is typically referred to as tempering. Tempering-based methods take advantage of the enhanced freedom of exploration at higher temperatures (sampling closer to the prior PDF) as popularized by the widely-used simulated annealing method for global optimization (*Kirkpatrick et al.*, 1983).

Parallel tempering (*Earl and Deem*, 2005) is a MCMC method in which many interacting chains target different power posteriors. Through proposed swaps of the states between chains, states sampled at higher temperatures can act as model proposals in the chains targeting the posterior distribution, also called unit temperature chains. Analogous to classical MCMC methods, PT approximates the posterior PDF by the states of the unit temperature chains sampled after burn-in. In the context of geophysical inversion, *Sambridge* (2014) demonstrated that PT can drastically improve sampling efficiency leading to an expanded exploration of the parameter space compared to standard MCMC. The PT method has lately been applied in a range of geoscientific problems such as landscape evolution (*Chandra et al.*, 2019), groundwater flow and transport (*Laloy et al.*, 2016; *Reuschen et al.*, 2020) and earthquake source inversion (*Gallovič et al.*, 2019).

A highly parallelizable alternative to MCMC is offered by particle methods such as annealed importance sampling (AIS, *Neal* (2001)) and sequential Monte Carlo (SMC, *Doucet and Johansen* (2011)), where the posterior distribution is approximated using a weighted sample of particle states. In these methods, importance sampling steps are performed sequentially along a sequence of power posteriors. What differentiates these two methods from each other is that SMC performs resampling of the particle population when the variance of the importance weights becomes too high. One outstanding feature of both methods is that the evidence, the normalizing constant in Bayes' theorem and the key quantity in Bayesian model selection, is estimated as well. Compared with its extensive use in science and engineering, SMC appears poorly explored in the Earth sciences (*Linde et al.*, 2017). *Zhou et al.* (2016) proposed an adaptive version of SMC (ASMC) that automatically tunes the step size (temperature reduction) between neighboring power posteriors. Recently, adaptive SMC algorithms were introduced and successfully implemented in geophysical applications for posterior PDF and evidence estimations (*Amaya et al.*, 2021; *Davies et al.*, 2023).

Realistic geological priors can often not be expressed by two-point geostatistical models (e.g., multivariate Gaussian), for example, when connectivity patterns play an essential role in determining the system response (*Gómez-Hernández and Wen*, 1998; *Renard and Allard*, 2013). Multiple-point statistics (MPS) is a sub-field of geostatistics aiming at generating conditional model realizations that honor higher-order statistics found in a so-called training image, a gridded 2-D or 3-D representation of the spatial field of interest that is built from generic or previous geological knowledge of the site (*Mariethoz and Caers*, 2014). To generate MPS-based candidate models within MCMC inversions, one popular approach is sequential geostatistical resampling (SGR) (*Ruggeri et al.*, 2015), in which model proposals are generated by re-simulating a random fraction of the current model conditioned to the remaining grid values. The SGR framework embraces two end-member strategies: either a randomly located boxed-shaped area is resimulated as in sequential Gibbs sampling by *Hansen et al.* (2012) and in blocking MCMC by *Fu and Gómez-Hernández* (2009), or random points throughout the model domain are resimulated as in iterative spatial resampling by *Mariethoz et al.* (2010a). Recently, hybrid methods determining an optimal combination of these end-members have been proposed by *Reuschen et al.* (2021). Other approaches rely on much faster model proposals offered, for instance, by graph cuts (*Zahner et al.*, 2016) or by encoding the complex priors in a much lower-dimensional space using deep generative networks, thereby, reducing

the number of inferred parameters from several thousands of unknowns to some tenths of latent variables (*Laloy et al.*, 2017, 2018).

In this paper, we consider the challenging groundwater transport inverse problem introduced by *Laloy et al.* (2016). It consists of a 2-D synthetic tracer experiment in which concentration is monitored at pumping wells and the aim is to recover the hydraulic conductivity field assuming a binary geological media (their case study 2). This test case is particularly challenging for three reasons: (i) the underlying non-linearity caused by long-range connectivity of high-conductivity zones and a conductivity ratio of 100 between permeable channels and less permeable matrix material, (ii) a large number of observations with a high signal-to-noise ratio and (iii) a true field that is designed such that the targeted posterior distribution is bimodal with the modes being located far from each other. *Laloy et al.* (2016) demonstrated how PT clearly outperforms standard MCMC when used within a SGR framework. Nevertheless, even if PT offered important improvements it did not sample any of the posterior modes and the simulated data of the generated model realizations did not fit the true data to the level of the added noise. Compared to PT, ASMC presents the following advantages: (i) adaptive determination of the temperature schedule, (ii) the model proposal scale is tuned adaptively using the acceptance rate at the previously considered temperature and, (iii) the evidence is calculated along the run with updates being made every time the temperature changes. In PT, the temperature schedule and the proposal scale need to be pre-defined. The evidence estimation in PT is reduced to a one-dimensional integral over the inverse temperature, which can imply large approximation errors if the temperatures are comparatively few or poorly chosen. We assess the performance of ASMC for this test case and compare the results with the PT results of *Laloy et al.* (2016). We further discuss the insights offered by analysing the results at intermediate temperatures corresponding *de facto* to assumptions of larger measurement noise. With respect to the geophysical ASMC study by *Amaya et al.* (2021), the present work considers a hydrogeological problem that is much more non-linear and the model parameterizations and model proposal schemes are entirely different. In ASMC-SGR we need to consider as many model parameters as there are pixels (7500 in our example) while the deep generative network used by *Amaya et al.* (2021) only considered a few tenths of unknowns.

3.2 Method

3.2.1 Bayes' theorem

It is often beneficial to pose inverse problems within a probabilistic framework using Bayes' theorem, in which the parameters to infer are treated as random variables. If we consider a conceptual model composed by parameters $\boldsymbol{\theta}$, the posterior pdf $\pi(\boldsymbol{\theta}|\mathbf{y})$ is given by:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta})\, p(\mathbf{y}|\boldsymbol{\theta})}{\pi(\mathbf{y})}. \tag{3.1}$$

The prior PDF $\pi(\boldsymbol{\theta})$ represents the a priori information concerning the model parameters. This information is then weighted by the likelihood function $p(\mathbf{y}|\boldsymbol{\theta})$ that expresses, for a given noise model, how probable it is that a particular set of parameter values have produced the observations $\mathbf{y}$. Assuming the noise on the data to be uncorrelated and normally distributed with a constant variance σ^2, the likelihood is expressed as:

$$p(\mathbf{y}|\boldsymbol{\theta}) = (\sqrt{2\pi\sigma^2})^{-m_d} \exp\left[-\frac{1}{2\sigma^2}\sum_{i}^{m_d}(y_i - F_i(\boldsymbol{\theta}))^2\right], \tag{3.2}$$

where m_d is the number of data points and $F(\boldsymbol{\theta})$ the simulated data given a set of model parameter values. It can be convenient to consider the variable component of the natural logarithm of the likelihood:

$$l(\mathbf{y}|\boldsymbol{\theta}) = -\frac{1}{2\sigma^2}\sum_{i}^{m_d}(y_i - F_i(\theta))^2, \tag{3.3}$$

which we refer to as the reduced log-likelihood as it ignores the constant terms.

The evidence, also known as the marginal likelihood, is the normalizing constant in Bayes' theorem. This quantity used to rank alternative conceptual models, defined by different prior models, represents how consistent a conceptual model is with the set of observations under consideration (*Kass and Raftery*, 1995). The evidence is a multidimensional integral over the parameter space:

$$\pi(\mathbf{y}) = \int \pi(\boldsymbol{\theta})\, p(\mathbf{y}|\boldsymbol{\theta})\, d\boldsymbol{\theta}, \tag{3.4}$$

making it very challenging to calculate for high-dimensional models. *Brunetti et al.* (2019) focus particularly on how to compute the evidence to compare different conceptual models within a MPS framework.

3.2.2 Sequential geostatistical resampling

Prior models are often represented by mathematical functions allowing any prior model realization to be evaluated in terms of its probability. Examples include uniform priors, multivariate Gaussian priors and latent space distributions learned by a deep generative neural network. However, such explicit prior model parameterizations are not always suitable, or possible, when seeking to encode realistic geological spatial heterogeneity (*Linde et al.*, 2015). As an alternative, one can instead consider realizations of MPS simulation tools (*Strebelle*, 2002) as samples drawn from the prior model. These realizations honor the higher-order

statistics of training images that can be built based on expected geological structures, outcrops, geophysical or borehole data. The downside of such prior sampling-based approaches is that one cannot calculate the prior probabilities of model realizations as needed in most MCMC algorithms.

Sequential geostatistical resampling is a mechanism allowing MCMC inference when model proposals are drawn using MPS algorithms that sample proposals proportionally to the prior density. It builds on the foundational paper by *Mosegaard and Tarantola* (1995) in geophysics. However, it is noteworthy that the underlying philosophy of such a prior sampling-based algorithm has more recently received strong theoretical backing in the context of infinite-dimensional inversion problems (e.g., *Cotter et al.* (2013)). At each MCMC iteration, a new model proposal is generated by re-simulating a random fraction of the current model realization using an MPS algorithm, conditioned to the remaining pixel values. There are two end-member approaches to determine the locations of the pixels that are to be resimulated: either a randomly located box-shaped area (*Alcolea and Renard*, 2010; *Hansen et al.*, 2012) or randomly located points (*Mariethoz et al.*, 2010a). In this study, we use boxes as it provided the best results in *Laloy et al.* (2016). We further rely on the DeeSse MPS algorithm (http://www.randlab.org/research/deesse/) that is, in turn, based on the direct sampling method by *Mariethoz et al.* (2010b). To re-simulate the value of a certain uninformed pixel, the algorithm scans the training image searching for patterns that agree with those found in the vicinity of this pixel. If a similar-enough pattern is found, it assigns the value of the pixel under consideration in the training image to the one in the new proposed model. This procedure is repeated for all the pixels that are to be re-simulated.

In MCMC algorithms, the Metropolis rule is used to accept or reject model proposals obtained from symmetric proposal distributions. The acceptance probability Γ to move from a current state $\boldsymbol{\theta}_c$ to a proposed state $\boldsymbol{\theta}_p$ is:

$$\Gamma(\theta_p, \theta_c) = min\left(1, \frac{\pi(\boldsymbol{\theta}_p)\, p(\mathbf{y}|\boldsymbol{\theta_p})}{\pi(\boldsymbol{\theta}_c)\, p(\mathbf{y}|\boldsymbol{\theta_c})}\right). \tag{3.5}$$

As mentioned above, this rule cannot be used with MPS algorithms such as DeeSse as $\pi(\boldsymbol{\theta})$ is unknown. Instead, MPS-based inversions often rely on the extended Metropolis (*Mosegaard and Tarantola*, 1995) method that is applicable if the model proposal mechanism generates samples drawn proportionally to the prior PDF. The acceptance probability is then reduced to:

$$\Gamma(\theta_p, \theta_c) = min\left(1, \frac{p(\mathbf{y}|\boldsymbol{\theta_p})}{p(\mathbf{y}|\boldsymbol{\theta_c})}\right) \tag{3.6}$$

and involves only likelihood ratios.

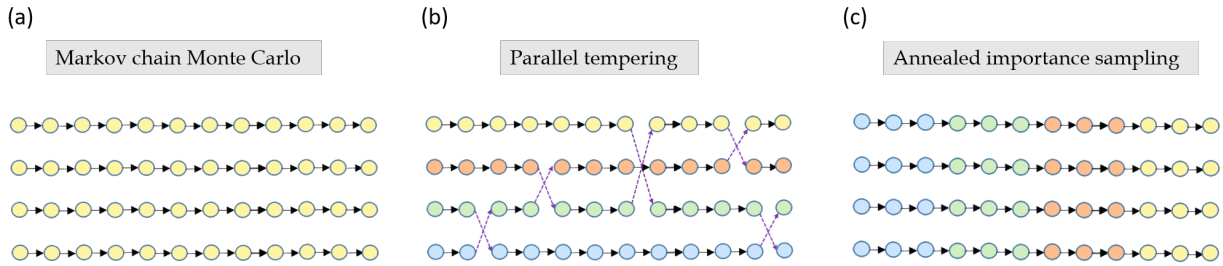

Figure 3.1: Diagram illustrating structural differences between probabilistic inversion methods with the circles representing the evolving states at each iteration and the colors denoting the temperatures of the targeted power posteriors. In (a) standard MCMC, all the chains target the posterior PDF at unit temperature (shown in yellow). (b) Parallel tempering uses a number of chains targeting different power posterior while allowing eventual swaps between them (shown as purple dashed lines), whereas in (c) annealed importance sampling, the targeted power posteriors change during the run in response to a gradually cooling sequence.

3.2.3 Adaptive sequential Monte Carlo

Power posteriors

Tempering consists in introducing a temperature variable flattening the likelihood function in equation 3.1. The corresponding tempered posterior PDFs are called power posteriors and can, in their unnormalized form, be expressed as:

$$\gamma_t(\boldsymbol{\theta}_t|\mathbf{y}) \equiv \pi(\boldsymbol{\theta}_t)\, p(\mathbf{y}|\boldsymbol{\theta}_t)^{\alpha_t}, \tag{3.7}$$

where the likelihood is raised to an inverse temperature $\alpha_t \in [0,1]$. The effect of increasing the temperature (decreasing α_t) is that the likelihood function becomes less peaky, that is, with less pronounced modes. Targeting these power posteriors, instead of only targeting the posterior PDF at unit temperature as in standard MCMC, increases the exploration capacity because the tempering process decreases the probability of getting trapped in local minima. A graphical explanation regarding the advantages of tempered exploration can be found in *Sambridge* (2014). Figure 3.1 illustrates the main structural differences between standard MCMC, and the methods of PT and AIS that both rely on tempering.

Annealed importance sampling (AIS)

Importance sampling is a Monte Carlo method used to estimate properties of a distribution that it is not possible to sample from (*Hammersley and Handscomb*, 1964). It relies on an auxiliary distribution $q(\boldsymbol{\theta})$ for drawing the samples, that must include and should ideally be slightly inflated with respect to the target distribution. For most applications, sampling from the prior distribution in order to estimate properties of the posterior PDF suffers from the curse of dimensionality, meaning that the computational effort needed to draw enough samples with a significant likelihood, as needed to enable reliable estimates, is unfeasible. In contrast, sampling from a well-chosen importance distribution allows focusing the sampling

in regions of high posterior probability. The samples drawn are then used to compute the desired property while correcting for the bias resulting from the chosen importance distribution. If the target distribution is the unnormalized posterior PDF $\pi(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})$, the importance weights are given by:

$$w = \frac{\pi(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{q(\boldsymbol{\theta})}. \tag{3.8}$$

Neal (2001) combined tempering and importance sampling to produce the AIS method. It uses N chains, each of them representing evolving particles that target sequentially a sequence of power posteriors at different temperatures ranging from the prior to the unnormalized posterior PDF of interest. The sequence is given by $\{\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})\}_{t=0}^{T}$, and it contains unnormalized power posteriors given by equation 3.7 with α_t ranging from $\alpha_{t=0} = 0$ (the prior) to $\alpha_{t=T} = 1$ (the unnormalised posterior PDF). The normalized power posteriors are given by:

$$\pi_t(\boldsymbol{\theta}_t|\mathbf{y}) = \frac{\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})}{Z_t}, \tag{3.9}$$

where Z_t is the normalizing constant of the distribution.

In AIS, importance sampling steps are performed sequentially between each pair of consecutive power posteriors. A subsequent power posterior $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$ is approximated by using the estimation of the previous power posterior $\gamma_{t-1}(\boldsymbol{\theta}_t|\mathbf{y})$ as the importance sampling distribution. In contrast to standard importance sampling were the samples are drawn directly from the importance distribution, in AIS the $\gamma_{t-1}(\boldsymbol{\theta}_t|\mathbf{y})$ samples are obtained by performing K MCMC iterations targeting this power posterior starting from the approximation $\gamma_{t-2}(\boldsymbol{\theta}_t|\mathbf{y})$. By performing multiple intermediate importance sampling steps between the prior and the posterior PDF, it is possible to ensure that each importance distribution is of high quality (slightly inflated with respect to the target) leading to estimates with low uncertainty (variance). After the importance sampling step (represented by the longer arrows in between different colored circles in Figure 3.2), again each of the N chains perform K MCMC steps targeting now $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$. This process is repeated until $\alpha_t = 1$

We refer to the importance weights (equation 3.8) resulting from each intermediate importance sampling step as the incremental weights. For a particle i at state $\boldsymbol{\theta}_{t-1}^{i}$, the incremental weight w_t^i that result from using $\gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})$ as an importance distribution for $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$ is:

$$w_t^i = \frac{\gamma_t(\boldsymbol{\theta}_{t-1}^{i}|\mathbf{y})}{\gamma_{t-1}(\boldsymbol{\theta}_{t-1}^{i}|\mathbf{y})}. \tag{3.10}$$

To calculate the total weight of a particle, one needs to account for all the intermediate importance sampling steps. To achieve this, the incremental weight w_t^i is used to update the normalized weight of particle i by:

$$W_t^i = \frac{W_{t-1}^i w_t^i}{\sum_{j=1}^N W_{t-1}^j w_t^j}, \tag{3.11}$$

where W_{t-1}^i is the normalized weight (that is, $\sum_{i=1}^N W_{t-1}^i = 1$) of the previous importance sampling step. The posterior PDF is then approximated through a particle approximation, in which the relative probabilities of the last N states of the particles are determined by the final normalized weights W_T^i. By saving intermediate normalized weights and corresponding particle states, the method allows also to approximate intermediate power posteriors that represent the solutions to the equivalent tempered problems.

Resampling

The variance of the particle weights influences strongly the quality of the importance sampling estimator (*Neal*, 2001). When using AIS, this variance may grow exponentially, resulting in poor estimations of the posterior PDF and the evidence. Sequential Monte Carlo (SMC) is a family of particle approaches that, as AIS, rely on sequential importance sampling. However, SMC incorporates also resampling (*Del Moral et al.*, 2006; *Doucet and Johansen*, 2011). In a resampling step, the states of the particles are replicated according to a probability that is proportional to their current normalized weights, and all the weights are re-set to $1/N$. The replacement of particles with lower weights and increasing those with higher weights results in two advantages: (i) it avoids the variance of the weights to grow indefinitely and (ii) it orients the exploration towards regions of higher posterior probability. Nevertheless, since the resampling process increases the variance of the estimates (*Douc and Cappé*, 2005), it is often better to perform resampling only when needed. The effective sample size (ESS) (*Kong et al.*, 1994) is expressed as:

$$ESS_t = \frac{(\sum_{i=1}^N W_{t-1}^i w_t^i)^2}{\sum_{j=1}^N (W_{t-1}^j)^2 (w_t^j)^2}. \tag{3.12}$$

It quantifies the number of effective samples in the particle approximation. The common approach is to monitor the ESS along the run, and perform resampling when it is lower than a specified threshold. In this paper, we rely on systematic resampling due to its good performance and easy implementation (*Doucet and Johansen*, 2011). Figure 3.2 shows a graphical example of SMC with $N = 5$ particles, in which the resampling step is indicated with red dashed lines.

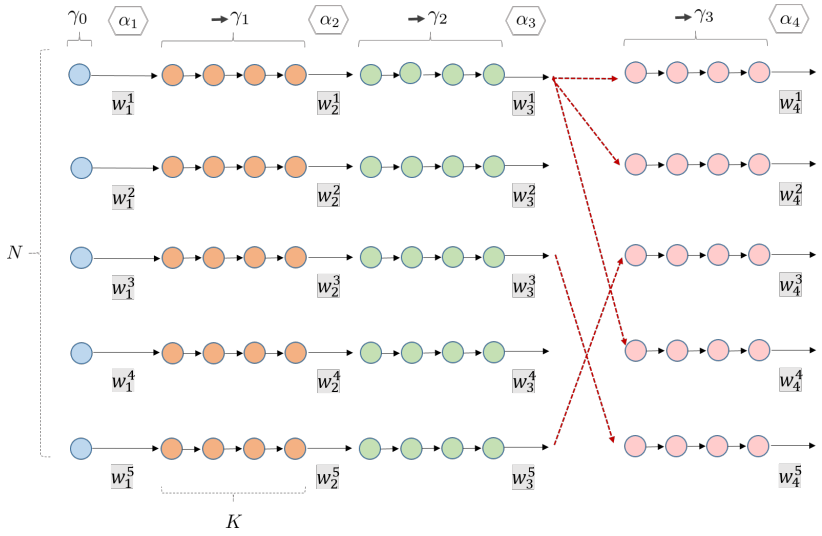

Figure 3.2: Schematic representation of sequential Monte Carlo (SMC) using $N = 5$ particles evolving in parallel. After the initial sampling from the prior PDF (blue circles), $K = 4$ Markov steps are performed to approximate one of the power posteriors γ_t. In these power posteriors, the likelihood is raised to an inverse temperature α_t that increases gradually. At the end of each approximation, an importance sampling step is performed to calculate an incremental weight w_t. Adaptive sequential Monte Carlo (ASMC) incorporates two modifications with respect to annealed importance sampling (AIS): (i) adaptive determination of the α_t-sequence defining the sequence of power posteriors and (ii) resampling when the variance of the particle weights becomes too large (indicated by red dashed lines).

Adaptive tempering schedule

One complication of AIS and SMC methods is the difficulty to pre-define a suitable tempering schedule (Figure 3.2). *Zhou et al.* (2016) propose an adaptive SMC method (ASMC) (their algorithm 4) in which an appropriate α-step-size increment is determined before each importance sampling step. To do so, they rely on the conditional effective sample size (CESS) quantifying the quality of using the particle approximation $\gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})$ as an importance distribution to estimate expectations for the $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$ arising for different choices of α_t, and is given by:

$$CESS = N\frac{(\sum_{i=1}^{N} W_{t-1}^{i} w_t^i)^2}{\sum_{j=1}^{N} W_{t-1}^{j} (w_t^j)^2}. \tag{3.13}$$

The ESS and $CESS$ are both obtained by a sample approximation of a Taylor expansion of the relative variance of the estimator (*Kong et al.*, 1994). The difference between them is that the ESS embraces the accumulated mismatch between the importance and target distributions, whereas the $CESS$ focuses on the quality of the current importance sampling step. If resampling was to be performed at every iteration, then the ESS and $CESS$ quantities would be equal. A detailed derivation of the $CESS$ can be found in the supplementary material of *Zhou et al.* (2016).

The $CESS$ depends on the incremental weights w_t that in turn depend on α_t. The strategy consists in finding the α-increment between consecutive power posteriors, that is, the $\Delta\alpha_t$ such that $\alpha_t = \alpha_{t-1} + \Delta\alpha_t$, giving the $CESS$ that is the closest to a pre-defined quality expressed by $CESS_{op}$. To find $\Delta\alpha_t$, we rely on a binary search within a sequence of possible $\Delta\alpha$ values. First, the $CESS$ is computed using the middle value of the $\Delta\alpha$ sequence and it is compared with $CESS_{op}$. Depending on if it is higher or lower, one of the two $\Delta\alpha$ half-intervals is kept. This procedure is repeated until the $\Delta\alpha$ that gives the $CESS$ that is the closest to $CESS_{op}$ is found.

If we increase $CESS_{op}$, we obtain higher-quality estimates as the number L of intermediate power posteriors increases, but at the expense of a longer ASMC run. The total number of iterations per particle is $L \times K$, with K the number of MCMC steps per intermediate power posterior. In practice, the ratio $CESS_{op}/N$ is chosen close to 1 in order to ensure high quality estimates. It has been suggested that it should be at least 0.99 to build a smooth α-sequence (*Amaya et al.*, 2021), but the optimal value is highly problem-dependent. The impact of the $CESS_{op}/N$ value on the resulting L is non-linear and not easy to predict.

ASMC-based evidence estimation

Evidence estimation is essential for Bayesian model selection. Considering two neighbouring distributions $\gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})$ and $\gamma_t(\boldsymbol{\theta}_t|\mathbf{y})$, we can express the ratio of their normalizing constants as:

$$\frac{Z_t}{Z_{t-1}} = \frac{\int \gamma_t(\boldsymbol{\theta}_t|\mathbf{y})\, d\boldsymbol{\theta}_t}{\int \gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})\, d\boldsymbol{\theta}_{t-1}}. \tag{3.14}$$

Del Moral et al. (2006) propose an approximation of this ratio as:

$$\frac{Z_t}{Z_{t-1}} \approx \sum_{i=1}^{N} W_{t-1}^i w_t^i. \tag{3.15}$$

The evidence $\pi(\mathbf{y})$ is the normalizing constant Z_T of the unnormalized posterior PDF, that is, the last distribution of the sequence when $\alpha_{t=T} = 1$. Considering that the prior PDF integrates to one, $Z_0 = 1$, we can express the evidence as the product of the normalizing constant ratios:

$$\pi(\mathbf{y}) = Z_T = \frac{Z_T}{Z_0} = \prod_{t=1}^{T} \frac{Z_t}{Z_{t-1}} \approx \prod_{t=1}^{T} \sum_{i=1}^{N} W_{t-1}^i {w_t}^i. \tag{3.16}$$

Consequently, the evidence can be updated along the run by accounting for the evolving particle weights.

3.2.4 Full ASMC-SGR algorithm

Our algorithm combining the SGR method for model proposals with ASMC for posterior PDF and evidence estimation is given in Algorithm 2. We denote this algorithm as ASMC-SGR (following the nomenclature in Laloy et al., 2016). In this study, the proposal scale ϕ indicates half of the side-length in meters of the box that is being re-simulated at each iteration. In addition to the previously mentioned advantages of adaptive tempering and evidence estimation, the algorithm also has the attractive feature that the proposal scale ϕ can be tuned on-the-go without violating detailed balance conditions as would be the case for MCMC. This is simply achieved by keeping track of the acceptance rate for the K MCMC steps at the previous α_{t-1} and then to use this information to adapt the proposal scale for the next number of K MCMC steps to ensure that it is within a pre-defined range. This saves a lot of time compared with standard MCMC and PT algorithms that often necessitate tuning using multiple time-consuming trial runs.

Algorithm 2: ASMC-SGR

The SGR section of the algorithm is adapted from *Laloy et al.* (2016) and the ASMC section from *Zhou et al.* (2016) (their algorithm 4).

Variables to pre-define:

Number of particles (N), optimal CESS ($CESS_{op}$), ESS threshold (ESS^*),
number of MCMC iterations at each intermediate distribution (K), minimal and maximal acceptance rate (AR_{min},AR_{max}),
minimal (ϕ_{min}) and maximal (ϕ_{max}) proposal scale and its percentage of change (f).

Initialization: Set $t = 0$

Set $\alpha = 0$

Sample θ_0 from the prior $\pi(\theta)$ N times

Set the N-dimensional vector of normalized weights $\mathbf{W}_0 = [\frac{1}{N}; \frac{1}{N}; ...; \frac{1}{N}]$

Set evidence $\pi(\mathbf{y}) = 1$

Iteration : Set $t = t + 1$

Search for incremental distribution

Do binary search for the increment $\Delta\alpha$ that gives the CESS (equation 3.13) that is the closest to $CESS_{op}$

Update $\alpha = min(1, \alpha + \Delta\alpha)$ and define the following intermediate distribution $\gamma_t(\theta_t|\mathbf{y}) = \pi(\theta_t)p(\mathbf{y}|\theta_t)^\alpha$

Perform the importance sampling step: compute the weight increments w_t^i (equation 3.10),
update and save the normalized weights W_t^i (equation 3.11) and the evidence $\pi(\mathbf{y}) = \pi(\mathbf{y})\sum_{i=1}^N W_{t-1}^i w_t^i$ (equation 3.16)

Resampling

Calculate ESS (equation 3.12), if $ESS < ESS^*$ do resampling: re-organize θ_t states and update $\mathbf{W}_t = [\frac{1}{N}; \frac{1}{N}; ...; \frac{1}{N}]$

Do K MCMC iterations for each of the N particles (chains):

Propose moves θ_p : randomly select the location of a box with dimensions $2\phi \times 2\phi$, and run the MPS simulation using
the points outside the box as conditioning points and accept or reject based on the extended Metropolis rule,
with an acceptance probability given by: $\Gamma(p, c) = min\left(1, \frac{p(\mathbf{y}|\boldsymbol{\theta_p})^\alpha}{p(\mathbf{y}|\boldsymbol{\theta_c})^\alpha}\right)$ (equation 3.6)

Save the N models θ and their likelihoods.

Set last state as θ_{t+1}

Tune proposal scale

If acceptance rate $AR < AR_{min}$ then decrease proposal scale factor: $\phi = \phi * (1 - \frac{f}{100})$

If acceptance rate $AR > AR_{max}$ then increase proposal scale factor: $\phi = \phi * (1 + \frac{f}{100})$

If $\phi < \phi_{min}$ then $\phi = \phi_{min}$, if $\phi > \phi_{max}$ then $\phi = \phi_{max}$

Repeat until α=1

3.3 Results

3.3.1 Test case

We consider the second test case from *Laloy et al.* (2016), in which the concentration of an injected tracer is measured at regular time intervals. The conceptual model is represented by a 250 × 250 categorical binary training image from *Strebelle* (2002) (Figure 3.3). The 2-D reference model is located in the x-y plane and has a dimension of 75 m × 100 m with a discretization cell size of 1 m. The hydraulic conductivity K is 0.01 m/s for the channels and 0.0001 m/s for the matrix. A conservative tracer with a concentration of 1 kg/m^3 is injected at 8 locations on the top and bottom of the model (Figure 3.4). The concentration is measured every 8 hours during 10 days at 11 pumping wells (a total of 330 observations) that extract 0.0005 m^2/s of water, and the facies at these points are assumed to be known. This test exhibits symmetry with respect to the x-axis, such that any model and its mirrored image produce the same simulated concentration data and, therefore, the same likelihood. Consequently, the posterior PDF is bi-modal with two distinct modes. Mode 1 of the reference model was obtained as a random realization from the DeeSse algorithm (Figure 3.4a) and then mirrored to obtain the mode 2 reference model (Figure 3.4b).

The simulations are performed using MaFloT, a finite-volume open-source code for transport simulations in porous media (*Künze and Lunati*, 2012). Fixed head boundaries of 0 m on the

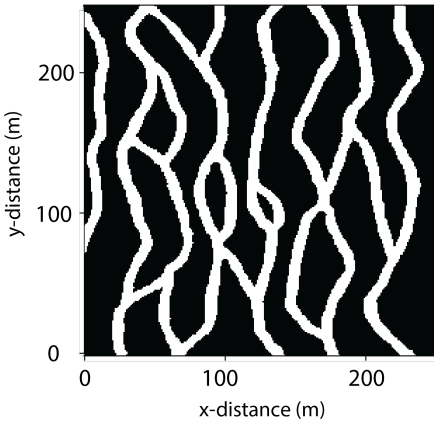

Figure 3.3: Channelized binary training image from *Strebelle* (2002) with the spatial dimensions used in the present study.

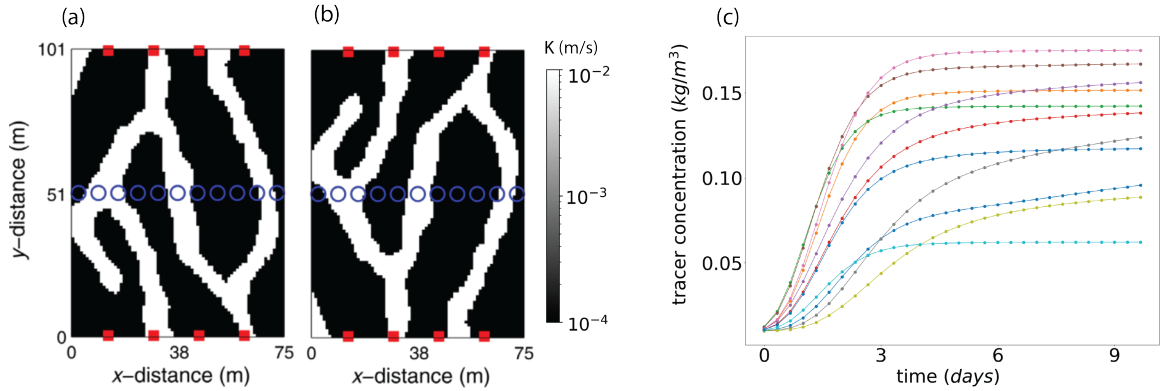

Figure 3.4: (a) Reference model [mode 1] and (b) mirrored reference model [mode 2], the red squares represent the points where the tracer is injected and the blue circles represent the pumping wells where measurements are made. Both models result in the same (c) simulated concentration over time (shown before being contaminated with uncorrelated Gaussian noise), where each color represents the observations at one pumping well.

top and bottom of the domain and no-flow boundaries on the sides are assumed to simulate steady-state groundwater flow. For the tracer transport, we assume open boundaries, an hydraulic dispersivity of 0.1 m and a background concentration of 0.01 kg/m^3. The simulated data were corrupted with uncorrelated Gaussian noise with a standard deviation of $\sigma = 0.003$ kg/m^3, approximately 3% of the mean concentration.

ASMC-SGR settings

The proposal scale ϕ used to create candidate models is tuned along the run (see Algorithm 2) by increasing or decreasing it by $f = 20\%$ to ensure that the acceptance rate stays within the range of $AR_{min} = 15\%$ and $AR_{max} = 35\%$. It is further constrained to be between $\phi_{max} = 50$ m and $\phi_{min} = 5$ m. For the DeeSse simulations, we follow *Laloy et al.* (2016) and use 75 neighbors, which implies that the patterns that are searched by the algorithm are composed of the 75 informed nodes that are the closest to the one being re-simulated. The fraction of

the training image that is scanned is 0.9 and the distance threshold to accept a pattern is 0.01 (*Mariethoz et al.*, 2010b).

3.3.2 ASMC-SGR results

Test 1: ASMC-SGR with 24 particles

We first compare the ASMC-SGR results with those obtained by *Laloy et al.* (2016) for a similar computational budget: 24 chains and 25,000 iterations per chain. To achieve this, we chose $N = 24$ particles and $CESS_{op}/N = 0.9997$ combined with $K = 18$, which resulted in 25,956 iterations per particle. The resampling threshold ESS^*/N was set to 0.3 (*Del Moral et al.*, 2006). The user-defined parameters and length of the run are summarized in Table 3.1 (ASMC-SGR 24p).

We first consider the evolution of the tempered log-likelihood, that is, the likelihood raised to the inverse temperature in the natural log-scale (Figure 3.5a). The tempered log-likelihood of each particle is seen to evolve according to the reference tempered log-likelihood curve (calculated using the assumed noise standard deviation $\sigma = 0.003$ kg/m^3). If $CESS_{op}/N$ or K would be too low, then the particles would have considerably lower tempered likelihoods than the reference curve, thereby, indicating that the sampled log-likelihoods are too low and that the associated computational budget is insufficient for the problem at hand. Consequently, this type of curve is a useful diagnostic plot allowing the user to terminate an ASMC run at an early stage if the tempered log-likelihoods falls below the reference curve.

The automatically tuned proposal scale ϕ (Figure 3.5e) enables the acceptance rate to stay within the pre-defined range (Figure 3.5c). The resulting α_t-sequence (Figure 3.5b) demonstrates that roughly half of the forward simulations are carried out with α_t-values less than 0.01, corresponding to temperatures above 100. The plot showing the evolution of the normalized weights (Figure 3.5d) illustrates the divergence of the weights between resampling steps and the re-alignment of the weights when the normalized effective sample size ESS/N (Figure 3.5f) reaches below the 0.3 threshold.

To compare these ASMC-SGR 24p results with those obtained by *Laloy et al.* (2016), we first consider the measure used in their study as an indicator of data fitting:

$$\Delta l(\mathbf{y}|\boldsymbol{\theta})[\%] = \frac{\bar{l}(\mathbf{y}|\boldsymbol{\theta}_T) - l(\mathbf{y}|\boldsymbol{\theta}_{ref})}{l(\mathbf{y}|\boldsymbol{\theta}_{ref})} \times 100, \tag{3.17}$$

where $l(\mathbf{y}|\boldsymbol{\theta}_{ref})$ is the reduced reference log-likelihood (equation 3.3) and $\bar{l}(\mathbf{y}|\boldsymbol{\theta}_T)$ is the mean sampled reduced log-likelihood. For MCMC and PT, this mean is simply the arithmetic average of the reduced log-likelihoods after burn-in (only considering unit temperature chains for PT), whereas for ASMC it is the weighted average of the N final reduced log-likelihoods. *Laloy et al.* (2016) demonstrated a drastic improvement when using PT-SGR compared with MCMC-SGR following *Hansen et al.* (2012). The indicator $\Delta l(\mathbf{y}|\boldsymbol{\theta}_t)$ was 9%; an

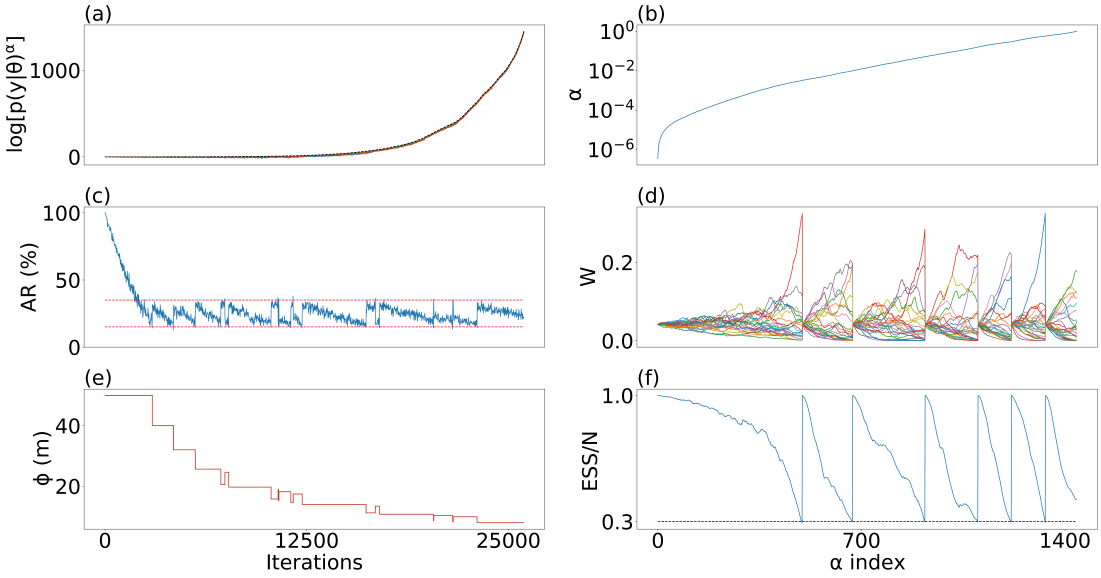

Figure 3.5: ASMC-SGR results using 24 particles: (a) tempered log-likelihood vs. iterations per particle, the colors represent different particles and the black dashed line indicates the reference tempered log-likelihood; (b) α-sequence vs. α index; (c) acceptance rate vs. iterations per particle, the dashed line indicates the pre-defined minimum and maximum range; (d) normalized weights vs. α index, the colors represents different particles; the (e) proposal scale vs. iterations per particle; (f) ESS/N vs. α index, the black dashed line indicates the 0.3 threshold below which resampling is made.

Table 3.1: User-defined parameters, resulting sequence length and data fitting for ASMC-SGR using 24 and 72 particles. The reduced reference log-likelihood $l(\mathbf{y}|\boldsymbol{\theta}_{ref})$ (equation 3.3) for this test case is -165. Using PT-SGR, *Laloy et al.* (2016) obtained a $\Delta l(\mathbf{y}|\boldsymbol{\theta}_t)$=9% for a numerical demand of 600,000 forward simulations.

	ASMC-SGR 24p	ASMC-SGR 72p
Particles (N)	24	72
$CESS_{op}/N$	0.9997	0.9997
ESS^*/N	0.3	0.3
AR_{min}	15%	15%
AR_{max}	35%	35%
K iterations	18	18
L power posteriors	1,442	1,533
Iterations per particle	25,956	27,594
Resampling times	6	6
Total number of forward simulations	622,944	1,986,768
$\Delta l(\mathbf{y}|\boldsymbol{\theta}_t)$	3.76%	1.5%
$l(\mathbf{y}|\boldsymbol{\theta}_T)$ range	[-196,-164]	[-188,-160]

important improvement of 70% on average compared with MCMC-SGR. Still, the reference reduced log-likelihood was actually not contained in the range of sampled reduced log-likelihoods with PT-SGR, indicating that these samples are not representative of the posterior PDF. For our ASMC-SGR 24p run, the indicator $\Delta l(\mathbf{y}|\boldsymbol{\theta}_t)$ is 3.76% and the log-likelihood range contains the reference value (Table 3.1).

Figures 3.6a-d show exemplary PT-SGR posterior samples from *Laloy et al.* (2016). These samples do not resemble either mode 1 or mode 2, even if Figure 3.6b has some structural similarities with mode 1 (Figure 3.4a). In contrast, the final states obtained by ASMC-SGR 24p (Figures 3.6e-h) recover models that resemble both reference modes: The realizations in Figures 3.6e-g resemble mode 2 (Figure 3.4b) and the one in Figure 3.6h resembles mode 1 (Figure 3.4a).

The reference mean (Figure 3.7a) is the mean of mode 1 (Figure 3.4a) and mode 2 (Figure 3.4b) of the reference model. The true posterior mean is unknown and it is likely to be slightly biased towards models resembling one of the modes. The reason for this is that even if mode 1 and 2 have the same likelihood, they do not have the same prior probability. This is a consequence of using the training image in Figure 3.3 that is likely to favour certain orientations of structures when generating prior samples. Nevertheless, Figure 3.7a provides a sensible point of comparison.

The ASMC-SGR 24p weights (Figures 3.6e-h) and the posterior mean corresponding to the weighted arithmetic mean of the samples (Figure 3.7b) suggest that the total weights given to the two “modes" is unbalanced with “mode 2" having a higher total weight than “mode 1".

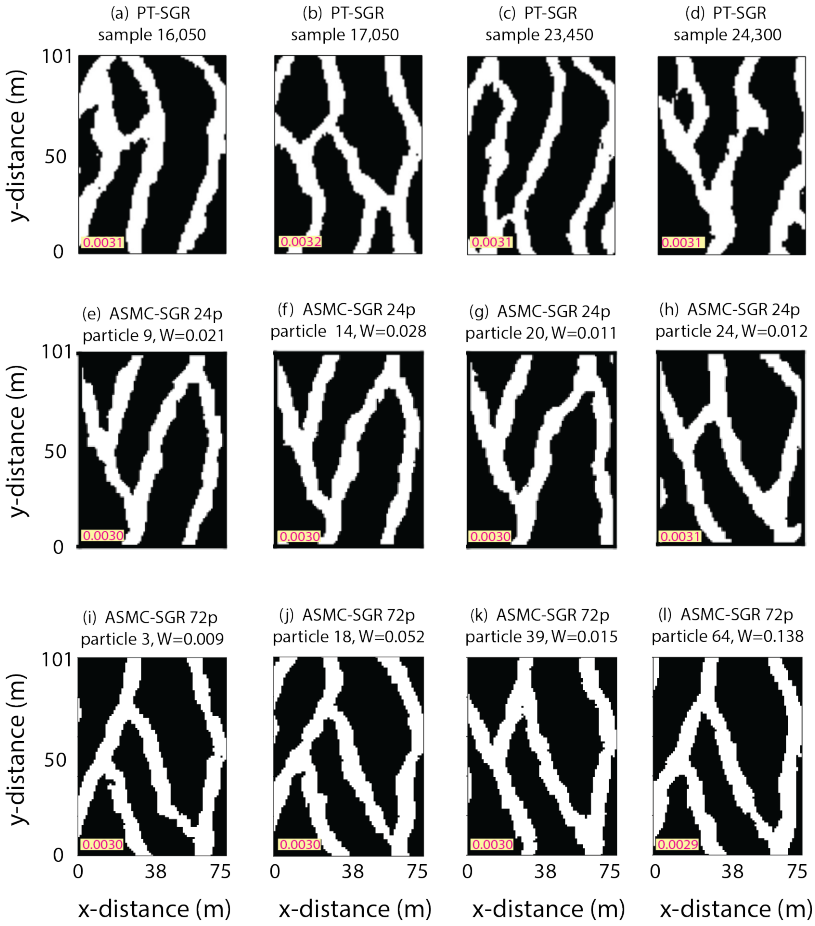

Figure 3.6: Samples from the posterior PDF obtained with: (a)-(d) PT-SGR by *Laloy et al.* (2016); (e)-(f) ASMC-SGR using 24 particles and (i)-(l) ASMC-SGR using 72 particles (the corresponding weights W of the particles are shown). The root mean square error (RMSE) without units is indicated for each sample; the corresponding value for the true model (modes 1 and 2) is 0.0030.

Still, these results show that ASMC-SGR can sample the two modes of this very challenging inverse problem and that the structures of the reference mean are partly recovered (unlike for the PT-SGR mean, see Figure 9b in *Laloy et al.* (2016)).

Test 2: ASMC-SGR with 72 particles

ASMC provides an approximation not only of the posterior PDF but also of every tempered intermediate power posterior. In this section, we focus on the evolution of the (unnormalized) power posteriors as α increases from the prior (α=0) to the posterior PDF (α=1). One way of interpreting these power posteriors is to consider them as posterior PDFs for different assumptions on the data error level. Indeed, decreasing the α-exponent has the same impact on the likelihood variable component as increasing the assumed standard deviation σ of the data noise ($\alpha \propto \frac{1}{\sigma^2}$ in Eqs. 3.3 and 3.7). Thus, the effect of tempering with a given α could also be achieved by considering an assumed standard deviation of $\sigma_\alpha = \sigma/\sqrt{\alpha}$, where σ is the original standard deviation of 0.003 kg/m^3 (for example, $\alpha = 0.25$ is analogous to assuming a standard deviation that is twice as large $\sigma_\alpha = 0.006$ kg/m^3). Since the objective in this

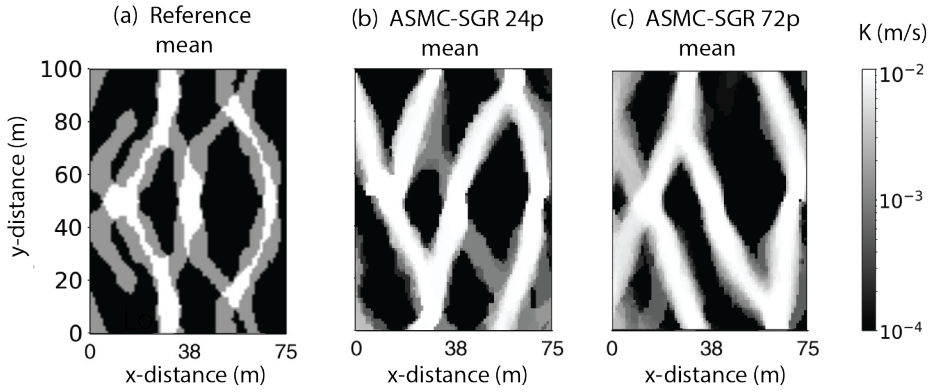

Figure 3.7: (a) Mean of the reference model's two modes; (b) ASMC-SGR 24p posterior mean and (c) ASMC-SGR 72p posterior mean obtained as a weighted mean of the final states of the particles.

section is no longer to compare the results with *Laloy et al.* (2016) for a similar computational budget, we now consider more particles. We increase the number of particles running in parallel from 24 to 72, thereby, aiming for improved approximations of the intermediate power posteriors, while keeping fixed the other user-defined parameters (ASMC-SGR 72p in Table 3.1). The number of power posteriors needed to honor the targeted $CESS_{op}$ are slightly higher compared to the ASMC-SGR 24p test. The indicator $\Delta l(\mathbf{y}|\boldsymbol{\theta})$[%] (equation 3.17) for ASMC-SGR 72p is 1.5%, that is, 60% less than for the 24 particles test. Furthermore, the likelihood range of the final particles is also reduced. The posterior mean for ASMC-SGR 72p (Figure 3.7c) and four samples from the posterior PDF (Figs. 3.6i-l) indicate that most of the samples resamble mode 1 instead of mode 2 of the reference model, that is, the opposite behavior compared with the ASMC-SGR 24p run.

The structural similarity index measure (SSIM) (*Wang et al.*, 2004) can be used to quantify the similarity between two images. It varies between -1 and 1, the higher the SSIM the more similar the two compared images are (SSIM=1 indicates identical images). The SSIM of the power posterior mean models with respect to the reference mean model (Figure 3.7a) initially increases before stagnating when α reaches 0.01 for both ASMC-SGR 24p (Figure 3.8a) and ASMC-SGR 72p (Figure 3.8b). For ASMC-SGR 24p, the SSIM values with respect to mode 2 continue to increase while the SSIM values with respect to mode 1 is even decreasing at the end of the run (Figure 3.8a). For ASMC-SGR 72p, the situation is the opposite with the SSIM values with respect to mode 1 being those that continue to increase for larger α-values (Figure 3.8b). For ASMC-SGR 24p, the SSIM remains the highest for mode 2 for all α-values above 0.001, while the SSIM values with respect to mode 2 for ASMC-SGR 72p only start to dominate for α-values above 0.1. This is a consequence of the larger number of particles and the corresponding increased ability to approximate the power posterior. The range of SSIM values between the particle realizations and the reference models are shown to decrease as the run progresses.

Figure 3.9 shows the posterior means and standard deviations at five stages of the ASMC-SGR 72p run. The mean of the prior models (Figure 3.9a) is computed from the initial DeeSse simulations using the facies at the pumping wells as conditioning data. At $\alpha = 2.0e-3$ (Figure

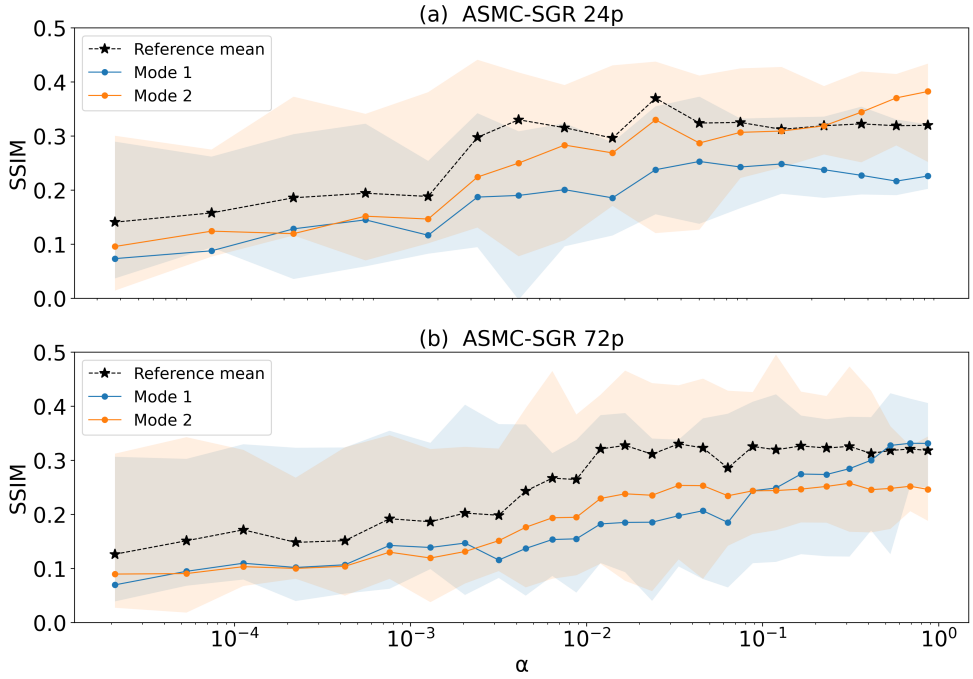

Figure 3.8: Structural similarity index measure (SSIM) of the weighted mean models for a subset of the estimated power posteriors with respect to the reference mean model (Figure 3.7a) and reference models (modes 1 and 2) (Figures 3.4a-b) vs. α. For modes 1 and 2, the range for all samples is indicated with shading. Results are shown for (a) ASMC-SGR 24p and (b) ASMC-SGR 72p.

3.9b), $\alpha = 1.7e-2$ (Figure 3.9c) and $\alpha = 8.8e-2$ (Figure 3.9d), the power posterior mean models already resembles patterns of the reference mean (Figure 3.7a). When $\alpha = 1$, the posterior mean model is dominated by mode 1 (Figure 3.9e). The standard deviations are initially high except in the vicinity of the conditioning points (Figure 3.9f) and they decrease as expectec with increasing α-values (Figures 3.9g-j) as the run evolves towards the posterior PDF. Four samples from the power posteriors corresponding to $\alpha = 2.0e-3$ (Figure 3.10e-h), $\alpha = 1.7e-2$ (Figure 3.10i-l) and $\alpha = 8.8e-2$ (Figure 3.10m-p) indicate that the variability among the realizations are high at the beginning with large corresponding RMSE values. As α increases, the variability among the realizations and the corresponding RMSE values decrease as the samples start resembling the modes and fit the data better.

Resampling and Eve indices

Resampling has the advantage of reducing the variance of the particle weights and focusing the sampling in regions of high posterior probability. However, the corresponding decrease in the variability of the sample realizations has also an adverse impact on the ASMC estimations. A conservative way of estimating the number of independent particles remaining in a run is to trace back the origin of the particles using the Eve indices. Before any resampling is perfomed, the Eve indices of the particles are $1:N$. As resampling implies re-organization and replication of particles, the Eve indices change along the run. At time t, each particle i has an Eve index E_t^i that denotes the original index of the particle that moved there (see *Lee and Whiteley* (2018) for a detailed and illustrative explanation).

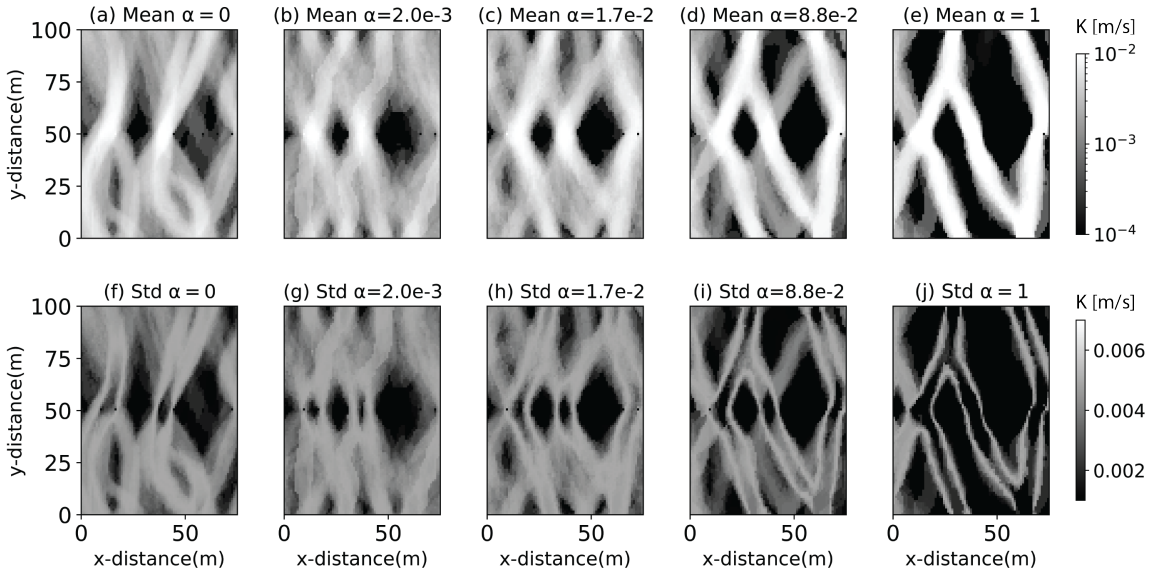

Figure 3.9: Posterior (a-e) means and (f-j) standard deviations of five different power posteriors for the ASMC-SGR 72p run: (a, f) $\alpha_t = 0$; (b, g) $\alpha_t = 0.002$; (c, h) $\alpha_t = 0.017$; (d, i) $\alpha_t = 0.088$ and (e, j) $\alpha_t = 1$.

The evolution of the Eve indices are shown for tests ASMC-SGR 24p (Figure 3.11a) and ASMC-SGR 72p (Figure 3.11b). The Eve indices are modified after each resampling step: particles with higher weights are more likely to be replicated, and as they bring their Eve indices (their origin) with them, these Eve indices are replicated as well, while other Eve indices corresponding to particle states with low weights are lost on the way. Consequently, the number of distinct Eve indices is reduced along the run due to resampling; the more resampling there is, the fewer surviving Eve indices at the end of the run. In each of our two example runs there is six resampling steps; this led to two surviving Eve indices out of 24 for ASMC-SGR 24p and only one surviving Eve index out of 72 for ASMC-SGR 72p. Of course, the particles with the same Eve indices are generally not identical as they develop independently after resampling in response to the MCMC proposal steps. Despite inherent randomness, a larger number of Eve indices are expected when reducing the number of resampling steps or increasing the number of particles. For our two test cases, the few surviving Eve indices indicate that a higher number of particles N, intermediate power posteriors or K steps would be beneficial.

Evidence estimation

The evidence $\pi(\mathbf{y})$ (equation 3.4), which can be used for Bayesian model selection and ranking, is obtained as a byproduct of the ASMC algorithm (equation 3.16). The log-evidence is shown to evolve similarly for the ASMC-SGR 24p and ASMC-SGR 72p (Figure 3.12a) runs. Both evidence curves have the same shape as α increases, and the final evidence estimates are close: $\pi(\mathbf{y}) = 1374.14$ for ASMC-SGR 72p and $\pi(\mathbf{y}) = 1371.06$ for ASMC-SGR 24p. For many model selection studies focused on conceptual model comparison, the differences in the evidence between conceptual models are often much larger (*Amaya et al.*, 2021; *Brunetti et al.*, 2017, 2019) than this discrepancy (Figure 3.12b), thereby, suggesting that only 24 particles

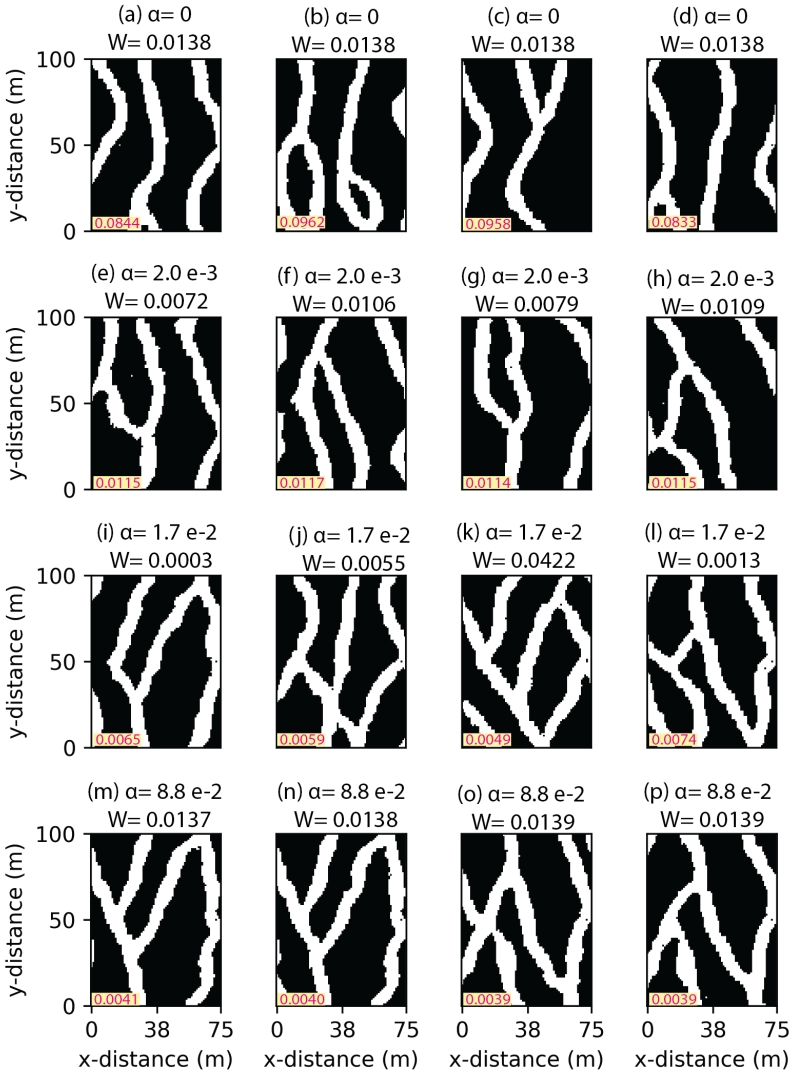

Figure 3.10: Four samples from different power posteriors sampled with the ASMC-SGR 72p run: (a)-(d) $\alpha_t = 0$; (e)-(h) $\alpha_t = 0.002$; (i)-(l) $\alpha_t = 0.017$ and (m)-(p) $\alpha_t = 0.088$. The corresponding weights W are shown and the root mean square error (RMSE) without units is indicated for each sample with the corresponding value of the reference model (modes 1 and 2) being 0.0030.

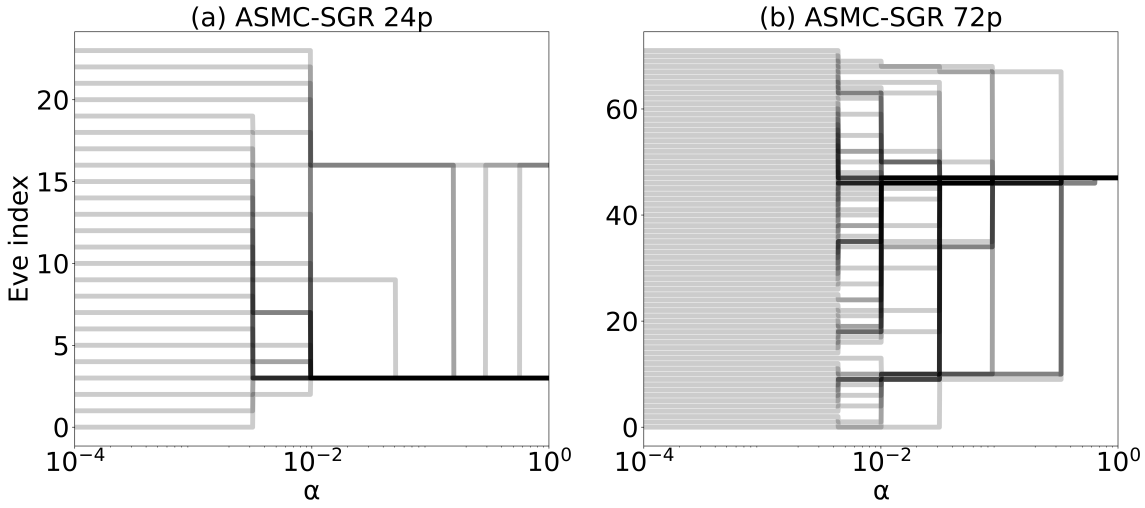

Figure 3.11: Evolution of the Eve indices for the (a) ASMC-SGR 24p and (b) ASMC-SGR 72p in the range from $\alpha = 10^{-4}$ until $\alpha = 1$. The opacity of the lines is proportional to the number of particles that have the same Eve index (same origin) at a given α.

would probably provide sufficiently accurate results. Analogous to the power posteriors, it is also possible to interpret the intermediate evidences as those corresponding to larger assumed σ-values. This necessitates a correction nevertheless, as the multiplicative term $((\sqrt{2\pi\sigma^2})^{-m_d}$ in equation 3.2) does not follow the proportionality $\alpha \propto \frac{1}{\sigma^2}$. The intermediate log-evidences $log[\pi(\mathbf{y},\alpha)]$ can be corrected to $log[\pi(\mathbf{y},\alpha)]_{corr}$ following:

$$log[\pi(\mathbf{y}),\alpha]_{corr} = log[\pi(\mathbf{y},\alpha)] + \alpha\, m_d\, log(\sqrt{2\pi}\sigma) - m_d\, log(\sqrt{2\pi}\sigma_\alpha), \tag{3.18}$$

where σ is the originally assumed standard deviation of 0.003 kg/m^3 and $\sigma_\alpha = \sigma/\sqrt{\alpha}$ is the standard deviation corresponding to that particular α. The results highlight that the estimated evidences depend very strongly on the assumed error level (Figure 3.12c).

3.4 Discussion

For a similar computational budget, ASMC-SGR has been shown to outperform PT-SGR in terms of data fitting (Table 3.1). Moreover, ASMC-SGR recover particle states (Figure 3.6e-h) that resemble both of the reference modes (Figure 3.4a-b), while none of them are recovered when using PT-SGR (Figure 3.6a-d). The ASMC algorithm adaptively tunes both the proposal scale and the α-sequence (inverse temperatures) along the run, which implies much less user effort compared to the tedious testing needed to make PT-SGR perform well. If conceptual model comparison is intended, ASMC becomes even more attractive as it provides evidence estimations (*Zhou et al.*, 2016) that are reliable and in agreement with unbiased estimations obtained using brute force Monte Carlo (*Amaya et al.*, 2021).

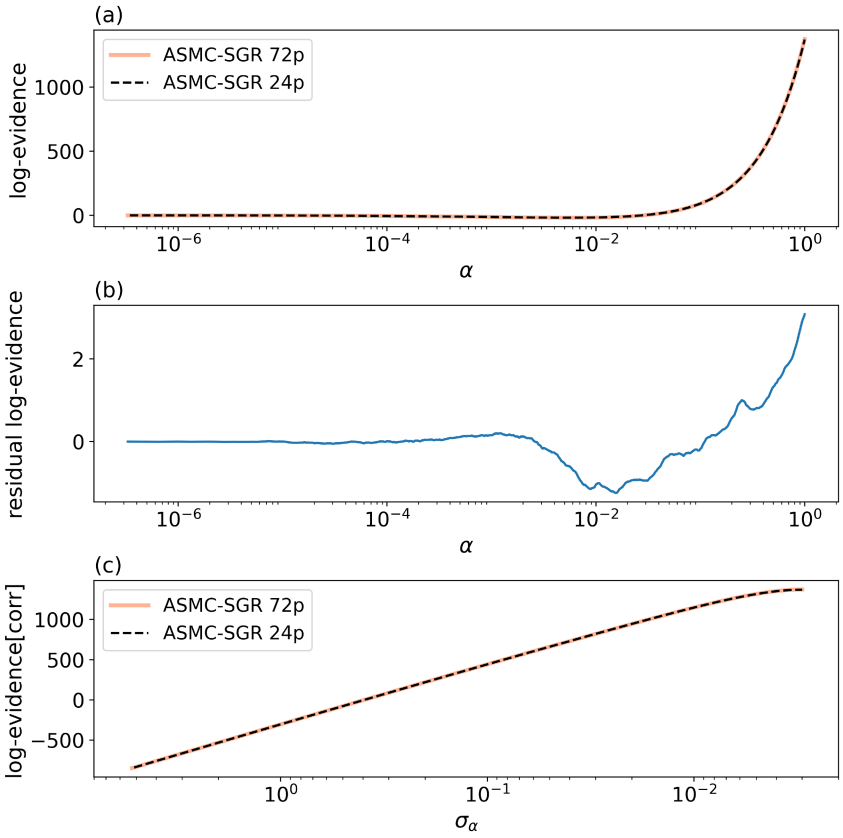

Figure 3.12: (a) Natural log-evidence evolution vs. α for ASMC-SGR using 24 and 72 particles; (b) difference of the log-evidence estimates for the two test cases vs. α; (c) corrected natural log-evidence (equation 3.18) vs. different assumptions of σ_α.

The intermediate power posterior approximations offered by the ASMC algorithm are highly instructive (Figures 3.9-3.10). When the ASMC-SGR algorithm starts considering α-values above a given threshold (lower for 24 particles than for 72 particles), the sampling tends to become unbalanced in our two example runs and there is one mode that ends up having a higher posterior probability than the other. This could be addressed by increasing the computational budget: either by considering a much larger number of particles (one could imagine using hundreds or thousands of particles), or by increasing $CESS_{op}$ or K that would reduce the number of resampling steps. Resampling plays the important role in particle methods of focusing the sampling towards high-probability regions by controlling the variance of the particle weights. Unfortunately, this advantage comes at the expense of losing the independence between the particles leading, in our case, to over-prediction of one of the posterior modes. In our test example, it would be straightforward to facilitate sampling of both modes simply by allowing for model proposals that would mirror the present state. However, this would not be possible in most realistic settings.

The test example was primarily designed to ensure that the posterior had two posterior modes located far from each other, thereby, enabling comparison of different probabilistic methods for a very challenging inverse problem. In order to allow a fair comparison between the previously published PT results and the new ASMC results, the training image and the DeeSse simulation parameters were the same as in *Laloy et al.* (2016). However, this implies that the prior probability of sampling modes 1 and 2 are different, and consequently that the two posterior modes have unequal posterior probabilities despite that the likelihoods are equivalent. To ensure that the true posterior has two modes of equal posterior probability, one could use a training image with two layers. The first layer would be the original training image and the second layer would be obtained by mirroring the training image similarly to how mode 2 was created. At each SGR step, the MPS algorithm would scan from either layer 1 or 2. Nevertheless, the fact that ASMC-SGR 24p primarily sampled mode 2 and ASMC-SGR 72p primarily sampled mode 1 suggests that the main limitation in the presented runs are the limited computational budgets that prohibit sampling the two posterior modes well during one ASMC run.

After publication of this study, we performed a test in which we increased the computational budget. We doubled the number of intermediate Markov steps in ASMC-SGR 72p, from $K = 18$ to $K = 36$. The results (Appendix A) show an important improvement both in the data fitting and in the balance of the two modes sampling. One option to reduce the computational time and, thereby, allow for longer runs would be to use faster algorithms for generating the candidate models: either newer versions of DeeSse, quick sampling (*Gravey and Mariethoz*, 2020), graph cuts (*Zahner et al.*, 2016), or by replacing MPS-based algorithms with deep learning-based generators as in the study by *Amaya et al.* (2021). Also, a computational gain could be achieved by replacing the expensive forward solver with a surrogate (e.g., by polynomial chaos expansion (*Laloy et al.*, 2013; *Meles et al.*, 2022)). This should not bias the results if the surrogate is only applied in the intermediate K Markov steps, while still using the expensive forward solver for the importance sampling steps.

The power posterior approximations can also be interpreted as posterior PDF approximations for different assumed data error levels (Figure 3.9). By raising the likelihood function to an inverse temperature α that is less that 1, the impact on the reduced log-likelihood is the

same as if increasing the assumed error level. That is, flattening the likelihood and, thereby, enhancing the freedom of the exploration. A similar effect is obtained by decreasing the number of data points considered m_d ($\alpha \propto m_d$ in equation 3.3): keeping a subset of the original observations will have a similar impact as reducing α or increasing σ^2. Tempering, assuming artificially high data errors or reducing the number of data are not uncommon in the literature when addressing challenging Bayesian inversions (e.g., *Juda and Renard* (2021)). This results in an easier to solve, but different, inverse problem that is conservative in the sense that the posterior mean is less informative and the posterior variance is larger than for the original problem. One important advantage of ASMC is that it explores all these intermediate problems, but also use the information gained to sample the original posterior PDF that is unfeasible for many other methods. Similarly, the evidence computations can be re-scaled to correspond to different assumptions of data error levels (Figure 3.12).

In field applications, the data error level is typically poorly known. ASMC can then be very helpful, as one could assume a noise level that is likely too low and then obtain approximations of several power posterior corresponding to different (larger) error assumptions. One could then consider choosing an optimal error level based on the ASMC intermediate results using the relationship between α and σ. For instance, one could perhaps choose the error level and the corresponding posterior (and evidence) approximations by considering the divergence between the reference target log-likelihood and the tempered log-likelihoods with increasing α. In Figure 3.2a there is no such divergence as the true data error level is assumed. This would be much more efficient than running multiple MCMC runs with different assumptions concerning σ.

An alternative and somewhat related method to solve inverse problems with SGR is Population Expansion (PoPEx) introduced by *Jäggli et al.* (2017, 2018). This method is similar to ASMC in the sense that the proposal distribution progressively evolves along the run towards the posterior PDF. These evolving distributions provide information maps built to efficiently select conditioning data for new SGR model proposals based on previously sampled high-likelihood models. The posterior PDF is approximated by iteratively expanding the set of models along the run. The corrected PoPEx algorithm by *Jäggli et al.* (2018) can be interpreted as an adaptive importance sampling algorithm (*Naylor and Smith*, 1988), in which the evolving proposal distribution is the importance distribution and the posterior PDF is the target distribution. This is different from ASMC where the importance sampling relies on consecutive power posteriors. Compared with PoPEx, ASMC also includes resampling steps, thereby, avoiding the degeneracy that often seems to plague PoPEx. To address this problem, *Jäggli et al.* (2018) artificially reweigh the weights in order to achieve a lower variance and, hence, a richer representation of the approximated posterior.

3.5 Conclusions

Tempering of likelihood functions is used in a wide variety of Bayesian methods to enhance posterior exploration and for evidence computations, particularly when confronted with high-dimensional and multimodal posterior PDFs that standard MCMC methods often struggle with. We demonstrate that adaptive sequential Monte Carlo (ASMC) outperforms parallel tempering (PT) when using sequential geostatistical resampling (a multiple-point statistics approach) as model proposal scheme in the context of a challenging synthetic groundwater transport inverse problem involving 7,500 model parameters with a bimodal posterior PDF. ASMC is found to be considerably more effective in locating the two posterior modes and to sample states with likelihoods that are in agreement with the data noise. The algorithm has a simple implementation and demands a minimal user effort in terms of tuning due to its adaptive features. Furthermore, it also estimates the evidence (marginal likelihood) at almost no additional computational cost. The intermediate results of the algorithm can be used to determine the posterior means, standard deviations and evidences corresponding to different assumptions of data errors. This can be very helpful as it avoids pre-defining one standard deviation on the noise (or doing many MCMC runs with different assumed errors) and it allows assessing how the posterior changes from the prior through a number of intermediate power posteriors to the targeted posterior PDF. The method is versatile, robust and very well suited for parallelization and could have wide applicability to solve inverse problems arising in the field of water resources using a wide range of model parameterizations, forward solvers and model proposal schemes. In the future, we will seek speed-ups through surrogate modeling to enable a larger number of particles or longer runs and, thereby, improve the posterior estimations further for a given computational cost. Indeed, our examples with 24 and 72 particles could locate the posterior modes, but the computational budgets were insufficient to robustly sample the two posterior modes during the same ASMC run.

3.6 Appendix A

ASMC-SGR test with increased number of Markov steps

We perform a third test doubling the number of K intermediate Markov steps approximating the power posteriors in ASMC-SGR 72p ($K = 36$ instead of $K = 18$), and keeping the rest of the ASMC parameters fixed. We refer to this test as ASMC-SGR 72p (K=36). The results show a strong improvement on the data fitting, the indicator $\Delta l(\mathbf{y}|\boldsymbol{\theta}_t)$ decreases from 1.5% ($K = 18$) to 0.1% ($K = 36$). The upper row in Figure 3.13 is an extension of Figure 3.7 with the additional Figure 3.13d showing the posterior mean of ASMC-SGR 72p (K=36). The patterns in Figure 3.13d are structurally more similar to the reference mean compared to Figure 3.13b and c. Figure 3.13e shows the SSIM evolution for ASMC-SGR 72p (K=36), as in Figure 3.8 for the K=18 tests. The orange and blue curves in Figure 3.13e indicate a more balanced sampling of the individual modes throughout the inversion compared to Figure 3.8a and b.

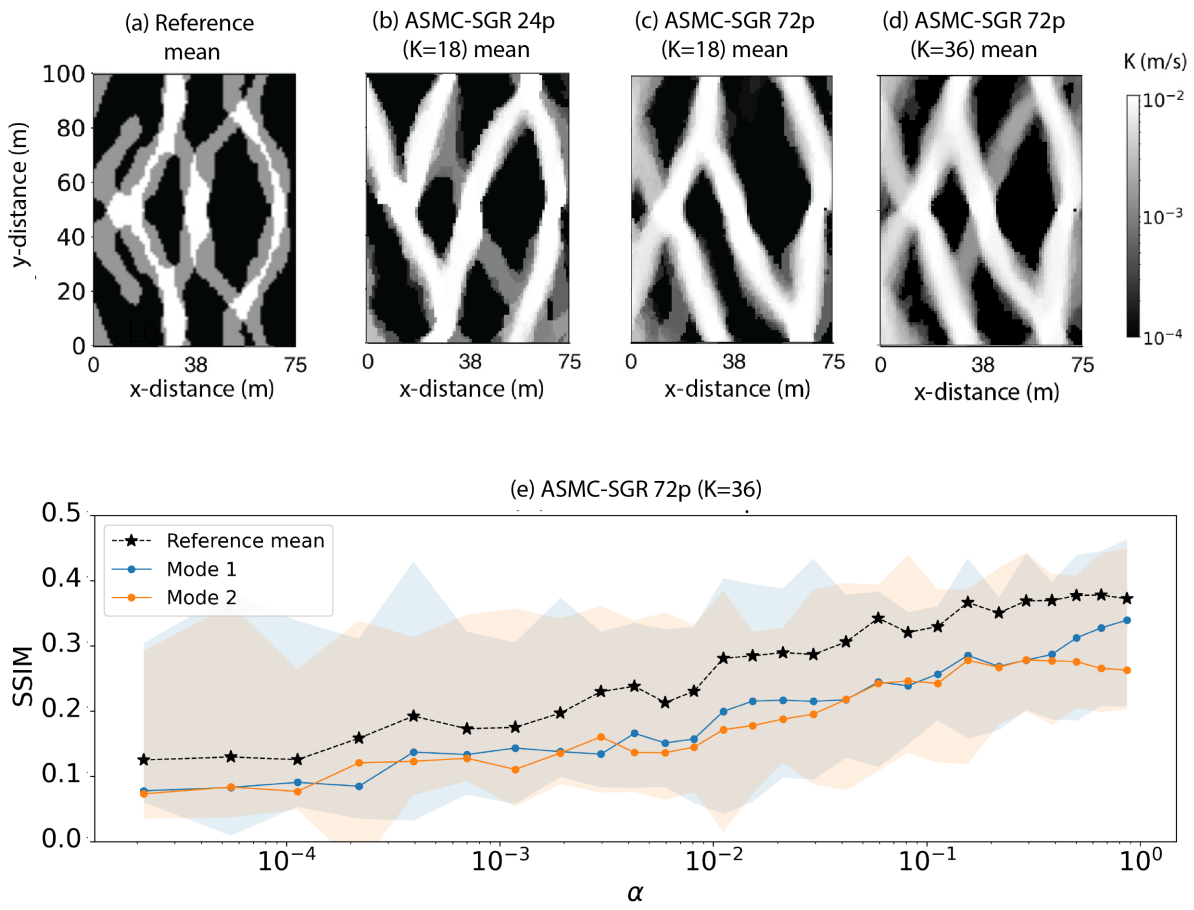

Figure 3.13: (a) Mean of the reference model's two modes; (b) ASMC-SGR 24p posterior mean using K=18, (c) ASMC-SGR 72p posterior mean using K=18, and (d) ASMC-SGR 72p posterior mean using K=36, obtained as a weighted mean of the final states of the particles. Subfigure (e) shows the structural similarity index measure (SSIM) of the weighted mean models for a subset of the estimated power posteriors with respect to the reference mean model and reference models (modes 1 and 2) vs. α. For modes 1 and 2, the range for all samples is indicated with shading.

Chapter 4

Multifidelity adaptive sequential Monte Carlo applied to geophysical inversion

---

Macarena Amaya, Giovanni Meles, Stefano Marelli and Niklas Linde,

Manuscript submitted to *Geophysical Journal International*.

Abstract

In the context of Bayesian inversion, Markov chain Monte Carlo (MCMC) methods are often used to estimate posterior probability density functions (PDFs). Standard MCMC methods can be inefficient for strongly non-linear, high-dimensional problems, with the chains getting trapped in local minima and failing to properly sample the posterior PDF. Some methods seek to overcome these issues by enhancing the ease of exploration through tempering, which re-scales the target function by reducing the weight of the likelihood function. We consider sequential Monte Carlo (SMC) methods that provide an approximation of the posterior PDF 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 sequential Monte Carlo (ASMC) allows tuning the temperature increments on-the-go. Another recurrent 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 modeling within ASMC and introduce an approach involving surrogate modeling only, in which either prior samples are used to train the surrogate, or the surrogate model is re-trained by updating the training set during the inversion. In our implementation, we rely on polynomial chaos expansions for surrogate modeling, 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 solvers. We find that the method based on re-training the surrogate during the inversion outperforms the results obtained when only considering prior samples. We then introduce a computationally more expensive multifidelity framework including a transition to the high-fidelity forward solver at the end of the ASMC run leading to even more accurate results. Both methods result in speed-ups that are larger than one order of magnitude compared to performing standard high-fidelity ASMC inversion.

4.1 Introduction

Bayesian inversion provides a comprehensive characterization of the uncertainty associated with inferred model parameters (*Sambridge and 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 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 and 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). Annealed importance sampling (AIS, *Neal* (2001)) and sequential Monte Carlo (SMC, *Doucet and Johansen* (2011)) are 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 decreasing temperatures. The difference between the two methods is that SMC incorporates resampling of the particle population when the variance of the importance weights becomes high. An important advantage of both methods is that they also allow estimating the evidence, the normalizing constant in Bayes' theorem and a crucial parameter in Bayesian model selection (*Kass and Raftery*, 1995; *Schöniger et al.*, 2014). 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 than state-of-the-art Markov chain Monte Carlo (MCMC) methods (*Amaya et al.*, 2021; *Davies et al.*, 2023). Furthermore, ASMC has shown a higher capacity to account for multi-modality than parallel tempering (*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. The term high-fidelity modeling 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. Surrogate modeling 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). Polynomial chaos expansion (PCE) is a type of surrogate that approximates models by their spectral representations in a suitable basis of polynomial functions (*Xiu and Karniadakis*, 2002; *Blatman and Sudret*, 2011), and has been successfully implemented within various Bayesian inversion frameworks (*Marzouk and Xiu*, 2009; *Marzouk et al.*, 2007; *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). The concept of multifidelity modeling was introduced in uncertainty quantification by *Ng and Eldred* (2012) and has been applied to different surrogate modeling frameworks (*Park et al.*, 2017). We use the term multifidelity inversion to refer to Bayesian inversion strategies that accelerate the computation 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 and Fox*, 2005). In this work, we propose an ASMC algorithm that operates within such a multifidelity framework. The use of evolving forward 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 increasing the spatial resolution considered in forward modeling and ASMC tempering was proposed by *Koutsourelakis* (2009), a scheme further automatized by *Latz et al.* (2018).

Our ASMC multifidelity inversion method relies on transitions to re-trained surrogate models and ultimately to the high-fidelity solver. The method is initialized with a low-fidelity solver and updates the surrogate as the inversion progresses. 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. 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 modified from *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 solver. We demonstrate that the multifidelity ASMC inversion with progressive surrogate updates can successfully reproduce the results obtained by traditional high-fidelity inversion, but at a fraction of its computational cost.

4.2 Method

4.2.1 Bayesian inference and model selection

A probabilistic formulation of the inverse problem is offered by Bayes' theorem, which expresses the posterior probability density function (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 $\mathbf{y}$:

$$\pi(\boldsymbol{\xi}|\mathbf{y}) = \frac{\pi(\boldsymbol{\xi})\, p(\mathbf{y}|\boldsymbol{\xi})}{\pi(\mathbf{y})}. \tag{4.1}$$

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 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 under consideration. A conceptual model refers here to the parameterization of the subsurface together with its prior PDF (*Brunetti et al.*, 2017). The evidence is needed to compare or rank different conceptual models (*Kass and Raftery*, 1995). However, this quantity can be challenging to compute for high-dimensional problems, as it is a multi-dimensional integral over the prior (*Schöniger et al.*, 2014):

$$\pi(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\xi})\pi(\boldsymbol{\xi})\, d\boldsymbol{\xi}. \tag{4.2}$$

Model parametrization and model reduction by principal components

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

$$\mathbf{y} = \mathscr{F}(g(\boldsymbol{\theta}_{full})) + \epsilon_y, \tag{4.3}$$

where $\mathbf{y}$ represents the observed data corresponding to the set of parameters $\boldsymbol{\theta}_{full}$ (e.g., a full decomposition by principal component analysis (PCA, *Jolliffe and Cadima* (2016)) as used in the present study) and ϵ_y the observational noise. The forward operator $\mathscr{F}(g(\boldsymbol{\theta}_{full}))$ can be expressed as a function composition $\mathscr{M} = \mathscr{F} \circ g$. In practice, one achieve dimensionality-reduction by using a truncated subset of, for example, PCA coordinates. If the function composition over the truncated set θ adequately approximate the forward operator, then:

$$\mathbf{y} = \mathscr{M}(\boldsymbol{\theta}) + \epsilon_y + \epsilon_{PCA}, \tag{4.4}$$

where ϵ_{PCA} is the error in the computed forward response associated with the PCA projection. In this paper 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).

4.2.2 Surrogate modeling

Surrogate models $\hat{\mathscr{M}}(\boldsymbol{\theta})$ 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*):

$$\hat{\mathscr{M}}(\boldsymbol{\theta}) \approx \mathscr{M}(\boldsymbol{\theta}). \tag{4.5}$$

Common examples of surrogate models used in applied sciences include Kriging, or Gaussian process modeling (*Santner et al.*, 2003), and polynomial chaos expansions (*Xiu and Karniadakis*, 2002; *Lüthen et al.*, 2022).

Polynomial chaos expansions

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 and Karniadakis*, 2002):

$$\mathcal{M}(\boldsymbol{\theta}) = \sum_{b \in \mathcal{N}^{\mathcal{M}}} a_b \psi_b(\boldsymbol{\theta}), \tag{4.6}$$

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:

$$\mathbf{y} = \hat{\mathcal{M}}(\boldsymbol{\theta}) + \epsilon_y + \epsilon_{PCA} + \epsilon_{PCE}, \tag{4.7}$$

where ϵ_{PCE} is the error associated with the surrogate modeling.

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 modeling error assumed to be zero. In this paper, we rely on PCE surrogates trained using the Matlab-based package UQLab (*Marelli and 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).

Accounting for errors in the likelihood function

A Bayesian inversion that incorporates dimensionality-reduction and surrogate modeling needs to consider the errors induced by these simplifications. If we assume the various error sources to be normally distributed, the corresponding likelihood function (c.f., equation 4.1) can be expressed as (*Tarantola*, 2005):

$$p(\mathbf{y}|\boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |C|^{-\frac{1}{2}} exp\left(-\frac{1}{2}(\hat{\mathcal{M}}(\boldsymbol{\theta}) - \mathbf{d_d} - \mathbf{y})^T C^{-1}(\hat{\mathcal{M}}(\boldsymbol{\theta}) - \mathbf{d_d} - \mathbf{y})\right), \tag{4.8}$$

where n is the number of data points, $\mathbf{y}$ the observed data, C the covariance matrix and $|C|$ its determinant. The $\mathbf{d_d}$ variable accounts for the modeling 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 modeling as well:

$$C = C_y + C_{PCA} + C_{PCE}, \tag{4.9}$$

where C_y is the data covariance matrix, C_{PCA} the covariance of the PCA projection error and C_{PCE} the covariance of the surrogate modeling error. Both C_{PCA} and C_{PCE} can be computed as:

$$C_{PCA/PCE} = \frac{1}{N} D_m D_m^T, \tag{4.10}$$

where D_m is a misfit matrix. In the case of C_{PCA}, 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 C_{PCE}, it is convenient in practice to derive D_m from the PCE training set using cross-validation, to avoid having to compute additional expensive high-fidelity responses. The j-th D_m column can be estimated with the output of a surrogate that is trained on all the input samples excepted the j-th (*Blatman*, 2009), we refer to this C_{PCE} as the leave-one-out (L_{OO}) covariance.

4.2.3 ASMC inversion

We rely on the ASMC approach introduced by *Zhou et al.* (2016), in which N particles evolve in parallel towards an approximation of the posterior PDF. The unnormalized 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 α_t:

$$\gamma_t(\boldsymbol{\theta}_t|\mathbf{y}) \equiv \pi(\boldsymbol{\theta}_t)\, p(\mathbf{y}|\boldsymbol{\theta}_t)^{\alpha_t}. \tag{4.11}$$

The ASMC 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 α_t) using an importance sampling step (IS, *Hammersley and 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:

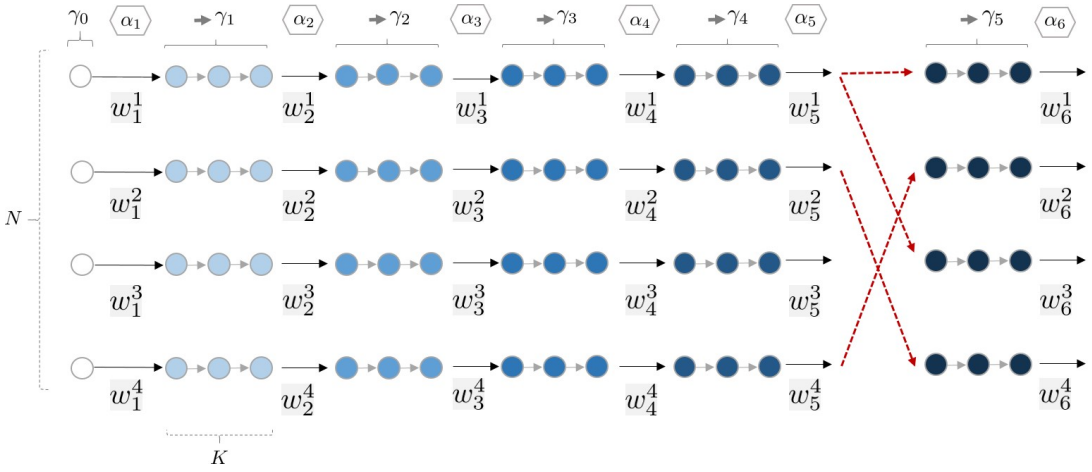

Figure 4.1: Schematic representation of ASMC for $N = 4$ particles. The circles represent the particles evolving states with the initial samples drawn from the prior PDF (white circles) and $K = 3$ MCMC steps are used to approximate the power posteriors γ_t with increasing inverse temperatures α_t. The α_t-increments are determined adaptively on-the-go. 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.

$$w_t^i = \frac{\gamma_t(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}{\gamma_{t-1}(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}, \tag{4.12}$$

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

$$W_t^i = \frac{W_{t-1}^i w_t^i}{\sum_{j=1}^N W_{t-1}^j w_t^j}, \tag{4.13}$$

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.

The ASMC method aims at building a suitable 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 equation 4.12, can be controlled. Figure 4.1 shows a graphical representation of the ASMC method using $N = 4$ particles.

Adaptive power posterior sequence

Pre-selecting a suitable sequence of α_t is very challenging. *Zhou et al.* (2016) propose a method that automatically defines the α_t-increments on-the-go based on the conditional effective sample size (CESS):

$$CESS = N\frac{(\sum_{i=1}^{N} W_{t-1}^{i} w_t^i)^2}{\sum_{j=1}^{N} W_{t-1}^{j} (w_t^j)^2}. \tag{4.14}$$

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 ($CESS_{op}$). The $CESS_{op}$ parameter controls the quality of the importance sampling steps and, hence, the quality of the resulting particle approximation. The $CESS_{op}/N$ 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 $CESS_{op}$ is non-linear, it is challenging to recommend $CESS_{op}$ values. Results by *Amaya et al.* (2021) suggest that $CESS_{op}/N >= 0.99$ provides a suitably smooth sequence.

Resampling

Sequential Monte Carlo (SMC) methods (*Doucet and Lee*, 2018), and hence ASMC, uses resampling. It refers to a re-organization of the particle states with replications that are proportional to the particle weights (red-dashed lines in Figure 4.1). 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 a 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 exploration of particles with significant posterior probabilities. However, the resampling process increases the variance of the SMC estimates unnecessarily if it is performed too often (*Douc and Cappé*, 2005). To make decisions on when to perform resampling, the effective sample (ESS) size is monitored throughout the run:

$$ESS_t = \frac{(\sum_{i=1}^{N} W_{t-1}^{i} w_t^i)^2}{\sum_{h=1}^{N} (W_{t-1}^{h})^2 (w_t^h)^2}. \tag{4.15}$$

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 and Johansen*, 2011). A derivation of equations 4.14 and 4.15 can be found in *Zhou et al.* (2016).

Evidence estimation

The normalizing constant in Bayes' theorem (equation 4.1) is the key quantity in Bayesian model selection and it is obtained as a byproduct of the ASMC method. *Del Moral et al.* (2006) show that the ratio of the normalizing constants of two consecutive power posteriors, $Z_t = \int \gamma_t(\boldsymbol{\theta}_t|\mathbf{y})d\boldsymbol{\theta}_t$ and $Z_{t-1} = \int \gamma_{t-1}(\boldsymbol{\theta}_{t-1}|\mathbf{y})d\boldsymbol{\theta}_{t-1}$, can be approximated as:

$$\frac{Z_t}{Z_{t-1}} \approx \sum_{i=1}^{N} W_{t-1}^i w_t^i. \tag{4.16}$$

As the prior PDF integrates to one ($Z_0 = 1$), the evidence (equation 4.2) can be expressed as the product of consecutive normalizing constant ratios and, hence, approximated using the evolving particle weights as:

$$\pi(\mathbf{y}) = Z_T = \frac{Z_T}{Z_0} = \prod_{t=1}^{T} \frac{Z_t}{Z_{t-1}} \approx \prod_{t=1}^{T} \sum_{i=1}^{N} W_{t-1}^i {w_t}^i. \tag{4.17}$$

Adaptive proposal scale

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

$$\Gamma(\boldsymbol{\theta}_p, \boldsymbol{\theta}_c) = min\left(1, \frac{\pi(\boldsymbol{\theta}_p)p(\mathbf{y}|\boldsymbol{\theta}_p)}{\pi(\boldsymbol{\theta}_c)p(\mathbf{y}|\boldsymbol{\theta}_c)}\right). \tag{4.18}$$

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 ϕ determining the proposal scale. An important advantage of ASMC is that ϕ can be tuned throughout the run to ensure that the acceptance rate AR is kept within a suitable range without violating detailed balance conditions, as would be the case for MCMC algorithms.

4.2.4 ASMC with surrogate updates

Performing ASMC inversion with surrogate modeling offers the opportunity to enhance the exploration, compared to standard MCMC, while also significantly reducing computational times when using ASMC with high-fidelity solvers only. However, when a high-fidelity forward solver is replaced with a low-fidelity solver, the resulting posterior PDF is different and this is to be avoided. As the inverse temperature reduces the influence of the likelihood, the 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 to direct the particles towards the posterior. 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. In this way, we aim to decrease the difference between the posterior PDF-estimates obtained with high-fidelity forward modeling and those obtained by surrogate modeling.

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 $\hat{\mathscr{M}}^{s_j}(\boldsymbol{\theta})$ and the covariance accounting for the three error sources in equation 4.10. The covariances associated with the data noise C_y and the PCA projection C_{PCA} are then constant throughout the inversion, while the covariance associated to the PCE C_{PCE} is re-calculated 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. In this case, the IS step is performed to account for the change in the forward operator and C_{PCE}. If the surrogate s_j with covariance $C_{PCE}^{s_j}$ is replaced with the updated surrogate s_{j+1} with covariance $C_{PCE}^{s_{j+1}}$, then the incremental weight resulting from the IS step will be:

$$w[s_j, s_{j+1}]_t^i = \frac{\gamma_t^{s_{j+1}}(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}{\gamma_{t-1}^{s_j}(\boldsymbol{\theta}_{t-1}^i|\mathbf{y})}. \tag{4.19}$$

Similarly to before, the α_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 equation 4.19 contribute to the particle importance weights and evidence in the same way as the standard ASMC incremental weights in equation 4.12. In contrast to those IS steps, we have often that α_t is smaller than α_{t-1}, because there is likely a reduction of the modeling 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 equation 5). If we consider an ASMC inversion in which the surrogate s_0 is unchanged, the sequence of incremental weight ratios (equation 4.12) for a particle is:

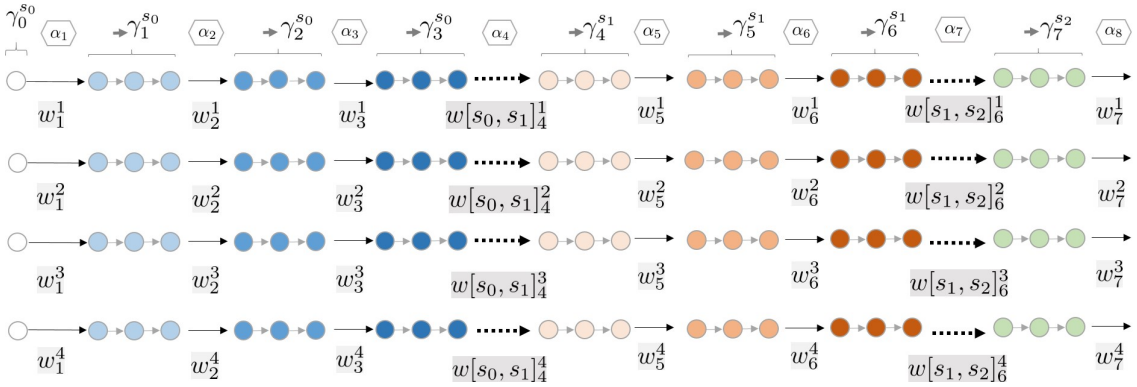

Figure 4.2: Modified schematic diagram of ASMC (Figure 4.1) with surrogate updating (ASMC-SURR). The black-dashed lines represent a re-training of the surrogate together with the corresponding importance sampling step to transition from one surrogate to the following one. The change in the color denotes a change of surrogate, whereas the change of the color tone denotes a change of temperature when keeping the surrogate solver fixed.

$$\prod_{t=1}^{T} w_t = \frac{\gamma_1^{s_0}(\boldsymbol{\theta}_0|\mathbf{y})}{\gamma_0^{s_0}(\boldsymbol{\theta}_0|\mathbf{y})} \frac{\gamma_2^{s_0}(\boldsymbol{\theta}_1|\mathbf{y})}{\gamma_1^{s_0}(\boldsymbol{\theta}_1|\mathbf{y})} \frac{\gamma_3^{s_0}(\boldsymbol{\theta}_2|\mathbf{y})}{\gamma_2^{s_0}(\boldsymbol{\theta}_2|\mathbf{y})} \frac{\gamma_4^{s_0}(\boldsymbol{\theta}_3|\mathbf{y})}{\gamma_3^{s_0}(\boldsymbol{\theta}_3|\mathbf{y})} \frac{\gamma_5^{s_0}(\boldsymbol{\theta}_4|\mathbf{y})}{\gamma_4^{s_0}(\boldsymbol{\theta}_4|\mathbf{y})} \frac{\gamma_6^{s_0}(\boldsymbol{\theta}_5|\mathbf{y})}{\gamma_5^{s_0}(\boldsymbol{\theta}_5|\mathbf{y})} \dots \frac{\gamma_T^{s_0}(\boldsymbol{\theta}_{T-1}|\mathbf{y})}{\gamma_{T-1}^{s_0}(\boldsymbol{\theta}_{T-1}|\mathbf{y})}, \quad (4.20)$$

where we have removed the particle index i to simplify the notation. Equation 4.20 illustrates the bridge created by the IS steps; the sequence starts with the prior γ_0 as an importance distribution and ends with the posterior γ_T as the target distribution. When considering surrogate updates, the sequence will also include the corresponding IS steps, for example:

$$\prod_{t=1}^{T} w_t = \frac{\gamma_1^{s_0}(\boldsymbol{\theta}_0|\mathbf{y})}{\gamma_0^{s_0}(\boldsymbol{\theta}_0|\mathbf{y})} \frac{\gamma_2^{s_0}(\boldsymbol{\theta}_1|\mathbf{y})}{\gamma_1^{s_0}(\boldsymbol{\theta}_1|\mathbf{y})} \frac{\gamma_3^{s_1}(\boldsymbol{\theta}_2|\mathbf{y})}{\gamma_2^{s_0}(\boldsymbol{\theta}_2|\mathbf{y})} \frac{\gamma_4^{s_1}(\boldsymbol{\theta}_3|\mathbf{y})}{\gamma_3^{s_1}(\boldsymbol{\theta}_4|\mathbf{y})} \frac{\gamma_5^{s_1}(\boldsymbol{\theta}_4|\mathbf{y})}{\gamma_4^{s_1}(\boldsymbol{\theta}_4|\mathbf{y})} \frac{\gamma_6^{s_2}(\boldsymbol{\theta}_5|\mathbf{y})}{\gamma_5^{s_1}(\boldsymbol{\theta}_5|\mathbf{y})} \dots \frac{\gamma_T^{s_J}(\boldsymbol{\theta}_{T-1}|\mathbf{y})}{\gamma_{T-1}^{s_J}(\boldsymbol{\theta}_{T-1}|\mathbf{y})}, \quad (4.21)$$

where the third and sixth ratio are associated with the transition between different surrogates (equation 4.19). To update the surrogate, we collect samples of the high-fidelity response as the ASMC sequence progresses. These samples are used to re-train the PCE surrogate after a pre-defined number of inverse temperature updates T_{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 Figure 4.2; we refer to this method as ASMC-SURR.

4.2.5 Multifidelity modeling 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:

$$w[s_J, HF]_t^i = \frac{\gamma_t^{HF}(\boldsymbol{\theta}_{t-1}|\mathbf{y})}{\gamma_{t-1}^{s_J}(\boldsymbol{\theta}_{t-1}|\mathbf{y})}, \tag{4.22}$$

where s_J is the last surrogate of the sequence. In this step the inverse temperature α_{t-1}, corresponding to the power posterior in the denominator $\gamma_{t-1}^{s_J}$ is equal to 1. The inverse temperature α_t associated to γ_t^{HF} can again be optimized. We seek an optimal value from $\alpha_t = F_{HF}\alpha_{t-1}$ with $F_{HF} \in [0.1, 1]$, such that it provides the highest possible $CESS$. We refer to F_{HF} as the α-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 Figure 4.3. A detailed description of the two algorithms is given in Appendix A.

4.2.6 Performance assessment

In our considered test cases, performing one iteration using surrogate solvers compared to using the high-fidelity solver implies a gain in computational time of 10 to 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 scores is preferred. A kernel density estimate of the posterior samples is used to approximate the posterior univariate distributions; we employ 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 being the highest similarity). We consider the SSIM with respect to the posterior mean model of the full high-fidelity inversion.

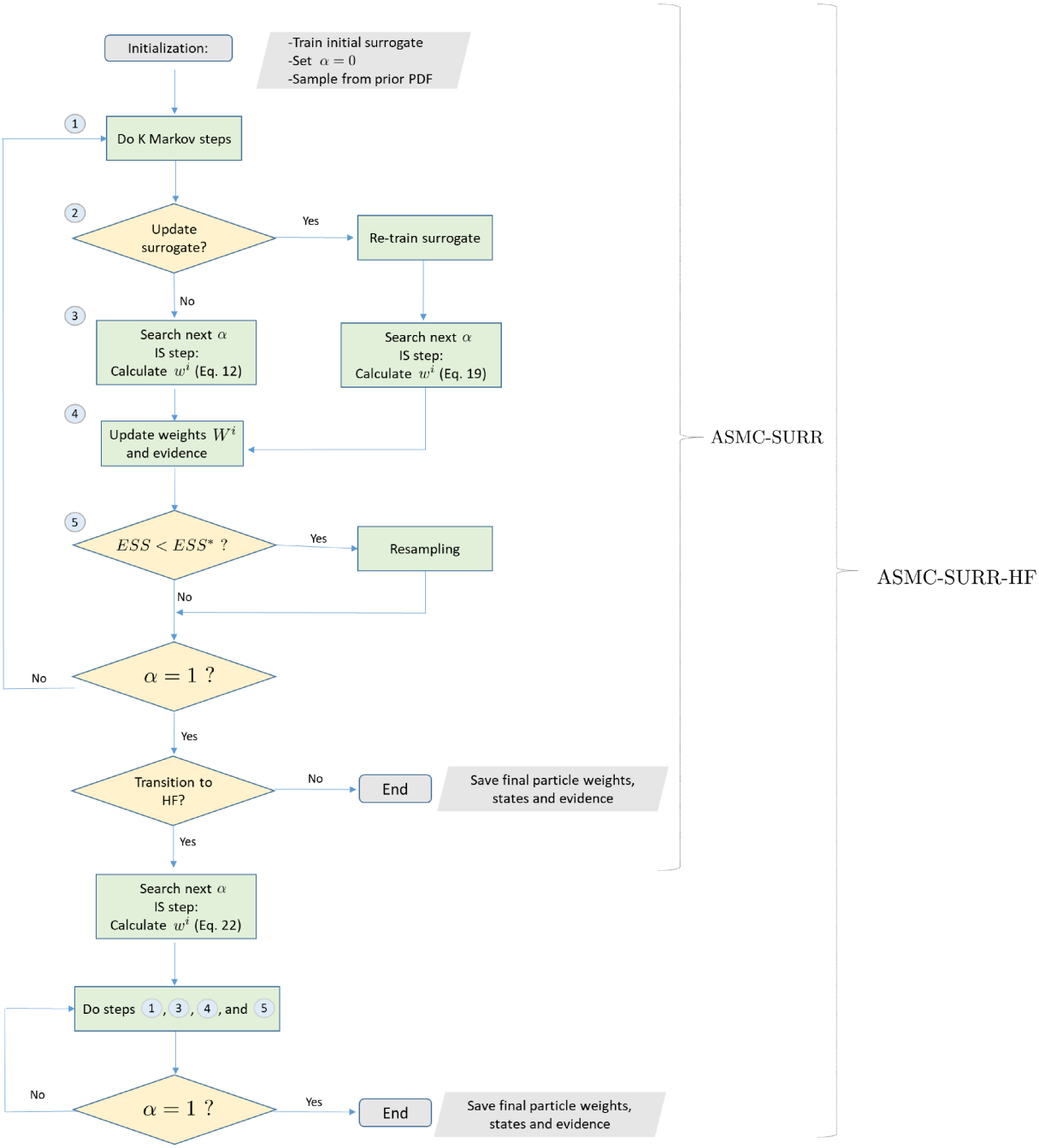

Figure 4.3: Simplified flow chart representing the main steps to perform ASMC inversion with surrogate updates with (ASMC-SURR-HF) and without (ASMC-SURR) transitioning to the high-fidelity solver. After the initialization, the sampling of the adaptive sequence of power posteriors at gradually increasing α_s starts, with importance sampling steps performed between consecutive power posteriors (see section 4.2.3). The surrogate can be updated followed by an importance sampling step to transition between surrogates (see section 4.2.4). If the ESS falls below the threshold ESS^*, resampling is performed (see section 4.2.3). The particle normalized weights and evidence (see section 4.2.3) 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 4.2.5).

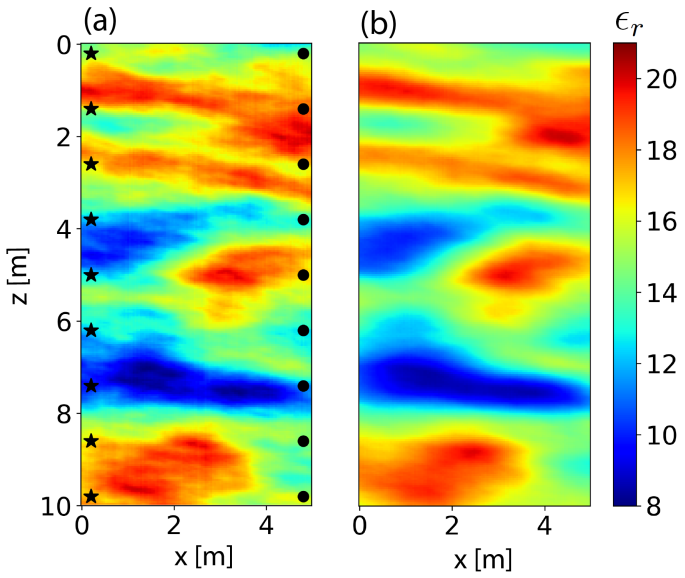

Figure 4.4: (a) Reference model and (b) its 100-PCA representation back-projected into the relative permittivity domain. The GPR sources are represented in (a) with black stars and the receivers with black circles.

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*).

4.3 Results

4.3.1 Test case

To evaluate the ASMC-SURR and the ASMC-SURR-HF methods, we consider a synthetic GPR crosshole tomography problem similar to *Meles et al.* (2022) with subsurface relative dielectric permittivity ϵ_r variations and constant electrical conductivity (Figure 4.4(a)). The dimension of the ϵ_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 travel times 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 travel times. The covariance structure is based on the 2-D Matérn geostatistical model (*Dietrich and 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.

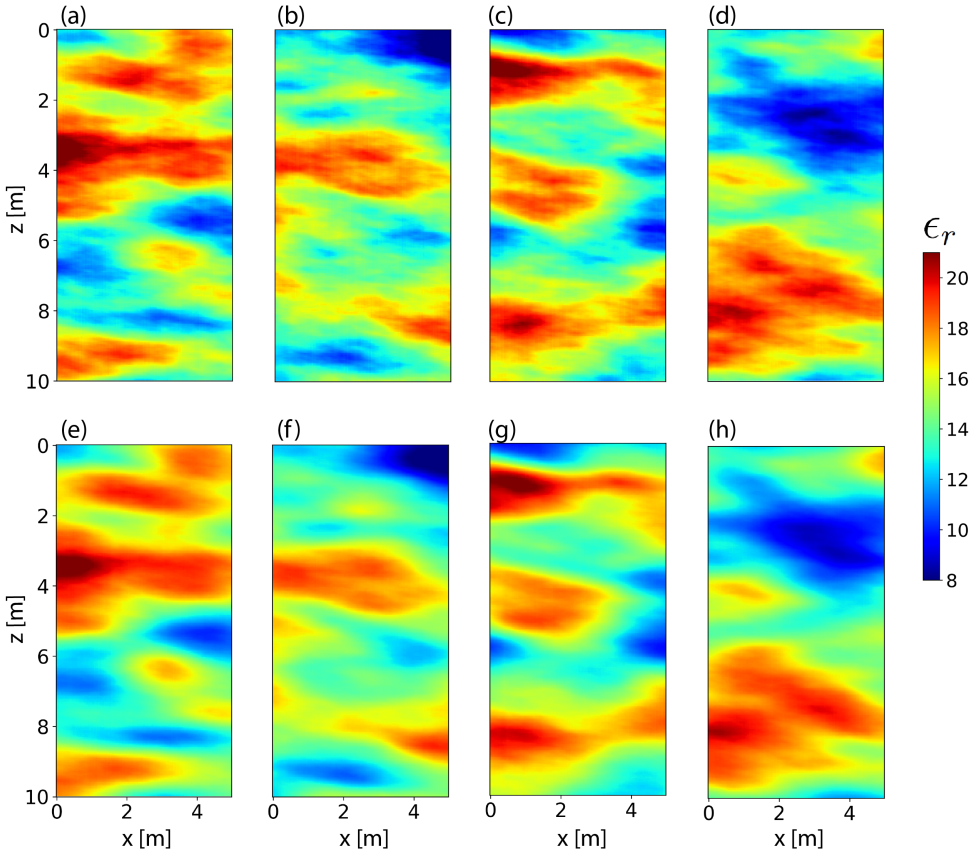

Figure 4.5: (a)-(d) Random samples obtained with the generative prior model and (e)-(h) their 100-PCA representation back-projected into the relative permittivity domain.

PCA representation

Following *Meles et al.* (2022), the PCA decomposition is learned from a set of 1000 samples drawn from the generative prior model. Figures 4.5(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 parameterize 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 Figure 4.4(b). Figures 4.5(e-h) show the resulting 100-PCA projection of the model realizations in Figures 4.5(a-d), back projected into the relative permittivity domain.

4.3.2 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 $CESS_{op}/N = 0.99$. When the acceptance rate gets lower than $AR_{min} = 15\%$ at α_{t-1}, the proposal scale ϕ is reduced a factor of $f = 20\%$ for α_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 its posterior PDF and the evidence using the high-fidelity solver only, thereby, allowing for comparisons with a reference solution. This reference solution is termed $\text{ASMC}_{\text{eikonal}}$ and it uses the *time2d* algorithm by *Podvin and 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 ($\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$). This configuration is also evaluated in the multifidelity setting ($\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-HF). We compare these results to those obtained when four re-training steps of the surrogate are performed ($\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$) and with a subsequent transition to the high-fidelity solver ($\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$-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 re-training is allowed. The final number of high-fidelity simulations needed for $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-HF and $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$-HF after switching to the high-fidelity solver depends on the resulting optimized α-correction F_{HF} (see section 4.2.5). 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 4.1 shows the resulting number of power posteriors, high-fidelity and surrogate evaluations for each test.

4.3.3 PCE surrogate training

The PCE surrogates are trained to learn the relationships between the PCA coefficients and the high-fidelity first-arrival travel times. 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 travel times computed on the 100-PCA back-projected relative permittivity models as output. In $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$, a set of 1000 prior input-output pairs are used to train the surrogate. In $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{updates}}$, 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) re-trained at regular intervals of $T_{up} = 35$ 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).

4.3.4 Covariance matrices

The data used for the various inversions are obtained by simulating the high-fidelity first-arrival travel times of the full reference model (Figure 4.4(a)), and then contaminating them with uncorrelated Gaussian noise with a standard deviation of $\sigma = 0.5$ ns. We consider a Gaussian likelihood function (equation 4.8) with a covariance matrix that incorporates the different error sources (equation 4.9). Figures 4.6(a) and (b) show the data error covariance matrix C_y and the PCA projection error covariance matrix C_{PCA}, respectively. The covariance matrix used in the likelihood calculation is the sum of the contributions C_y and C_{PCA} in the cases cases when the forward simulations are performed using the high-fidelity solver ($\text{ASMC}_{\text{eikonal}}$). Figure 4.6(c) shows the covariance matrix of the PCE errors C_{PCE} for the surrogate used in $\text{AMSC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$. When performing simulations using the PCE surrogate, we assume the covariance matrix in the likelihood function to be the sum of the three contributions C_y, C_{PCA} and C_{PCE}. The diagonals of the covariance matrices for this case are shown in Figure 4.6(d).

4.3.5 High-fidelity reference solution with eikonal solver

The inversion using the high-fidelity forward solver ($\text{ASMC}_{\text{eikonal}}$) resulted in $L = 192$ power posteriors, a total number of 4.8 million high-fidelity forward simulations and two resampling steps (Table 4.1). Figures 4.7(a) and (h) show the corresponding weighted posterior mean and standard deviation, respectively, while Figure 4.8(a) shows the logarithmic score values on the relative permittivity domain. Table 4.1 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* σ, 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 (Figure 4.6). The evidence estimation is -110.81.

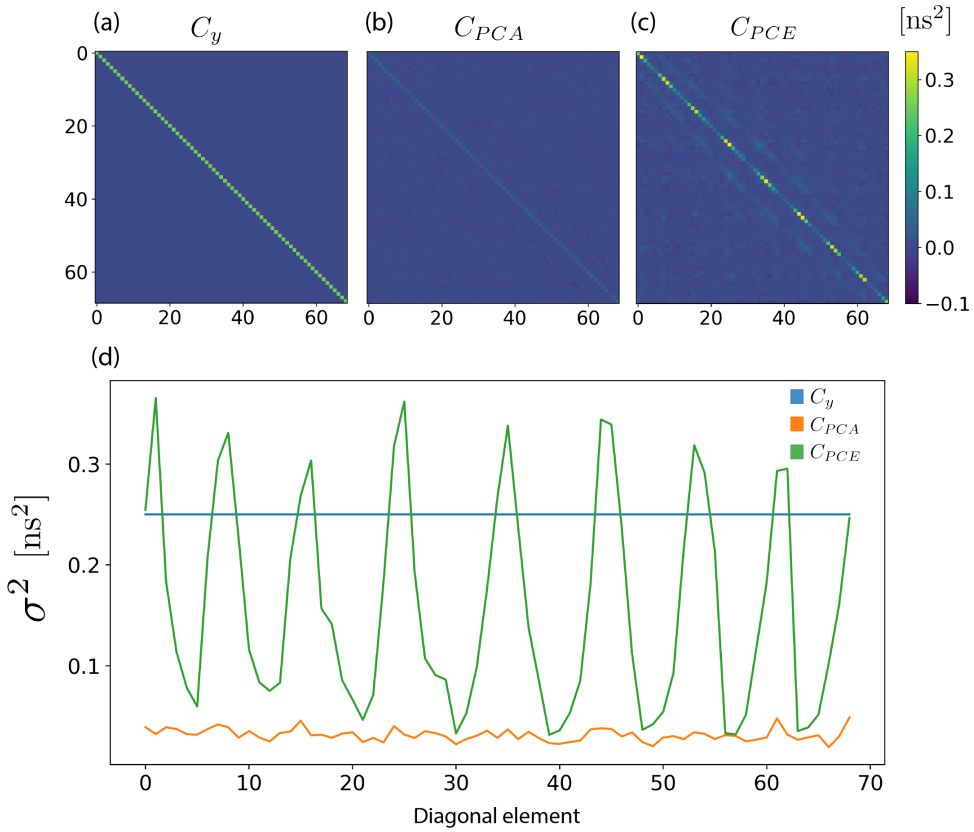

Figure 4.6: (a) Data error covariance matrix, (b) PCA projection error covariance matrix, (c) PCE surrogate modelling error covariance matrix for a surrogate trained using 1000 prior samples, and (d) their diagonal elements.

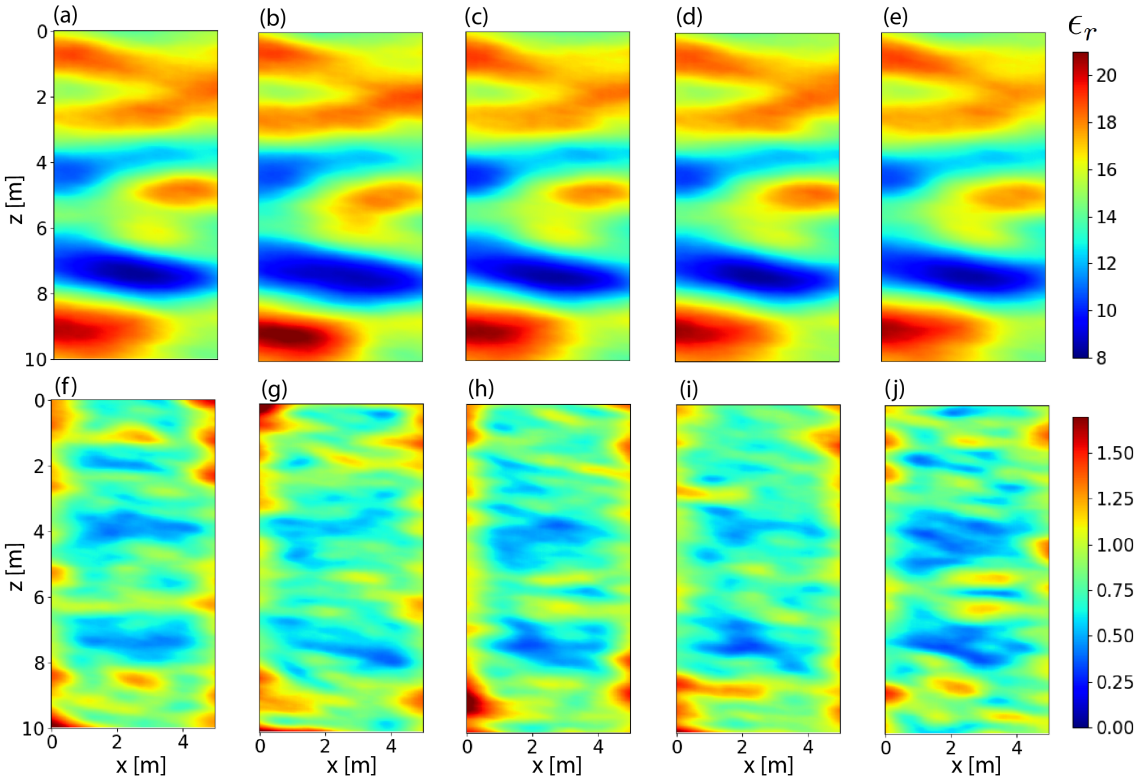

Figure 4.7: Posterior weighted mean model and standard deviation in, respectively, (a) and (f) $\text{ASMC}_{\text{eikonal}}$, (b) and (g) $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$, (c) and (h) $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-HF, (d) and (i) $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$, (e) and (j) $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{updates}}$-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)).

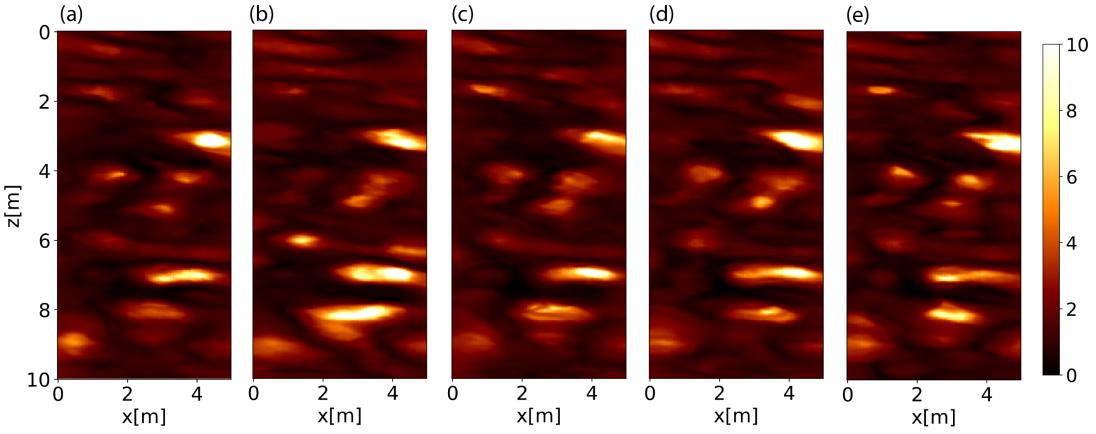

Figure 4.8: Logarithmic score values for (a) $\text{ASMC}_{\text{eikonal}}$, (b) $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$, (c) $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-HF, (d) $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{updates}}$, (e) and (l)$\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{updates}}$-HF

Table 4.1: Lengths of the ASMC runs and their performance metrics. The total number of iterations per particle is given by L, the number of power posteriors (inverse temperatures) arising from the adaptive nature of the algorithm (section 4.2.3), multiplied by the K MCMC iterations used to approximate each power posterior. All runs are performed using $N = 50$, $CESS_{op}/N = 0.99$, and $K = 500$ (except for the reduction to $K = 50$ in the high-fidelity stage of $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{updates}}$-HF).

	$\text{ASMC}_{\text{eikonal}}$	$\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$	$\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-HF	$\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{updates}}$	$\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{updates}}$-HF	$\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{updates}}$	$\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{updates}}$-HF
L	192	173	251	180	203	178	191
HF sim$[\times 10^3]$	$\tilde{4}800$	1	1950	1	576	1	33.5
HF sim relative	100%	0.02%	40.63%	0.02%	12%	-	-
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

4.3.6 Surrogate-based ASMC solutions with eikonal high-fidelity solver

Figures 4.7(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 Figures 4.8(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 metrics of Table 4.1, we see for the surrogate-only inversions (ASMC-SURR) that updating the surrogates during the inversion ($\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$) outperforms the case when the surrogate is based on prior realizations only ($\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$). Comparing these two cases, it is seen that the *SSIM* is 0.91 vs. 0.84, the *Mean log-score* 1.6 vs. 1.73 and the *RMSE output* 0.65 ns vs. 0.84 ns. Both runs underestimate the evidence when compared to the high-fidelity inversion, but the $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$ approximation is closer to the reference (-114.3 vs. -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 re-training 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 4.1 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 4.1 suggest that $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-HF and $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$-HF results are very similar. The main difference between the two inversions is that the

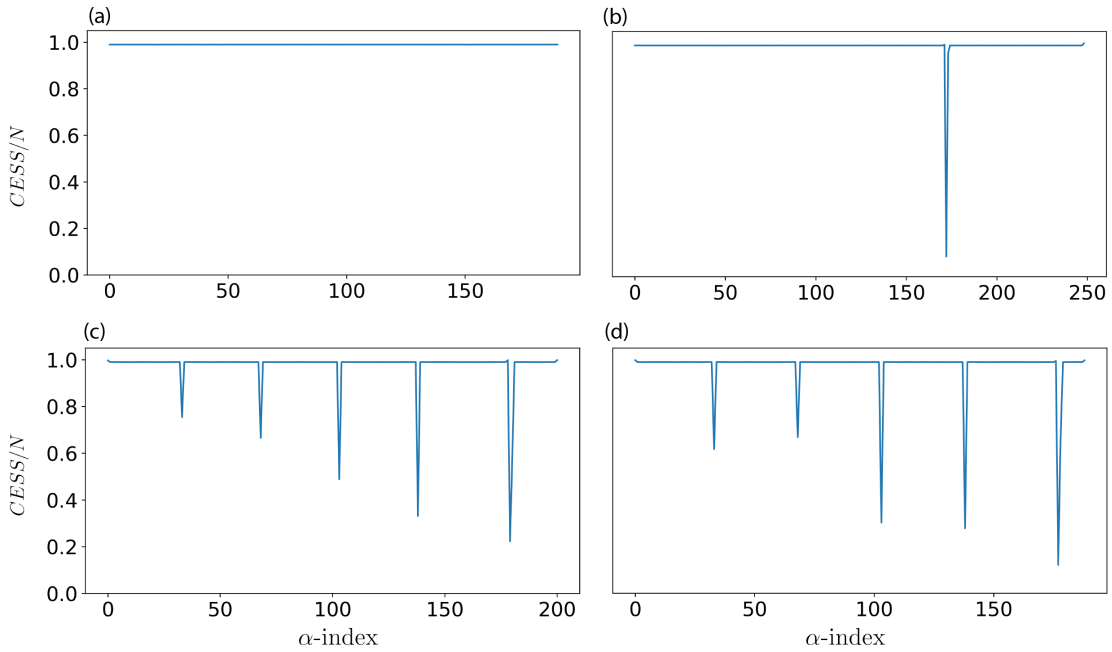

Figure 4.9: Conditional effective sample size over the number of particles ($CESS/N$) vs. inverse temperature index for (a) $\text{ASMC}_{\text{eikonal}}$, (b) $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-HF, (c) $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$-HF and (d) $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$-HF. The low values correspond to $CESS$-decreases occuring when updating the surrogate model or when transitioning to the high-fidelity forward solver.

α-correction factor is $F_{HF} = 0.53$ for $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$-HF compared to $F_{HF} = 0.15$ for $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-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 $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-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 the α-correction factor F_{HF}. Hence, updating the surrogate sequence results in a smoother transition to the high-fidelity solver requiring less computing time.

Figures 4.9(a-c) show the evolution of $CESS/N$ (equation 4.14) for $\text{ASMC}_{\text{eikonal}}$, $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-HF, and $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$-HF, respectively. The values are close to the optimal $CESS_{op} = 0.99$ except for the IS steps used to transition between forward solvers for which the $CESS/N$ is much lower. The diagonal elements of the modeling error covariance matrix C_{PCE} is shown in Figure 4.10(a). As the training set grows in $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$-HF, the mean error level in the C_{PCE} diagonals decreases. After the third update, the surrogate presents a lower error level than the surrogate trained using prior samples.

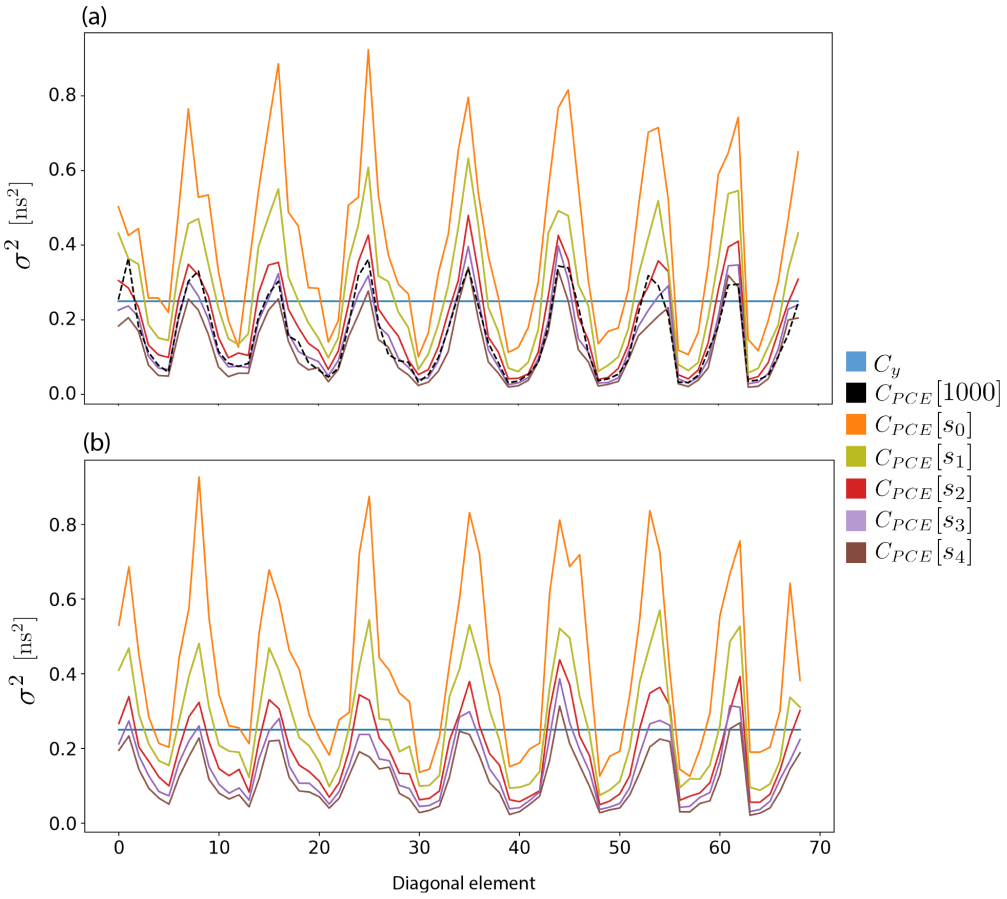

Figure 4.10: Evolution of the diagonal of the covariance modelling error matrix C_{PCE} for the four surrogate updates in (a) ASMC$_{\text{eikonal}}$-SURR$_{\text{update}}$ and (b) ASMC$_{\text{FDTD}}$-SURR$_{\text{update}}$. The initial surrogate s_0 is trained using 200 samples, and 200 more are added 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 C_y, and the black-dashed line in (a) is the C_{PCE} diagonal of the eikonal surrogate trained with 1000 prior samples used in ASMC$_{\text{eikonal}}$-SURR$_{\text{prior}}$.

4.3.7 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 and Knight*, 2006). Following *Meles et al.* (2022), automatic determination of travel times is performed by applying a threshold based on the relative maximum amplitude of each source-receiver pair. Running the ASMC inversion using always this expensive solver would require several months of computing (one parallelized simulation of 50 particles takes around 3.5-5 minutes); see also (*Hunziker et al.*, 2019). We test our proposed method using FDTD as the high-fidelity solver ($\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$) while keeping the same ASMC inversion setting as for the $\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{update}}$ case. The inversion needed 178 power posteriors and two resampling steps (Table 4.1). The correction F_{HF} for the transition to the high-fidelity stage is 0.73, which resulted in only 13 extra inverse temperatures ($\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$-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, at the cost of having a less accurate approximation of the importance distributions. The reduction to $K = 50$ for the steps involving HF makes it feasible to run the high-fidelity stage of the inversion. Figure 4.11 shows the posterior weighted mean, standard deviation and logarithmic score values for $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$ and $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$-HF. Figure 4.10(b) shows the evolution of the covariance matrices in the PCE updates and Figure 4.14 the $CESS/N$ values with the low values corresponding to transitions between forward solvers, respectively. The performance metrics in Table 4.1 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 ns 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 $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$-HF is 0.638 compared to 0.62 for $\text{ASMC}_{\text{eikonal}}$ and the *Mean log-score* is 1.09 compared to 1.46).

4.4 Discussion

Over the past 20 years, surrogate modeling 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 modeling 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 the corresponding modeling error covariance operator (*Li and 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 only from the prior. With ASMC-SURR there is an outstanding

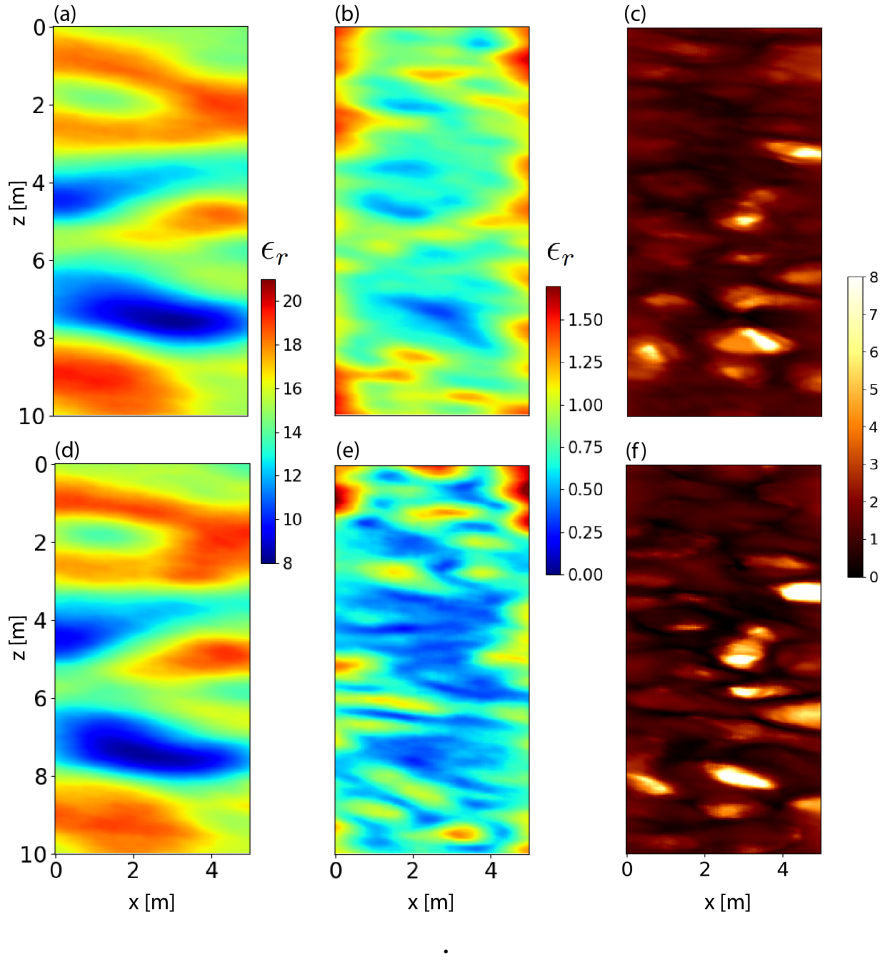

Figure 4.11: Posterior weighted mean model for (a) $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$ and (d) $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$-HF, posterior standard deviation for (b) $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$ and (e) $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$-HF, and logarithmic score value for (c) $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$ and (f) $\text{ASMC}_{\text{FDTD}}$-$\text{SURR}_{\text{update}}$-HF. Note that the standard deviations only consider those associated with the first 100 PCA components (see *Meles et al.* (2022))

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 then be used to quickly obtain posterior estimations for applications that do not require 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 the transition to high-fidelity. In this 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 overall computational gain due to the surrogate updates exceeds a factor of three. For this reason, re-training the surrogate is preferred over training only with prior samples.

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-dimensional 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 travel time for any given source-receiver pair) requires its own PCE model, with the training cost for each PCE depending nonlinearly 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 that of the whole inversion depends on the length of the ASMC process and the computational burden of high-fidelity modeling. In our ASMC-SURR-HF examples, the training represented less than 10% 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 $CESS_{op}$ on the total number of forward simulations is nonlinear and challenging to predict in advance. For this reason, it is recommended to initially select a sufficiently large $CESS_{op}$ and then adjust the value of K, with $CESS_{op}$ exceeding $0.99N$ (*Amaya et al.*, 2021). In this study we decided to start with this minimum recommended $CESS_{op}$ and increase 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 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 the less resources needed, as the high-fidelity solvers (ideally) require to run in parallel as many cores as the number of particles.

The surrogate updates based on regular intervals led to improvements compared to training based on prior samples only. The design of alternative and more elaborate surrogate re-training 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. Second, 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.

ASMC estimations are sensitive to the quality of the intermediate importance-based sampling steps as reflected in the variance of the importance weights. When the importance sampling steps are of poor quality, fewer particles contribute 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 centered 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.

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 (*Latz et al.*, 2018; *Beskos et al.*, 2016). When the computational resources are available and the study requires ascertained unbiased evidence estimations, one alternative is to perform a second non-adaptive ASMC run using the temperature sequence obtained in the first one (*Dai et al.*, 2022). In Appendix B, we provide the metrics obtained when re-running 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 (c.f., Tables 4.1 and 4.2).

The ASMC results vary slightly because of the stochastic nature of the algorithm, implying that multiple ASMC runs with the same setting should be performed to get a detailed assessment of the errors (e.g., *Amaya et al.* (2021)). In this study, we focus on the relative improvement and computational gain of the proposed method employing 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 travel-time 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)).

4.5 Conclusion

Using a GPR crosshole tomography example, we demonstrate that surrogate modeling can accelerate the ASMC inversion by orders of magnitude while still providing results that are highly similar to those obtained when using the high-fidelity solver only. We demonstrate that a multifidelity approach combining surrogate modeling 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 accuracy. When using surrogates only and considering the same high-fidelity budget for training, we show that re-training 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 re-training 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 parameterizations or GPR examples, but that it can be adapted to other surrogates, model parameterizations and geophysical problem settings.

4.6 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_{fp} inverse temperatures, and the models together with the responses are saved. The PCE surrogate is re-trained at regular intervals of T_{up} inverse temperatures. During an interval between surrogate updates, a total amount of $\tau_0 = N\lfloor T_{up}/T_{fp}\rfloor$ are collected. The initial surrogate s_0 is trained using τ_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 τ_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 modeling error covariance matrix C_{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, C_{PCE} is set to zero, as no modeling errors are assumed.

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 ($CESS_{op}$), ESS threshold (ESS^*),

number of MCMC iterations at each intermediate distribution (K), minimal acceptance rate (AR_{min}),

percentage of change (f) of the proposal scale ϕ,

period for evaluating and saving high-fidelity forward simulations for surrogate update (T_{fp}),

period of surrogate re-training (T_{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 $t_{fp} = 0$

Set $t_{up} = 0$

Iteration : Set $t = t + 1$

$t_{fp} = t_{fp} + 1$,

$t_{up} = t_{up} + 1$.

Do K MCMC iterations for each of the N particles (chains):

Propose moves $\boldsymbol{\theta}_p$ and accept or reject based on acceptance criterion (equation 4.18).

Tune proposal scale

If acceptance rate $AR < AR_{min}$ then decrease proposal scale factor: $\phi = \phi * (1 - \frac{f}{100})$

If $\phi < \phi_{min}$ then $\phi = \phi_{min}$.

Search for next power posterior

If $t_{up} < T_{up}$ do binary search for the increment $\Delta\alpha$ that gives the CESS (equation 4.14) that is the closest to $CESS_{op}$

Update $\alpha = min(1, \alpha + \Delta\alpha)$ and define the following power posterior.

Perform the IS step: compute the weight increments w_t (equation 4.12),

update and save the normalized weights W_t^i (equation 4.13) and the evidence $\pi(\mathbf{y})$ (equation4.17)

Evaluate high-fidelity - Update surrogate

If $t_{fp} = T_{fp}$ compute the N current high-fidelity forward responses and save them for training, set $t_{fp} = 0$.

If $t_{up} = T_{up}$, re-train surrogate and replace, update covariance matrix (equation 4.9 and 4.10), set $t_{up} = 0$.

Search for the α that yields the highest $CESS$.

Perform IS step to transition to the updated surrogate solver, compute the weight increments $w[s_j, s_{j+1}]_t^i$ (equation 4.19),

update and save the normalized weights W_t^i (equation 4.13) and the evidence $\pi(\mathbf{y})$ (equation4.17).

Resampling

Calculate ESS (equation 4.15), 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 α=1

Algorithm 1 extension: ASMC-SURR-HF

Transition to the high-fidelity solver

Search for the α that yields the highest $CESS$.

Perform IS step to transition to the high-fidelity forward solver, compute weight increments $w[s_j, HF]_t^i$ (equation4.22).

Update and save the normalized weights W_t^i (equation 4.13) and the evidence $\pi(\mathbf{y})$ (equation4.17).

Update covariance matrix.

Continue -Iteration- loop using the high-fidelity solver without *Evaluate high-fidelity - Update surrogate* step, until α=1.

Table 4.2: Resulting length of the ASMC runs and performance metrics of the re-runs 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 4.1).

	$\text{ASMC}_{\text{eikonal}}$	$\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$	$\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{prior}}$-HF	$\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{updates}}$	$\text{ASMC}_{\text{eikonal}}$-$\text{SURR}_{\text{updates}}$-HF
L	192	173	251	180	203
HF sim$[\times 10^3]$	$\tilde{4}800$	1	1950	1	576
HF sim relative	100%	0.02%	40.63%	0.02%	12%
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 σ	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

4.7 Appendix C

Re-runs 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 re-runs (Table 4.2) in which we keep fixed the adaptively-determined sequence of α_s of the eikonal runs of Table 4.1. 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.

Chapter 5

Conclusions and Outlook

5.1 Conclusions

Interpreting geophysical and hydrogeological inversion results obtained without considering uncertainty and inherent non-uniqueness can lead to inaccurate and misleading findings. On the other hand, performing Bayesian inversions that properly account for uncertainties and successfully sample the posterior distribution is often a difficult task. Two key current challenges in geophysical and hydrogeological inverse problems are (i) to account for prior geological realism, and (ii) to reduce the computing time associated with expensive high-fidelity forward solver. In this thesis, we have found that ASMC enhances the efficiency and freedom of exploration by sequentially sampling power posteriors at decreasing temperatures. Both deep generative models and MPS were used to encode and sample complex prior information in Bayesian inversions, something that is impossible to achieve using traditional geostatistical techniques. On the other hand, the ASMC method was combined with adaptive surrogate modeling strategies that significantly reduced the computing time while obtaining a comparable accuracy as for applications relying on high-fidelity solvers only.

In Chapter 2, ASMC showed to be more efficient and reliable in estimating the posterior PDF than a state-of-the art adaptive MCMC algorithm in a GPR crosshole tomography test case. The non-linearity of the inverse problem was enhanced by the use of deep generative networks that encode a geologically-realistic prior in its latent space. Both the ASMC-derived evidence and posterior estimation were insensitive to the model proposal schemes tested. Moreover, the evidence estimation was found to be robust with respect to ASMC parameters, and comparable in a high-noise setting with the unbiased Brute Force Monte Carlo estimator. When assessing the uncertainty in the evidence estimations, we found that our proposed single-run estimator is only reliable when enough intermediate Markov steps are used. The results of this chapter suggest that ASMC is particularly advantageous to solve non-linear geophysical problems for which MCMC approaches struggle.

In Chapter 3, we evaluated ASMC in a high-dimensional and bimodal hydrogeological inversion relying on an MPS-based prior and proposal scheme. For a similar computational budget, ASMC was more effective in locating the two high-probability regions of the posterior and fitted the data better than parallel tempering, a state-of-the-art MCMC-based tempering method. The adaptive nature of the ASMC algorithm reduces the complications and efforts of tuning the temperature and proposal scale needed in parallel tempering. The results indicate that further increasing the ASMC computational budget improves the balance on the sampling of the two modes. We also showed that intermediate ASMC estimations associated to the power posteriors can be related to higher noise assumptions in the observations, which can be useful in real applications when the data error level is unknown. This study highlights the potential of ASMC to target multimodal posterior distributions.

In Chapter 4, we explore the combination of ASMC with surrogate modeling in a GPR cross-hole tomography example. The PCE surrogate solvers successfully emulated the behavior of eikonal and FDTD high-fidelity solvers, drastically reducing the computational times compared to the full high-fidelity inversions. Re-training the surrogate as the inversion progresses was advantageous compared to training the surrogate using only prior samples. Depending on the accuracy needed and purposes of a given study, the inversion can be performed using only surrogates or it can be extended to a multifidelity scheme by incorporating a final high-fidelity stage. We conclude that using an updated surrogate scheme in a multifidelity ASMC framework substantially reduces both the computing resources needed for the original inversion (otherwise very difficult or even impossible to be solved given realistic computational times) and the loss of accuracy associated to inversions using surrogate solvers only.

In these studies (Chapters 2-4), the ASMC method has shown to be efficient in estimating the posterior PDF and the evidence. It manifested a remarkable ability to explore the posterior, outperforming widely used MCMC inversion methods. The resampling steps play an important role in explaining ASMC performance, as they limit weight degeneracy and focus the sampling on high-likelihood regions. Without these steps, the particles would very likely spend long times sampling very low-likelihood regions, making the process much less efficient. Overall, the method is attractive not only due to its efficiency, robustness and versatility, but also because of its ease of implementation. ASMC reduces the work and tuning needed to find proper algorithmic settings that often plague standard approaches. Moreover, its highly-parallelizable nature makes it convenient for computationally efficient implementations.

It is essential to develop and improve mathematical and statistical techniques that can effectively handle the challenges posed by complex modern hydrogeological and geophysical problems. ASMC has only recently been introduced in geophysics and hydrogeology and its use in these scientific fields is still limited. Through our studies, we conclude that ASMC is a powerful and practical method, that can be used especially when non-linearity, high-dimensionality or multi-modality challenge state-of-the-art MCMC-based approaches.

5.2 Limitations and Outlook

Even if the ASMC algorithm decreases the need of manual intervention through its adaptive strategies, choosing a suitable set of inversion parameters requires testing and tuning (*Dai et al.*, 2022). Providing guidance concerning the algorithmic parameter values to use is tricky, as it depends on the problem at hand. In Chapters 2-4, we provide practical recommendations that can help constrain and guide the choices. In this thesis, we relied on systematic resampling, but future studies could compare the performance of alternative resampling schemes (*Douc and Cappé*, 2005). It is important to keep in mind that estimations obtained with ASMC inversions are reliable as long as the intermediate power posteriors are well approximated. Generally, in Bayesian inversions, performing enough forward simulations such that the high-likelihood regions are well explored is computationally demanding, thereby, motivating future research in how to better integrate surrogate modeling.

As it is not trivial to build appropriate surrogate updating schemes within ASMC, it would be interesting to study how alternative strategies perform. Adaptive schemes to decide when to update the forward solver should be explored (*Latz et al.*, 2018), for example, based on the correspondence of the surrogate accuracy level with the tempering level (*Catanach et al.*, 2020). A better understanding of the trade-off between computational effort and accuracy could lead to more efficient choices and particular recommendations concerning updating schemes. It would also be useful to explore ASMC with other types of surrogate models than PCE based on, for example, Gaussian process regression (*Santner et al.*, 2003; *Rasmussen*, 2004) or deep learning (*Jin et al.*, 2020). On the other hand, we know that approximations of the surrogate modeling error in the inversion can lead to inaccurate estimations. Therefore, it is important to dedicate more effort to assess strategies to derive reliable estimations of the modeling error, for example, using neural networks (*Levy et al.*, 2021).

Developing robust and efficient inversion algorithms that can face the increasingly complex inverse problems arising in hydrogeology and geophysics is crucial. Realistic geophysical and hydrogeological inverse problems often involve models with a high number of parameters. As SMC methods have manifested their ability to explore high-dimensional posteriors (*Jasra et al.*, 2011; *Schäfer and Chopin*, 2013; *Kantas et al.*, 2014), they can be well-suited for tackling these problems. In this thesis, the ASMC method has shown to successfully address complex inverse problems in synthetic settings. The natural next step would be to test the method in real-case geophysical and hydrogeological scenarios. It would be interesting to assess the performance in particularly challenging problems with large data sets and/or three-dimensional models, for which the value and limitations of ASMC are still unknown. We also encourage using ASMC in model selection studies and compare the quality and speed of evidence computations with other state-of-the-art numerical methods. Even if standard SMC evidence estimations are unbiased (*Del Moral*, 2004), the impact of the adaptive strategies in the evidence estimations should be further explored, to better understand the extent of the bias induced by them (*Beskos et al.*, 2016).

Incorporating geologically-realistic priors in Bayesian inversions can significantly enhance the reliability and usefulness of the inferred subsurface properties and system predictions.

The integration of complex priors into Bayesian inversion is challenging, in this thesis we have tested methods based in categorical training images within ASMC. We would like to stress that continuous training images could be used as well, and conditioning data and further sources of information could be accounted for when training deep generative networks or included within the MPS algorithms. The development of realistic priors that can easily encode complex information without compromising the efficiency of inverse algorithms is a challenge where further progress would be highly useful in geophysics and hydrogeology.

Since SMC methods were originally intended for data assimilation, studies relying on data sets that become available progressively in time could also be considered. These methods can naturally handle such time-series because of their sequential nature, making them well-suited for this type of problems. The ASMC method could be tested in real-time geophysical and hydrogeological data assimilation problems, for example for monitoring and making real-time decisions.

Finally, we would like to highlight that the complexity of the statistical and numerical methods needed to tackle modern inverse problems can be a limitation for its application and popularity. It is important to link the problems for which these complex methods would be useful with the right algorithms to tackle them. It is relevant then, to dedicate effort in converting the algorithms into practical tools than can be easily accessed by the public. Simplifying their use, explaining their basic functioning and proving guidelines on how to properly address particular problems are key factors to facilitate and broaden their applicability.

Bibliography

Agapiou, S., O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart (2017), Importance sampling: Intrinsic dimension and computational cost, *Statistical Science, 32*(3), 405–431.

Alcolea, A., and P. Renard (2010), Blocking moving window algorithm: Conditioning multiple-point simulations to hydrogeological data, *Water Resources Research, 46*(8), W08511.

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

Amaya, M., N. Linde, and E. Laloy (2022), Hydrogeological multiple-point statistics inversion by adaptive sequential monte carlo, *Advances in Water Resources, 166*, 104252.

Angluin, D., and C. H. Smith (1983), Inductive inference: theory and methods, *ACM Computing Surveys (CSUR), 15*(3), 237–269.

Arridge, S. R. (1999), Optical tomography in medical imaging, *Inverse Problems, 15*(2), R41.

Bai, Y., R. V. Craiu, and A. F. Di Narzo (2011), Divide and conquer: a mixture-based approach to regional adaptation for MCMC, *Journal of Computational and Graphical Statistics, 20*(1), 63–79.

Bailer-Jones, C. (2015), A general method for bayesian time series modelling, *Technical report, Max Planck Institute for Astronomy, Heidelberg*.

Bayes, T. (1763), An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions, 53*(1763), 370–418.

Bergen, K. J., P. A. Johnson, V. Maarten, and G. C. Beroza (2019), Machine learning for data-driven discovery in solid Earth geoscience, *Science, 363*, eaau0323.

Beskos, A., A. Jasra, N. Kantas, and A. Thiery (2016), On the convergence of adaptive sequential Monte Carlo methods, *The Annals of Applied Probability, 26*(2), 1111–1146.

Blatman, G. (2009), Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis, Ph.D. thesis, Clermont-Ferrand 2.

Blatman, G., and B. Sudret (2011), Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of Computational Physics, 230*(6), 2345–2367.

Bodin, T., and M. Sambridge (2009), Seismic tomography with the reversible jump algorithm, *Geophysical Journal International*, *178*(3), 1411–1436.

Bolić, M., P. M. Djurić, and S. Hong (2004), Resampling algorithms for particle filters: A computational complexity perspective, *EURASIP Journal on Advances in Signal Processing*, *15*, 2267–2277.

Bond, C. E., A. D. Gibbs, Z. K. Shipton, S. Jones, et al. (2007), What do you think this is? "Conceptual uncertainty" in geoscience interpretation, *GSA Today*, *17*(11), 4.

Box, G. E. (1979), Robustness in the strategy of scientific model building. In R. L. Launer and G. N. Wilkinson (Eds.), *Robustness in Statistics*, pp. 201–236. Academic Press.

Brown, D., and A. Neal (1991), The analysis of the variance and covariance of products, *Biometrics*, *47*(2), 429–444.

Brunetti, C. (2018), Bayesian model selection in hydrogeophysics and hydrogeology, Ph.D. thesis, Université de Lausanne, Faculté des géosciences et de l'environnement.

Brunetti, C., N. Linde, and J. A. Vrugt (2017), Bayesian model selection in hydrogeophysics: Application to conceptual subsurface models of the South Oyster Bacterial Transport Site, Virginia, USA, *Advances in Water Resources*, *102*, 127–141.

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

Calvetti, D., and E. Somersalo (2007), *An introduction to Bayesian Scientific Computing: Ten Lectures On Subjective Computing*, vol. 2, Springer Science & Business Media.

Cappé, O., E. Moulines, and T. Ryden (2005), *Inference in Hidden Markov Models*, Springer Series in Statistics, Springer, New York.

Carter, J., and D. White (2013), History matching on the Imperial College fault model using parallel tempering, *Computational Geosciences*, *17*, 43–65.

Catanach, T. A., H. D. Vo, and B. Munsky (2020), Bayesian inference of stochastic reaction networks using multifidelity sequential tempered Markov chain Monte Carlo, *International Journal for Uncertainty Quantification*, *10*(6), 515–542.

Chan, H. P., and T. L. Lai (2013), A general theory of particle filters in hidden Markov models and some applications, *The Annals of Statistics*, *41*(6), 2877–2904.

Chandra, R., R. D. Müller, D. Azam, R. Deo, N. Butterworth, T. Salles, and S. Cripps (2019), Multicore parallel tempering Bayeslands for basin and landscape evolution, *Geochemistry, Geophysics, Geosystems*, *20*(11), 5082–5104.

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

Christen, J. A., and C. Fox (2005), Markov chain Monte Carlo using an approximation, *Journal of Computational and Graphical statistics*, *14*(4), 795–810.

Cotter, S. L., M. Dashti, J. C. Robinson, and A. M. Stuart (2009), Bayesian inverse problems for functions and applications to fluid mechanics, *Inverse Problems*, *25*(11), 115,008.

Cotter, S. L., G. O. Roberts, A. M. Stuart, and D. White (2013), MCMC methods for functions: modifying old algorithms to make them faster, *Statistical Science*, *28*(3), 424–446.

Craig, I., and J. Brown (1986), *Inverse Problems in Astronomy: A Guide to Inversion Strategies for Remotely Sensed Data*, Adam Hilger, Bristol.

Curtis, A., and A. Lomax (2001), Prior information, sampling distributions, and the curse of dimensionality, *Geophysics*, *66*(2), 372–378.

Dai, C., J. Heng, P. E. Jacob, and N. Whiteley (2022), An invitation to sequential Monte Carlo samplers, *Journal of the American Statistical Association*, *117*(539), 1587–1600.

Davies, L., A. Ley-Cooper, M. Sutton, and C. Drovandi (2023), Bayesian detectability of induced polarisation in airborne electromagnetic data, *Geophysical Journal International*, *ggad073*.

Del Moral, P. (2004), *Feynman-Kac Formulae: Genealogical and interacting particle systems with applications*, Springer, New York.

Del Moral, P., A. Doucet, and A. Jasra (2006), Sequential Monte Carlo samplers, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(3), 411–436.

Del Moral, P., A. Doucet, and A. Jasra (2012), An adaptive sequential Monte Carlo method for approximate Bayesian computation, *Statistics and Computing*, *22*, 1009–1020.

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

Domenico, P. A., and F. W. Schwartz (1997), *Physical and Chemical Hydrogeology*, John Wiley & sons.

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

Doucet, A., and A. M. Johansen (2011), A tutorial on particle filtering and smoothing: Fifteen years later. In: Crisan, D. and Rozovskii, B., Eds., *The Oxford Handbook of Nonlinear Filtering. Oxford University Press, New York.*, pp. 656–704.

Doucet, A., and A. Lee (2018), Sequential Monte Carlo methods. In: Marloes, M., Drton, M., Lauritzen, S., and Wainwright, M., Eds., *Handbook of Graphical Models,CRC Press*, pp. 165–189.

Doucet, A., N. De Freitas, N. J. Gordon, et al. (2001), *Sequential Monte Carlo Methods in Practice*, vol. 1, Springer.

Earl, D. J., and M. W. Deem (2005), Parallel tempering: Theory, applications, and new perspectives, *Physical Chemistry Chemical Physics*, *7*(23), 3910–3916.

Evans, M., and T. Swartz (1995), Methods for approximating integrals in statistics with special emphasis on bayesian integration problems, *Statistical science*, pp. 254–272.

Everett, M. E. (2013), *Near-Surface Applied Geophysics*, Cambridge University Press.

Falcioni, M., and M. W. Deem (1999), A biased Monte Carlo scheme for zeolite structure solution, *The Journal of Chemical Physics*, *110*(3), 1754–1766.

Fetter, C. W. (2018), *Applied Hydrogeology*, Waveland Press.

Friel, N., and A. N. Pettitt (2008), Marginal likelihood estimation via power posteriors, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(3), 589–607.

Fu, J., and J. J. Gómez-Hernández (2009), A blocking Markov chain Monte Carlo method for inverse stochastic hydrogeological modeling, *Mathematical Geosciences*, *41*(2), 105–128.

Gallovič, F., L. Valentová, J.-P. Ampuero, and A.-A. Gabriel (2019), Bayesian dynamic finite-fault inversion: 1. Method and synthetic test, *Journal of Geophysical Research: Solid Earth*, *124*(7), 6949–6969.

Gelman, A., and X.-L. Meng (1998), Simulating normalizing constants: From importance sampling to bridge sampling to path sampling, *Statistical Science*, *13*(2), 163–185.

Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, *7*(4), 457–472.

Geweke, J. (1989), Bayesian inference in econometric models using Monte Carlo integration, *Econometrica: Journal of the Econometric Society*, (6), 1317–1339.

Geyer, C. J. (1991), Markov chain Monte Carlo maximum likelihood, *Computing Science and Statistics: Proc. 23rd Symp. Interface*, p. 156–163.

Geyer, C. J., and E. A. Thompson (1995), Annealing Markov chain Monte Carlo with applications to ancestral inference, *Journal of the American Statistical Association*, *90*(431), 909–920.

Gideon, S. (1978), Estimating the dimension of a model, *The Annals of Statistics*, *6*(2), 461–464.

Gilks, W. R., S. Richardson, and D. Spiegelhalter (1995), *Markov chain Monte Carlo in practice*, CRC press.

Gómez-Hernández, J. J., and X.-H. Wen (1998), To be or not to be multi-Gaussian? a reflection on stochastic hydrogeology, *Advances in Water Resources*, *21*(1), 47–61.

Good, I. J. (1992), *Rational Decisions*, pp. 365–377, Springer Series in Statistics.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014), Generative adversarial nets, in *Advances in Neural Information Processing Systems*, pp. 2672–2680.

Goodfellow, I., Y. Bengio, and A. Courville (2016), *Deep Learning*, MIT Press.

Gravey, M., and G. Mariethoz (2020), QuickSampling v1.0: a robust and simplified pixel-based multiple-point simulation approach, *Geoscientific Model Development*, *13*(6), 2611–2630.

Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, *82*(4), 711–732.

Haario, H., M. Laine, A. Mira, and E. Saksman (2006), DRAM: efficient adaptive MCMC, *Statistics and Computing*, *16*, 339–354.

Hammersley, J., and D. Handscomb (1964), *Monte Carlo Methods*, Springer Netherlands.

Hammersley, J. M. (1960), Monte Carlo methods for solving multivariable problems, *Annals of the New York Academy of Sciences*, *86*(3), 844–874.

Hansen, T. M., K. S. Cordua, and K. Mosegaard (2012), Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling, *Computational Geosciences*, *16*(3), 593–611.

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

Horowitz, J. L. (2014), Ill-posed inverse problems in economics, *Annual Review of Economics*, *6*(1), 21–51.

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

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

Jäggli, C., J. Straubhaar, and P. Renard (2017), Posterior population expansion for solving inverse problems, *Water Resources Research*, *53*(4), 2902–2916.

Jäggli, C., J. Straubhaar, and P. Renard (2018), Parallelized adaptive importance sampling for solving inverse problems, *Frontiers in Earth Science*, *6*, 203.

Jasra, A., D. A. Stephens, A. Doucet, and T. Tsagaris (2011), Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo, *Scandinavian Journal of Statistics*, *38*(1), 1–22.

Jefferys, W. H., and J. O. Berger (1992), Ockham's razor and Bayesian analysis, *American Scientist*, *80*(1), 64–72.

Jeremiah, E., S. Sisson, L. Marshall, R. Mehrotra, and A. Sharma (2011), Bayesian calibration and uncertainty analysis of hydrological models: A comparison of adaptive Metropolis and Sequential Monte Carlo samplers, *Water Resources Research*, *47*(7), W07547.

Jeremiah, E., S. A. Sisson, A. Sharma, and L. Marshall (2012), Efficient hydrological model parameter optimization with sequential Monte Carlo sampling, *Environmental Modelling & Software*, *38*, 283–295.

Jetchev, N., U. Bergmann, and R. Vollgraf (2016), Texture synthesis with spatial generative adversarial networks, *arXiv preprint arXiv:1611.08207*.

Jin, Z. L., Y. Liu, and L. J. Durlofsky (2020), Deep-learning-based surrogate model for reservoir simulation with time-varying well controls, *Journal of Petroleum Science and Engineering*, *192*, 107,273.

Jolliffe, I. T., and J. Cadima (2016), Principal component analysis: a review and recent developments, *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150,202.

Juda, P., and P. Renard (2021), An attempt to boost posterior population expansion using fast machine learning algorithms, *Frontiers in Artificial Intelligence*, *4*, 25.

Kaipio, J., and E. Somersalo (2007), Statistical inverse problems: discretization, model reduction and inverse crimes, *Journal of Computational and Applied Mathematics*, *198*(2), 493–504.

Kantas, N., A. Beskos, and A. Jasra (2014), Sequential Monte Carlo methods for high-dimensional inverse problems: A case study for the Navier–Stokes equations, *SIAM/ASA Journal on Uncertainty Quantification*, *2*(1), 464–489.

Karpatne, A., I. Ebert-Uphoff, S. Ravela, H. A. Babaie, and V. Kumar (2018), Machine learning for the geosciences: Challenges and opportunities, *IEEE Transactions on Knowledge and Data Engineering*, *31*(8), 1544–1554.

Kass, R. E., and A. E. Raftery (1995), Bayes factors, *Journal of the American Statistical Association*, *90*(430), 773–795.

Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal (1998), Markov chain Monte Carlo in practice: a roundtable discussion, *The American Statistician*, *52*(2), 93–100.

Kersaudy, P., B. Sudret, N. Varsier, O. Picon, and J. Wiart (2015), A new surrogate modeling technique combining Kriging and polynomial chaos expansions-Application to uncertainty analysis in computational dosimetry, *Journal of Computational Physics*, *286*, 103–117.

Khaki, M., M. S. Filmer, W. E. Featherstone, M. Kuhn, L. K. Bui, and A. L. Parker (2019), A sequential Monte Carlo framework for noise filtering in inSAR time series, *IEEE Transactions on Geoscience and Remote Sensing*, *58*(3), 1904–1912.

Kingma, D. P., and M. Welling (2013), Auto-encoding variational Bayes, *arXiv preprint arXiv:1312.6114*.

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

Kitanidis, P. K. (1997), *Introduction to Geostatistics: Applications in Hydrogeology*, Cambridge University Press.

Knuth, K. H., M. Habeck, N. K. Malakar, A. M. Mubeen, and B. Placek (2015), Bayesian Evidence and Model Selection, *Digital Signal Processing*, *47*, 50–67.

Koltermann, C. E., and S. M. Gorelick (1996), Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches, *Water Resources Research*, *32*(9), 2617–2658.

Kong, A., J. S. Liu, and W. H. Wong (1994), Sequential imputations and Bayesian missing data problems, *Journal of the American Statistical Association*, *89*(425), 278–288.

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

Krüger, F., S. Lerch, T. Thorarinsdottir, and T. Gneiting (2021), Predictive inference based on Markov chain Monte Carlo output, *International Statistical Review*, *89*(2), 274–301.

Künze, R., and I. Lunati (2012), An adaptive multiscale method for density-driven instabilities, *Journal of Computational Physics*, *231*(17), 5557–5570.

Laloy, E., and J. A. Vrugt (2012), High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing, *Water Resources Research*, *48*(1).

Laloy, E., B. Rogiers, J. A. Vrugt, D. Mallants, and D. Jacques (2013), Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion, *Water Resources Research*, *49*(5), 2664–2682.

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

Laloy, E., N. Linde, D. Jacques, and G. Mariethoz (2016), Merging parallel tempering with sequential geostatistical resampling for improved posterior exploration of high-dimensional subsurface categorical fields, *Advances in Water Resources*, *90*, 57–69.

Laloy, E., R. Hérault, J. Lee, D. Jacques, and N. Linde (2017), Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network, *Advances in Water Resources*, *110*, 387–405.

Laloy, E., R. Hérault, D. Jacques, and N. Linde (2018), Training-image based geostatistical inversion using a spatial generative adversarial neural network, *Water Resources Research*, *54*(1), 381–406.

Laloy, E., N. Linde, C. Ruffino, R. Hérault, G. Gasso, and D. Jacques (2019), Gradient-based deterministic inversion of geophysical data with generative adversarial networks: Is it feasible?, *Computers & Geosciences*, *133*, 104,333.

Latz, J., I. Papaioannou, and E. Ullmann (2018), Multilevel sequential^2 Monte Carlo for Bayesian inverse problems, *Journal of Computational Physics*, *368*, 154–178.

LeCun, Y., Y. Bengio, and G. Hinton (2015), Deep learning, *Nature*, *521*(7553), 436–444.

Lee, A., and N. Whiteley (2018), Variance estimation in the particle filter, *Biometrika*, *105*(3), 609–625.

Lee, B. S., M. Haran, R. W. Fuller, D. Pollard, and K. Keller (2020), A fast particle-based approach for calibrating a 3-D model of the Antarctic ice sheet, *Annals of Applied Statistics*, *14*(2), 605–634.

Levy, S., J. Hunziker, E. Laloy, J. Irving, and N. Linde (2021), Using deep generative neural networks to account for model errors in Markov chain Monte Carlo inversion, *Geophysical Journal International*, *228*(2), 1098–1118.

Lewis, S. M., and A. E. Raftery (1997), Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator, *Journal of the American Statistical Association*, *92*(438), 648–655.

Li, J., and Y. M. Marzouk (2014), Adaptive construction of surrogates for the Bayesian solution of inverse problems, *SIAM Journal on Scientific Computing*, *36*(3), A1163–A1186.

Linde, N. (2014), Falsification and corroboration of conceptual hydrological models using geophysical data, *Wiley Interdisciplinary Reviews: Water*, *1*(2), 151–171.

Linde, N., P. Renard, T. Mukerji, and J. Caers (2015), Geological realism in hydrogeological and geophysical inverse modeling: A review, *Advances in Water Resources*, *86*, 86–101.

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

Lowrie, W., and A. Fichtner (2020), *Fundamentals of Geophysics*, Cambridge University Press.

Lu, D., M. Ye, and S. P. Neuman (2011), Dependence of Bayesian model selection criteria and Fisher information matrix on sample size, *Mathematical Geosciences*, *43*, 971–993.

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

Lüthen, N., S. Marelli, and B. Sudret (2022), Automatic selection of basis-adaptive sparse polynomial chaos expansions for engineering applications, *International Journal for Uncertainty Quantification*, *12*(3), 10.1615.

MacKay, D. J. (2003), *Information Theory, Inference and Learning Algorithms*, Cambridge University Press.

Marelli, S., and B. Sudret (2014), UQLab: A framework for uncertainty quantification in Matlab. In: Beer, M., Au, S. and Hall, J.W., Eds., *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, pp. 2554–2563.

Marelli, S., P.-R. Wagner, C. Lataniotis, and B. Sudret (2021), Stochastic Spectral Embedding, *International Journal for Uncertainty Quantification*, *11*(2), 25–47.

Marelli, S., N. Lüthen, and B. Sudret (2022), UQLab user manual – Polynomial chaos expansions, *Tech. rep.*, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, report UQLab-V2.0-104.

Mariethoz, G., and J. Caers (2014), *Multiple-Point Geostatistics: Stochastic Modeling with Training Images*, John Wiley & Sons.

Mariethoz, G., P. Renard, and J. Caers (2010a), Bayesian inverse problem and optimization with iterative spatial resampling, *Water Resources Research*, *46*(11), W11530.

Mariethoz, G., P. Renard, and J. Straubhaar (2010b), The direct sampling method to perform multiple-point geostatistical simulations, *Water Resources Research*, *46*(11), W11,536.

Marinari, E., and G. Parisi (1992), Simulated tempering: a new Monte Carlo scheme, *Europhysics letters*, *19*(6), 451.

Marzouk, Y., and D. Xiu (2009), A stochastic collocation approach to Bayesian inference in inverse problems, *Communications in Computational Physics*, *6*(4), 826–847.

Marzouk, Y. M., H. N. Najm, and L. A. Rahn (2007), Stochastic spectral methods for efficient bayesian solution of inverse problems, *Journal of Computational Physics*, *224*(2), 560–586.

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

Metropolis, N., and S. Ulam (1949), The Monte Carlo method, *Journal of the American Statistical Association*, *44*(247), 335–341.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *The Journal of Chemical Physics*, *21*(6), 1087–1092.

Miltenberger, A., L. Wang, T. Mukerji, and J. Caers (2023), Formulating and solving the data-consistent geophysical inverse problem for subsurface modeling applications, *EarthArXiv preprint*.

Mohamed, L., B. Calderhead, M. Filippone, M. Christie, and M. Girolami (2012), Population MCMC methods for history matching and uncertainty quantification, *Computational Geosciences*, *16*, 423–436.

Mosegaard, K., and M. Sambridge (2002), Monte Carlo analysis of inverse problems, *Inverse problems*, *18*(3), R29.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *Journal of Geophysical Research: Solid Earth*, *100*(B7), 12431–12447.

Mosser, L., O. Dubrule, and M. J. Blunt (2017), Reconstruction of three-dimensional porous media using generative adversarial neural networks, *Physical Review E*, *96*(4), 043–309.

Mosser, L., O. Dubrule, and M. J. Blunt (2020), Stochastic seismic waveform inversion using generative adversarial networks as a geological prior, *Mathematical Geosciences*, *52*(1), 53–79.

Mueller, J. L., and S. Siltanen (2012), *Linear and Nonlinear Inverse Problems with Practical Applications*, SIAM.

Naylor, J., and A. Smith (1988), Econometric illustrations of novel numerical integration strategies for bayesian inference, *Journal of Econometrics*, *38*(1-2), 103–125.

Neal, R. M. (2001), Annealed importance sampling, *Statistics and Computing*, *11*(2), 125–139.

Newton, M. A., and A. E. Raftery (1994), Approximate Bayesian inference with the weighted likelihood bootstrap, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *56*(1), 3–26.

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

Nicolson, H., A. Curtis, B. Baptie, and E. Galetti (2012), Seismic interferometry and ambient noise tomography in the british isles, *Proceedings of the Geologists' Association*, *123*(1), 74–86.

Nissinen, A., L. Heikkinen, and J. Kaipio (2007), The Bayesian approximation error approach for electrical impedance tomography-experimental results, *Measurement Science and Technology*, *19*(1), 015,501.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), Verification, validation, and confirmation of numerical models in the Earth sciences, *Science*, *263*(5147), 641–646.

Park, C., R. T. Haftka, and N. H. Kim (2017), Remarks on multi-fidelity surrogates, *Structural and Multidisciplinary Optimization*, *55*, 1029–1050.

Parker, R. L. (1994), *Geophysical Inverse Theory*, Princeton University Press.

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

Peterson, J. E., Jr (2001), Pre-inversion corrections and analysis of radar tomographic data, *Journal of Environmental & Engineering Geophysics*, *6*(1), 1–18.

Pike, E. R., J. G. McWhirter, M. Bertero, and C. De Mol (1984), Generalised information theory for inverse problems in signal processing, in *IEE Proceedings F (Communications, Radar and Signal Processing)*, vol. 131, pp. 660–667, IET.

Pirot, G., J. Straubhaar, and P. Renard (2015), A pseudo genetic model of coarse braided-river deposits, *Water Resources Research*, *51*(12), 9595–9611.

Pirot, G., E. Huber, J. Irving, and N. Linde (2019), Reduction of conceptual model uncertainty using ground-penetrating radar profiles: Field-demonstration for a braided-river aquifer, *Journal of Hydrology*, *571*, 254–264.

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

Poeter, E., and D. Anderson (2005), Multimodel ranking and inference in ground water modeling, *Groundwater*, *43*(4), 597–605.

Popper, K. (2005), *The Logic of Scientific Discovery*, Routledge.

Rasmussen, C. E. (2004), Gaussian processes in machine learning, in *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures*, pp. 63–71, Springer.

Renard, P., and D. Allard (2013), Connectivity metrics for subsurface flow and transport, *Advances in Water Resources*, *51*, 168–196.

Reuschen, S., T. Xu, and W. Nowak (2020), Bayesian inversion of hierarchical geostatistical models using a parallel-tempering sequential Gibbs MCMC, *Advances in Water Resources*, *141*, 103–614.

Reuschen, S., F. Jobst, and W. Nowak (2021), Efficient discretization-independent Bayesian inversion of high-dimensional multi-Gaussian priors using a hybrid MCMC, *Water Resources Research*, *57*(8), WR030051.

Reynolds, J. M. (2011), *An Introduction to Applied and Environmental Geophysics*, John Wiley & Sons.

Rossat, D., J. Baroth, M. Briffaut, and F. Dufour (2022), Bayesian inversion using adaptive polynomial chaos Kriging within subset simulation, *Journal of Computational Physics*, *455*, 110986.

Roth, K., R. Schulin, H. Flühler, and W. Attinger (1990), Calibration of time domain reflectometry for water content measurement using a composite dielectric approach, *Water Resources Research*, *26*(10), 2267–2273.

Rothman, D. H. (1985), Nonlinear inversion, statistical mechanics, and residual statics estimation, *Geophysics*, *50*(12), 2784–2796.

Rothman, D. H. (1986), Automatic estimation of large residual statics corrections, *Geophysics*, *51*(2), 332–346.

Rubin, Y., and S. S. Hubbard (2006), *Hydrogeophysics*, vol. 50, Springer Science & Business Media.

Ruggeri, P., J. Irving, and K. Holliger (2015), Systematic evaluation of sequential geostatistical resampling within MCMC for posterior sampling of near-surface geophysical inverse problems, *Geophysical Journal International*, *202*(2), 961–975.

Sambridge, M. (2014), A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophysical Journal International*, *196*(1), 357–374.

Sambridge, M., and K. Mosegaard (2002), Monte Carlo methods in geophysical inverse problems, *Reviews of Geophysics*, *40*(3), 3–1–3–29.

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

Scales, J. A., and L. Tenorio (2001), Prior information and uncertainty in inverse problems, *Geophysics*, *66*(2), 389–397.

Schäfer, C., and N. Chopin (2013), Sequential Monte Carlo on large binary sampling spaces, *Statistics and Computing*, *23*, 163–184.

Schobi, R., B. Sudret, and J. Wiart (2015), Polynomial-chaos-based Kriging, *International Journal for Uncertainty Quantification*, *5*(2), 171–193.

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

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

Sisson, S. A. (2005), Transdimensional Markov chains: A decade of progress and future perspectives, *Journal of the American Statistical Association*, *100*(471), 1077–1089.

Skilling, J. (2004), Nested sampling, in *AIP Conference Proceedings*, vol. 735, pp. 395–405, American Institute of Physics.

Skilling, J. (2006), Nested sampling for general Bayesian computation, *Bayesian Analysis*, *1*(4), 833–860.

Strebelle, S. (2002), Conditional simulation of complex geological structures using multiple-point statistics, *Mathematical geology*, *34*(1), 1–21.

Stuart, A. M. (2010), Inverse problems: a Bayesian perspective, *Acta Numerica*, *19*, 451–559.

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

Torre, E., S. Marelli, P. Embrechts, and B. Sudret (2019), Data-driven polynomial chaos expansion for machine learning regression, *Journal of Computational Physics*, *388*, 601–623.

Van Leeuwen, P. J. (2009), Particle filtering in geophysical systems, *Monthly Weather Review*, *137*(12), 4089–4114.

Veettil, D. R. A., and K. Clark (2020), Bayesian geosteering using sequential Monte Carlo methods, *Petrophysics*, *61*(01), 99–111.

Volpi, E., G. Schoups, G. Firmani, and J. A. Vrugt (2017), Sworn testimony of the model evidence: Gaussian Mixture Importance (GAME) sampling, *Water Resources Research*, *53*(7), 6133–6158.

Von Neumann, J., et al. (1951), Various techniques used in connection with random digits, *Applied Math Series*, *12*, 768–770.

Vrugt, J. A. (2016), Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation, *Environmental Modelling & Software*, *75*, 273–316.

Vrugt, J. A., C. J. ter Braak, C. G. Diks, and G. Schoups (2013), Hydrologic data assimilation using particle Markov Chain Monte Carlo simulation: Theory, concepts and applications, *Advances in Water Resources*, *51*, 457–478.

Wagner, P.-R., R. Fahrni, M. Klippel, A. Frangi, and B. Sudret (2020), Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels, *Engineering structures*, *205*, 110063.

Wagner, P.-R., S. Marelli, and B. Sudret (2021), Bayesian model inversion using stochastic spectral embedding, *Journal of Computational Physics*, *436*, 110141.

Wang, L., H. Li, and Y. Fan (2019), Bayesian inversion of logging-while-drilling extra-deep directional resistivity measurements using parallel tempering Markov Chain Monte Carlo sampling, *IEEE Transactions on Geoscience and Remote Sensing*, *57*(10), 8026–8036.

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

Werner, M. J., K. Ide, and D. Sornette (2011), Earthquake forecasting based on data assimilation: sequential Monte Carlo methods for renewal point processes, *Nonlinear Processes in Geophysics*, *18*, 49–70.

Xie, W., P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen (2011), Improving marginal likelihood estimation for Bayesian phylogenetic model selection, *Systematic Biology*, *60*(2), 150–160.

Xiu, D., and G. E. Karniadakis (2002), The Wiener–Askey polynomial chaos for stochastic differential equations, *SIAM Journal on Scientific Computing*, *24*(2), 619–644.

Zahner, T., T. Lochbühler, G. Mariethoz, and N. Linde (2016), Image synthesis with graph cuts: a fast model proposal mechanism in probabilistic inversion, *Geophysical Journal International*, *204*(2), 1179–1190.

Zeng, X., M. Ye, J. Wu, D. Wang, and X. Zhu (2018), Improved nested sampling and surrogate-enabled comparison with other marginal likelihood estimators, *Water Resources Research*, *54*(2), 797–826.

Zhou, Y., A. M. Johansen, and J. A. Aston (2016), Toward automatic model comparison: an adaptive sequential Monte Carlo approach, *Journal of Computational and Graphical Statistics*, *25*(3), 701–726.

Zhu, G., X. Li, J. Ma, Y. Wang, S. Liu, C. Huang, K. Zhang, and X. Hu (2018), A new moving strategy for the sequential Monte Carlo approach in optimizing the hydrological model parameters, *Advances in Water Resources*, *114*, 164–179.

Zinn, B., and C. F. Harvey (2003), When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields, *Water Resources Research, 39*(3), 1051.

Acknowledgements

Here comes a key section of this work, as it would never have been possible without the people behind it. It is with great joy that I would like to acknowledge the individuals who were a part of this journey.

My supervisor and mentor, **Niklas Linde**. For giving me the opportunity and trust to take over this project. For always being an outstanding example, showing us an admirably ethical and critical approach to research. It is unmeasurable how much I have learned from you thanks to your constant commitment, willingness, insightfulness and patience along the way.

Eric Laloy, **Giovanni Meles** and **Stefano Marelli**, whose contributions greatly enriched this work, but above all, for the good energy and predisposition that made each of our discussions an enlightening learning experience. **Arnaud Doucet**, whose expertise motivated this dive into sequential Monte Carlo methods, and **David Ginsbourger's group**, for the shared meetings and the useful suggestions for these projects.

My **collegues and friends** in Geopolis, for making the office a great place to be. And especially **Shiran** and **Lea**, the best two PhD sailing partners. Thank you both for sharing every little step of the way and make it an unforgettable adventure.

My **mother**, my **father** and my **sister**, for being the very inner and solid core of who I am. For your unwavering strength, perspective and support. And for everything else, that I doubt I would ever be able to put down in words.

My **family**. My grandma, for her immense wisdom combined with limitless love. My aunts, uncles, cousins and nieces, who never failed to encourage and cheer me in the most joyful ways. My **friends in Argentina**, for always finding the most creative approaches to being part of this process despite the distance, and Paula Vega, for her smart and perceptive counsel.

Stade Lausanne field hockey club and its people, who welcomed me to the place that ending up being my second home in Lausanne. Especially, my amazing team, for all the trainings and trips to play around Switzerland that were my best source of energy.

My **friends in this beautiful city**, for the uncountable adventures, laughs, melodies and landscapes shared during these years. Caro, Vale, Marti, Nacho, Maite, Chuli, Agos, Lu, Cata, Santi, Tomi, Agus, Marian, Clari, Marqui, Dai, Rodri, Queque, Anita, Mau, Ricky, Rodo y Mel. You are all the joy and support I could have ever imagined.

My deepest and infinite thanks to all of you.

- This work was funded by the Swiss National Science Foundation, project number 184574 -

Macarena **Amaya**

Birthdate: 27-09-1991
Address: Rue du Villars 15, 1024 Ecublens, Switzerland
(+41) 762309240 | amayamacarena@gmail.com | macarena-amaya-94804017b

Education

University of Lausanne (UNIL) — *Lausanne, Switzerland*
PhD student at Institute of Earth Sciences — *October 2023*
- Thesis: *Geophysical and Hydrogeological Bayesian Inversion by Adaptive Sequential Monte Carlo*
- Founding: Swiss National Science Foundation

National University of La Plata (UNLP) — *La Plata, Arg.*
Geophysics [BSc + MSc eq.] at Faculty of Astronomical and Geophysical Sciences — *December, 2018*
- Thesis: *Identification and characterization of paleochannels using electrical resistivity tomographies*
- Founding: Geoscientists without Borders, Society of Exploration Geophysicists

Work Experience

University of Lausanne (UNIL) - Institute of Earth Sciences — *Lausanne, Switzerland*
PhD Research and Assistant — *Nov. 2019 - Oct. 2023*
- Project: GEOFACES

National University of La Plata (UNLP) – Faculty of Astronomical and Geophysical Sciences — *La Plata, Arg.*
Assistant — *March 2016 - March 2019*
- Courses: Geomatics and General Geophysics

Scientific Research Committee of Buenos Aires (CIC-BA) — *La Plata, Arg.*
Scientific Training — *Jan 2018- Jan 2019*
- Project: *Groundwater resources in Chaco province, Argentina.*

National Institute of Scientific Research (INRS) — *Québec, Canada*
Research internship — *Jan. 2018 - May 2018*
- Project: *High resolution mapping of ice mass loss in the Gulf of Alaska.*

Laval University (ULAVAL) — *Québec, Canada*
Research internship — *Jan. 2018 - May 2018*
- Project: *Integrated water resources management in the production of cranberries*

Honors, Awards & Scholarships

2018	**Gary & Lorene Servos Scholarship**, Society of Exploration Geophysicists	*Tulsa, USA*
2018	**Scientific Training Scholarship**, CIC-BA.	*Buenos Aires, Arg*
2017	**Emerging Leaders in the Americas Program Scholarship**, Global Affairs Canada	*Canada*
2019	**Distinguished graduate student**, National University of La Plata	*La Plata, Arg.*

Conference Presentations

Conference of the International Association for Mathematical Geosciences (IAMG)

Nancy, France

Macarena Amaya, Niklas Linde, Eric Laloy. *2022*

- *Hydrogeological multiple-point statistics inversion by adaptive sequential Monte Carlo.*

American Geophysical Union (AGU) Fall Meeting

San Francisco, USA

Macarena Amaya, Niklas Linde, Eric Laloy. *2022*

- *Adaptive sequential Monte Carlo in the context of hydrogeological inverse problems with complex spatial priors.*

European Geosciences Union (EGU) General Assembly

Vienna, Austria

Macarena Amaya, Niklas Linde, Eric Laloy. *2022*

- *Hydrogeological inference by adaptive sequential Monte Carlo with geostatistical resampling model proposals.*

American Geophysical Union (AGU) Fall Meeting

Online

Macarena Amaya, Niklas Linde, Eric Laloy, Arnaud Doucet. *2020*

- *Adaptive sequential Monte Carlo (SMC) for posterior inference and model selection among complex geological priors encoded through deep generative neural networks.*

Society of Exploration Geophysicists (SEG) Annual Meeting

Anaheim, USA

Luis Guarracino, Macarena Amaya, Santiago Perdomo, Amilcar Juarez, Gabriel Gelpi, Jonatan Pendiuk, Soledad Lagos, José Gómez, Claudia Tocho, and Jerónimo Ainchil. *2018*

- *Identification and evaluation of groundwater resources for small rural and aboriginal communities in Chaco, Argentina.*

Publications

Geophysical Journal International

Macarena Amaya, Giovanni Meles, Stefano Marelli, Niklas Linde. *2023*

- *A multifidelity adaptive sequential Monte Carlo approach applied to geophysical inversion (under review)*

Geophysical Journal International

Giovanni Meles, Macarena Amaya, Shiran Levy, Stefano Marelli, Niklas Linde. *2023*

- *Efficient Bayesian travel-time tomography with geologically-complex priors using sensitivity-informed polynomial chaos expansion and deep generative networks (under review)*

Advances in Water Resources

Macarena Amaya, Niklas Linde, Eric Laloy. *2022*

- *Hydrogeological multiple-point statistics inversion by adaptive sequential Monte Carlo*

Geophysical Journal International

Macarena Amaya, Niklas Linde, Eric Laloy. *2021*

- *Adaptive sequential Monte Carlo for posterior inference and model selection among complex geological priors*

Frontiers in Earth Science

Cheick Dumbia, Pascal Castellazzi, Alain N Rousseau, Macarena Amaya. *2019*

- *Ice mass loss mapping in the Gulf of Alaska from constrained forward modelling of GRACE data*